Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New Caledonia

Mathieu Quenu¹, Steven A. Trewick¹, Fabrice Brescia² and Mary Morgan-Richards¹

¹Ecology, College of Science, Massey University, Private Bag 11-222, 4442 Palmerston North, New Zealand; and
²Institut Agronomique Neo-Caledonien (IAC), Port-Laguerre, Paita, BP73 98890 Noumea, New Caledonia

Correspondence: M. Quenu; e-mail: mathieuquenu@hotmail.fr

(Received 2 April 2019; editorial decision 28 August 2019)

ABSTRACT

Size and shape variations of shells can be used to identify natural phenotypic clusters and thus delimit snail species. Here, we apply both supervised and unsupervised machine learning algorithms to a geometric morphometric dataset to investigate size and shape variations of the shells of the endemic land snail *Placostylus* from New Caledonia. We sampled eight populations of *Placostylus* from the Isle of Pines, where two species of this genus reportedly coexist. We used neural network analysis as a supervised learning algorithm and Gaussian mixture models as an unsupervised learning algorithm. Using a training dataset of individuals assigned to species using nuclear markers, we found that supervised learning algorithms could not unambiguously classify all individuals of our expanded dataset using shell size and shape. Unsupervised learning showed that the optimal division of our data consisted of three phenotypic clusters. Two of these clusters correspond to the established species *Placostylus fibratus* and *P. porphyrostomus*, while the third cluster was intermediate in both shape and size. Most of the individuals that were not clearly classified using supervised learning were classified to this intermediate phenotype by unsupervised learning, and most of these individuals came from previously unsampled populations. These results may indicate the presence of persistent putative-hybrid populations of *Placostylus* in the Isle of Pines.

INTRODUCTION

Species are defined using many different criteria, in order to classify individuals into distinctive groups (Darwin, 1859; Mayr, 1942; Mahner, 1993; Mallet, 1995; De Queiroz, 2007). However, the continuity of evolution means that probably no single species concept can effectively serve all circumstances. Instead, it has been argued that practical and useful definitions (operational criteria) should be prioritized when applying empirical data to the issue of species delimitation (Dubois, 2011; Vaux, Trewick & Morgan-Richards, 2016). The definition of species as morphological or genotypic clusters of individuals (Mallet, 1995) allows for the presence of individuals that are morphologically or genetically intermediate between two species, as long as these individuals are relatively uncommon. In this view, species are groups of individuals that interact and reproduce mostly together, and morphological and genotypic clustering is a consequence of the correlation of their traits. This definition accommodates evolutionary flux and so allows speciation, gene flow and selection to be studied without the circularity of defining a species by reproductive isolation. In this framework, we need to be able to delimit groups of individuals using character data and, if intermediate individuals between clusters exist, be able to determine whether they are strongly under-represented (Mallet, 1995). Robust analytical tools are necessary to both identify groups and detect intermediates. Recent advances in machine learning using supervised and unsupervised learning approaches allow identification of specimens with and without a priori labels and so provide less subjective ways of delimitating species’ limits.

In parallel with advances in machine learning, since the 1980s, there have been developments in the analysis of morphological variation (reflecting the development of a formal theory of shape and geometric morphometric methods) (Kendall, 1986; Mitteroecker & Gunz, 2009). Landmark-based geometric morphometries use the relationships between a set of landmarks to study the shape of organisms, rather than a set of traditional distance measurements (Bookstein, 1991). In this approach, shape differences between individuals are defined as any variation found among a set of homologous landmarks after they have been scaled, rotated and translated to the same criteria, a process called generalized
Procrustes analysis. Two types of landmarks are the most commonly used. Fixed landmarks sample homologous features in different individuals, whereas semi-landmarks can be used to sample the same positions on any surface located between two fixed landmarks (Zelditch, Swiderski & Sheets, 2004). Statistical analyses can then quantify the differences in shape between individuals by considering variation in a set of aligned landmarks, typically after steps of ordination and dimension reduction (Zelditch et al., 2004). Geometric methods are more sensitive to shape variation than traditional morphometry (Rohlf & Marcus, 1993; Maderbacher et al., 2006) and have been usefully applied to the study of shell shape variation of molluscs (Carvajal-Rodriguez, Conde-Padín & Roñán-Alvarez, 2005; Cruz, Pante & Rohlf, 2012; Gustafson et al., 2014; Dillon & Jaffrin, 2013; Gustafson & Bolek, 2016; Vaux et al., 2017, 2018; Rao et al., 2018; Verhaegen et al., 2018). In the present study, we combine the use of geometric morphometry, with supervised and unsupervised learning algorithms, to test whether the current species-level taxonomy of the land snail *Placostylus* on the Isle of Pines, New Caledonia, corresponds with patterns of shell morphological variation.

*Placostylus* is a genus of large terrestrial pulmonate snails that is native to the islands of the western Pacific Ocean (Breure, Grognenberg & Schulthuizen, 2010). In New Caledonia, six species are recognized according to current taxonomy (Neubert, Chérel-mora & Bouchet, 2009), but a combination of genotypic and geometric data indicate that the two most abundant nominal species-level taxa *Placostylus fimbriatus* (Martyn, 1709) and *P. porphyrostomus* (Pfeiffer, 1851) may contain additional undescribed species on the main island, Grande Terre (Dowle et al., 2015). These two species are also locally common on the Isle of Pines, a small island (152 km²) south of Grande Terre, where *P. fimbriatus* is still harvested for food (Brescia, 2011).

Analysis of 26 specimens from the Isle of Pines has shown that shell morphology seems sufficient to discriminate the two sympatric *Placostylus* species, with concordance between shell morphology and nuclear genetic clusters being unequivocal (Dowle et al., 2015). Shell differences in sympathy indicate that the size and shape of adult *Placostylus* shells are largely controlled by genetic factors, rather than environmental differences. *Placostylus fimbriatus* is larger and has a wider ventral aperture than *P. porphyrostomus*, which is usually smaller with a thicker lip around a smaller aperture. Although shell morphology was concordant with two genetic clusters in a previous study (Dowle et al., 2015), and this may contain additional undescribed species on the main island, Grande Terre, where *P. porphyrostomus* is larger and has a wider ventral aperture than *P. fimbriatus*, the two recognized species on the island and thus difficult to identify (Brescia, 2011). Available mt (mitochondrial) DNA sequence data suggest some historical genetic exchange consistent with hybridization between the two species (Dowle et al., 2015), and this may explain why these populations are intermediate in size to *P. fimbriatus* and *P. porphyrostomus*.

Based on geometric morphometric analyses of data on shell size and shell shape data, the aim of this study is to apply supervised and unsupervised learning algorithms to test whether the recognition of two *Placostylus* species on the Isle of Pines is sufficient and to quantify the frequency of potential novel phenotypes.

MATERIAL AND METHODS

**Rationale**

Supervised learning is a type of algorithm that is ‘taught’ to create a function that will link a set of features (inputs) to a set of labels (outputs) (Hastie, Tibshirani & Friedman, 2009). This is done by providing example data, called the training dataset, for which features are already associated with some known labels. A training process is then used to find the optimal function that links feature values (shell shape and size) to label values (species identification). Here, we use a class of artificial neural networks called a multi-layer perceptron for supervised learning (Ripley, 1994). Multi-layer perceptrons are recognized for their predictive efficiency and rely on the use of gradient descent and backpropagation algorithms. Associations of these algorithms with geometric morphometrics can be used to classify individuals into different categories based on shape and size information (Baylac, Villemant & Simbolotti, 2003; Dubey et al., 2006; Soda, Slice & Naylor, 2017). We chose to use this algorithm as it requires few assumptions and can be applied to a wide range of data types (Ripley, 1994). A sample of snails where we have both morphometric and genetic information from a previous study was used as a training dataset (Dowle et al., 2015). These shells form two separate morphological groups that could also be separated using population genetic tools with nuclear loci; using the genotypic cluster definition, they are regarded as two species. If the current taxonomy of *Placostylus* reflects biological reality and shell morphology can be used to separate individuals into two distinct species, we expect the supervised learning algorithm to classify without ambiguities all (or almost all) individuals in our expanded sample.

We also use Gaussian mixture models (GMMs), a class of unsupervised learning algorithm, to find the optimal number of morphological groups in our new (expanded) dataset. Unsupervised learning is a class of machine learning algorithms, which infer structure in datasets that do not have a priori labels (Hastie et al., 2009). In the case of GMMs, the dataset is considered to be a mixture of population samples with different Gaussian distributions, and the modelling process divides the data into different classes, each one corresponding to a single Gaussian distribution (Fraley & Raftery, 2006). We use GMMs because this approach can model any kind of data, as long as the data can be separated into different Gaussian components. Different models are created for a different number of hypothetical clusters. Bayesian information criteria (BIC) scores are used to determine which model is the best fit to the data, and in this way, the optimal number of clusters in our sample is estimated. If the current shell-based taxonomy of *Placostylus* in the Isle of Pines is correct, we expect an optimal GMM that divides our dataset into two groups, corresponding to the two currently recognized species.

**Sampling strategy**

A total of 337 *Placostylus* snails were sampled across eight sites on the Isle of Pines in April 2015 (Fig. 1). Sample sites were arrayed along a set of elevation gradients from the coast to the centre of the island (Fig. 1). All live snails within a 20 m radius of a fixed point were collected by hand, at three locations along each of eight transects. Two-dimensional geometric morphometric data were gathered from the 337 shells sampled in 2015 and from the 26 shells that were collected and analysed previously (Dowle et al., 2015). Among the eight population samples examined, five had previously been sampled (Vao, Gadji, Kere, Youay, Touete), and three were sampled for the first time (Comagna, Wapan, Waatchia) (Fig. 1). Once snails are sexually mature, age-related variation in shell shape and size are considered to be minimal (F. M. Brescia et al., 2008). Only shells of adult snails (identified by the thickened lip) were photographed for shape analysis. As these snails are hermaphroditic, it was not necessary to consider sexual dimorphism.

**Geometric morphometrics**

The set of two-dimensional landmarks used for this study was derived from a set of optimal landmarks and semi-landmarks established in an analysis of New Zealand *Placostylus* (Daly, 2016). A total of 40 landmarks and semi-landmarks were used, consisting of nine fixed landmarks and 31 sliding/semi-landmarks (Fig. 2). Digital images of the ventral surface of each shell were obtained using a Canon EOS 600D with EF100 mm f/2.8 USM macro lens after careful positioning of the shell in a bed of sand of contrasting...
For subsequent analyses.

variance of the rest of the dataset, which we deemed to be negligible.

Placostylus nized

of the sum of the arrays coming from the centroid position of

Coordgen8. This size estimation is calculated as the square root

the 'centroid size' tool included in the Procrustes analysis in

2004). Principal component analysis (PCA) was executed on the

and semi-landmarks was conducted using the programme tpsDig2

section of the lips and outer shell (Fig. 2). Digitization of landmarks

their teeth perpendicular to a line from the shell apex to the inter-

'combs' were placed manually using Adobe Photoshop CS6, with

orderto placethesemi-landmarksalongthecurvesof theshells,two

personto minimize operator error (Schilthuizen & Haase, 2010). In

Errors linked to shell manipulation and digitization were assessed

by experimental replication; this involved taking five photographs

of the same shell and comparing the variance attributed to the

repeat process with the variance of the whole dataset using the

mclust. For GMMs, we used the R package mclust v. 5.4 (Fraley &

Raftery, 2006). GMMs were built using shell size and the

morphol.disparity function in the package geomorph v. 3.1.2 (Adams,

Collyer & Kaliontzopoulou, 2018) in R v.3.6.1 (R Development

Core Team & R Core Team, 2017). Overall, the variance linked to

manipulation and digitization error corresponded to <0.3% of the

variance of the rest of the dataset, which we deemed to be negligible

for subsequent analyses.

Supervised learning algorithms

For the training process of the supervised learning algorithm, we

used 26 shells that had been genetically assigned to one of two

species (our training dataset). These specimens had been unequiv-

ically assigned by Dowle et al. (2015) to one of the two recog-

nized Placostylus species on the basis of a nuclear DNA dataset

comprising 661 SNPs. Neural network models were built using the

neuralnet function in the R package neuralnet v. 1.44.2 (Fritsch

et al., 2016). Input data (shape and size) were normalized and

scaled using a minmax function. Neural network models con-

taining two hidden layers of five and three neurons were used

(Fig. 3).

We first tested the efficacy of different neural network models

using exhaustive cross validation on our training dataset (\(n = 26\)).

For model validation purposes, this dataset was split into differ-

ent train and test datasets of various proportions, using different

numbers of input variables (for split proportions, number of input

variables and other details; see Supplementary Material Table S1).

The objectives of this process were to estimate whether we had a

sufficiently large training dataset and to determine the optimal

number of input variables. Efficacy of the models was assessed by

looking at the mean proportion of individuals assigned correctly in

the test datasets for a large number of replicates. We found that the

optimal models used only two PCs and one size of a variable. In

these models, the cross validation correctly identified all snails most

of the time (mean proportion of correct assignment was between

98.8 and 99.8% for different-sized training datasets; Supplementary

Material Table S1).

We then trained neural network models using the whole training

dataset and used trained models on our new dataset of 337 shells

to assign each snail to one of the two species. We repeated this

process for a large number of neural network models (\(x 1,000\))

to avoid overfitting issues arising from single model interpreta-

tion (Zhou, 2009), retaining the average value of the neural net-

works outputs as assignment score for each individual (ensemble

learning).

Unsupervised learning algorithms

For GMMs, we used the R package mclust v. 5.4 (Fraley &

Raftery, 2006). GMMs were built using shell size and the

same number of PCs as in the supervised learning analyses

(1 and 2). Different models were built for a range of a priori

clusters in the dataset, with the optimal model being selected

on the basis of BIC scores (Supplementary Material Fig S1).

Results of the modelling process with additional input variables

(PC1–5, shell size) were also computed and yielded similar

results.
RESULTS

First and second PCs of variation of the geometric morphometric analysis accounted for 35.3 and 15.5% respectively of the total variance in shell shape. Warp transformation grids indicate that spire height, aperture size and lip thickness varied the most (Fig. 2).

Supervised learning algorithms

Using the supervised learning algorithm, individual specimens were assigned to one or other of the two species, *Placostylus fibratus* or *P. porphyrostomus*, if their assignment probability was > 95%. On the basis of this criterion, 224 individual snails were identified as *P. fibratus* (67%) and 61 as *P. porphyrostomus* (18%); 51 (15%) could not be assigned to either species (Table 1, Figs 2B, 4A). Most of the unassigned individuals came from two populations: Comwagna (28 individuals) and Wapan (15 individuals). The six other population samples contained individuals that could be confidently identified to either *P. fibratus* or *P. porphyrostomus*, with only a few unassigned individuals (Table 1). Youaty was the only population sample with just one snail species (*P. fibratus*).

Unsupervised learning algorithms

The shell size and shape variations had an optimal fit to a GMM with three clusters (ellipsoidal, equal volume, shape and orientation (EEE) model; Supplementary Material Fig. S1). Assignment probabilities to the three phenotypic clusters were computed for each of the 337 individuals (Table 1, Fig. 4). Two of the phenotypic clusters identified using shell size and shape variables correspond to the recognized species *P. fibratus* and *P. porphyrostomus*. A third cluster was also observed. This consisted mostly of individuals from Comwagna and Wapan, which were unassigned by the supervised learning algorithm (Fig. 4, Supplementary Material Table S1). The snails in the third cluster were intermediate in shell size and shape to *P. fibratus* and *P. porphyrostomus* (Fig. 5).

Using GMMs, 226 (67%) individuals were assigned to the *P. fibratus* cluster, 56 (17%) to the *P. porphyrostomus* cluster and 54 (16%) to the intermediate group. The assignment of individuals was similar to the result produced by the supervised learning algorithm, with specimen assignment probabilities <0.95 (from supervised learning) placed in the third cluster (Fig. 4B). Only for 11 of the 337 snails (3.27%) did the classification differ between the two methods: the three individuals not assigned by supervised learning were assigned by GMM to *P. fibratus* (two individuals) and *P. porphyrostomus* (one individual), whereas the eight shells classified as *P. porphyrostomus* by the supervised learning algorithm were classified as either *P. fibratus* or the intermediate type by GMM. The geographic source of the snails that contribute to the three morphological clusters identified by GMM suggests that the intermediate shell phenotype has a limited distribution. The Youaty population sample consisted only of *P. fibratus*, but the Touete, Waatchia, Gadji, Kere and Vao population samples contained both *P. fibratus* and *P. porphyrostomus*. The Wapan population sample contained examples of all three clusters, while the Comwagna population sample consisted mostly of individuals assigned to the intermediate type, with a few individuals of *P. porphyrostomus* (Table 1).
Table 1. Numbers of *Placostylus* in the three phenotype clusters (species) as assigned by supervised and unsupervised analysis of shell size and shape for eight population samples from the Isle of Pines, New Caledonia.

<table>
<thead>
<tr>
<th>Population</th>
<th>Comwagna</th>
<th>Gadjj</th>
<th>Kere</th>
<th>Touete</th>
<th>Vao</th>
<th>Waatchia</th>
<th>Wapan</th>
<th>Youaty</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised learning classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. fibratus</em></td>
<td>0</td>
<td>29</td>
<td>31</td>
<td>26</td>
<td>52</td>
<td>38</td>
<td>23</td>
<td>25</td>
<td>224</td>
</tr>
<tr>
<td><em>P. porphyrostomus</em></td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>16</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>Not assigned</td>
<td>28</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>Unsupervised learning classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. fibratus</em></td>
<td>0</td>
<td>29</td>
<td>32</td>
<td>26</td>
<td>53</td>
<td>38</td>
<td>23</td>
<td>25</td>
<td>226</td>
</tr>
<tr>
<td><em>P. porphyrostomus</em></td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>Intermediate phenotype</td>
<td>32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>54</td>
</tr>
</tbody>
</table>

Individuals were accepted as assigned to a cluster only if their assignment score was >0.95.

Figure 4. Assignment probabilities calculated for shell shape and size variation in eight population samples of *Placostylus* from the Isle of Pines. A. Neural network supervised learning analysis for two phenotypes (the two recognized species). Colour coding: grey, *P. fibratus*; white, *P. porphyrostomus*. B. An unsupervised analysis based on an optimal Gaussian mixture models that found three phenotypes (the two recognized species and the intermediate phenotype). Colour coding: grey, *P. porphyrostomus*; black, *P. fibratus*; white, intermediate phenotype.

Figure 5. A. Relative frequency of three phenotypes of *Placostylus* in eight population samples from the Isle of Pines, New Caledonia, as indicated by the optimal GMM, using the first two PCs of variation in shell shape (50.8% of total shape variation) and shell size. B. Gaussian mixture model–based classification of *Placostylus* in relation to variation of shell size and shape along PC 1 (35.3%). Colour coding: red, *P. porphyrostomus*; blue, *P. fibratus*; yellow, intermediate phenotype.
DISCUSSION

Shell morphology, separable into shape and size components through geometric morphometrics, is controlled by a suite of heritable characters that can be influenced by environmental factors. Aquatic gastropods have provided evidence that phenotypic plasticity of shell size and shape is widespread (Bourdeau et al., 2015). The New Zealand land snail Placostylus ambiguus shows size variation that is, in part, due to environmental factors (Parrish, Stringer & Sherley, 2014). However, previous work on the two New Caledonian Placostylus species P. fibratus and P. porphyrostomus has shown that they are genetically distinct, despite coexisting at many locations on the Isle of Pines (Brescia et al., 2008; Dowle et al., 2015). Distinct shell traits of sympatric species are unlikely to be due to phenotypic plasticity, as the snails experience the same environment.

When applied to our new dataset, supervised learning algorithms failed to unambiguously assign all snails into one or other of the two species that are recognized on the basis of traditional shell-based taxonomy. Most of the specimens with low assignment probabilities originate from two newly sampled locations on the Isle of Pines. The presence of many unassigned snails suggests that the classification into two morphological groups (species) is too simplistic in the context of our expanded dataset. At the same time, unsupervised learning algorithms indicate the presence of a third group, and this result is consistent with our finding that supervised learning failed to classify some individuals. This third group is intermediate in both shape and size to the two recognized species (Fig. 5).

The snails from the Comwagna district, which constitute the majority of intermediate-type individuals, have never been harvested for commercial purposes due to their smaller size. To date, these individuals have been considered to be a subgroup of P. fibratus (Brescia et al., 2008). Our analysis suggests that they are distinct and could potentially represent an unusual population of hybrid individuals. While available data on mt DNA haplotype distributions suggest historical hybridization between P. fibratus and P. porphyrostomus, the presence of morphologically and genetically distinct sympatric populations does not indicate ongoing hybridization (Dowle et al., 2013). However, here, our morphological data suggest hybridization could go beyond simple mitochondrial introgression and could be the origin of an exclusively hybrid population at Comwagna and rare hybrids at other localities.

Hybridization can be viewed as a gradient that extends from rare introgression through stable hybrid zones to the formation of exclusively hybrid populations (Nieto Feliner et al., 2017). Introgression can generate gene flow between distinct species, but parental lineages can remain distinct when hybrids normally backcross to one of the parental forms (Anderson, 1953; Folk et al., 2018). Observations of five populations of Placostylus from the Isle of Pines (Gadji, Kere, Touete, Waatchia, Vao) seem to be consistent with the rare introgression hypothesis suggested by mt DNA haplotype data; this is because of the presence of just two distinctive parental phenotypic clusters and a small number (1.4–2.9%) of individuals with intermediate shell phenotype in these five populations (Fig. 5). Given the absence of intermediates at localities where P. fibratus and P. porphyrostomus are sympatric, we could apply the genotypic cluster definition (Mallet, 1995) and recognize two species. On the other hand, the numerical dominance of the intermediate phenotype in the population samples from Wapan and Comwagna suggests that if hybridization explains this phenotype, then hybridization has been locally common and has led to the loss of distinct parental phenotypic clusters at these sites.

Environmental conditions instead of interspecific hybridization may potentially explain the prevalence of individuals with an intermediate phenotype. Of the eight locations where snails were sampled, Comwagna is situated on red ferrallitic (=literite) soils, whereas all the others are on uplifted coral reef terrain (Lagarde & Ouetcho, 2016). Experiments have shown a positive correlation between calcium availability and shell growth in some terrestrial gastropods (Bechy & Richmond, 2007). Phenotypic plasticity is likely to account for a part of the morphological diversity observed for New Caledonian Placostylus (Brescia et al., 2008; Dowle et al., 2015). Lower calcium availability on ferrallitic soils is, therefore, a plausible explanation for the intermediate phenotype at Comwagna, but this is challenged by the presence of that phenotype on calcium-rich substrate at Wapan. Although hybridization and recent gene flow between P. fibratus and P. porphyrostomus seems to be the most likely explanation for the presence of the intermediate phenotype, at this stage we cannot rule out the possibility that the intermediate phenotype represents a novel lineage distinct from P. fibratus and P. porphyrostomus, or the possibility of an explanation based on environmental factors.

Here, we used two approaches to classify samples: one uses a training dataset, the other does not. The first method seeks to classify individuals within a framework of knowledge, while the second explores a dataset to find optimal divisions. Both methods are subject to overfitting issues (Adams, Rohlf & Slice, 2004). Supervised learning can misclassify individuals that are not inside the range of the training data, and unsupervised learning can underestimate or overestimate the number of clusters in a dataset. We suggest that low support values for individual classification from supervised learning can be biologically interesting. While low support assignment can arise from technicalities such as a small training dataset or mislabelling, here, the concordance of supervised learning results with unsupervised learning classification suggests our models were not lacking power. Our results, therefore, support the use of both methods, rather than either of them individually. Overall, our work joins the growing number of studies that have demonstrated that the association of geometric morphometrics, and machine learning can be useful in addressing biological questions (Dubey et al., 2006; Bochlaer & Schultheiß, 2010; Mapp et al., 2017; Nattier et al., 2017; Soda et al., 2017; Fang et al., 2018).

ACKNOWLEDGEMENTS

We thank the New Caledonian people, the Institut Agronomique New-Caledonien and members of the Phoenix group for Evolutionary Ecology and Genetics (evolves.massey.ac.nz). This research was supported by the Institute Agronomique Neo-Caledonien and a Massey University PhD scholarship (awarded to MQ).

REFERENCES


