Hide and seek shark teeth in Random Forests: Machine learning applied to *Scyliorhinus canicula* populations

Fidji Berio[1,2] [*], **Yann Bayle**[3] [*], **Daniel Baum**[4], **Nicolas Goudemand**[2], **and Mélanie Debiais-Thibaud**[1]

[1]**ISEM, CNRS, IRD, EPHE, Univ. Montpellier, Montpellier, Place Eugène Bataillon, 34090 Montpellier, France**
[2]**Institut de Génomique Fonctionnelle de Lyon, UMR 5242, Univ. Lyon, École Normale Supérieure de Lyon, Centre National de la Recherche Scientifique, 46 Allée d'Italie, F-69364 Lyon Cedex 07, France**
[3]**Université de Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR5800, F-33400 Talence, France**
[4]**Zuse Institute Berlin, Department of Visual and Data-Centric Computing, Takustr. 7, 14195 Berlin, Germany**
[*]**These authors contributed equally to this work.**

Corresponding author:

Fidji Berio[1,2]

Email address: fidji.berio@gmail.com

ABSTRACT

Shark populations that are distributed alongside a latitudinal gradient often display body size differences at sexual maturity and vicariance patterns related to their number of tooth files. Previous works have demonstrated that *Scyliorhinus canicula* populations differ between the northeastern Atlantic Ocean and the Mediterranean Sea based on biological features and genetic analysis. In this study, we sample more than 3,000 teeth from 56 *S. canicula* specimens caught incidentally off Roscoff and Banyuls. We investigate population differences based on tooth shape and form by using two approaches. Classification results show that the classical geometric morphometric framework is outperformed by an original Random Forests-based framework. Visually, both *S. canicula* populations share similar ontogenetic trends and timing of gynandric heterodonty emergence but the Atlantic population has bigger, blunter teeth, and less numerous accessory cusps than the Mediterranean population. According to the models, the populations are best differentiated based on their lateral tooth edges, which bear accessory cusps, and the tooth centroid sizes significantly improve classification performances. The differences observed are discussed in light of dietary and behavioural habits of the populations considered. The method proposed in this study could be further adapted to complement DNA analyses to identify shark species or populations based on tooth morphologies. This process would be of particular interest for fisheries management and identification of shark fossils.

INTRODUCTION

The recognition of disjunct shark populations has opened to new questionings on marine ecosystem connectivity and consequences of gene flows on species evolution. Shark population distributions are structured by ecological habits of species (e.g., degree of habitat fidelity), reproductive strategies (e.g., use of nursery areas by females), dispersal ability (e.g., type of reproduction, migratory behaviour), and environmental barriers to gene flow (e.g., oceanic basins, geological climatic events) (Lucifora et al., 2003; Rodríguez-Cabello et al., 2004; Portnoy et al., 2010; Veríssimo et al., 2010; Karl et al., 2012; Kousteni et al., 2015). Populations of a given shark species sometimes display vicariance patterns of number of vertebrae (Gruber and Compagno, 1981) and tooth files (McEachran and Martin, 1977; Templeman, 1984; Lucifora et al., 2003). In some species, the total length of a specimen at sexual maturity also differs between populations, as in bonnethead sharks *Sphyrna tiburo* (Parsons, 1993; Lombardi-Carlson et al., 2003), shortspine spurdogs *Squalus mitsukurii* (Taniuchi and Tachikawa, 1997), starspotted smooth-hounds *Mustelus manazo* (Yamaguchi et al., 1998, 2000), and cloudy catsharks *Scyliorhinus torazame* (Horie and Tanaka, 2002). Such differences in specimen size at sexual maturity that are reported among shark populations have been hypothesized to result from genetic or environmental constraints, or both, but the combination between these factors is difficult to evaluate (Lombardi-Carlson et al., 2003).

When observed, these size differences at sexual maturity are often distributed alongside a latitudinal gradient and shark populations inhabiting higher and colder latitudes are significantly bigger (Leloup and Olivereau, 1951; Parsons, 1993; Taniuchi and Tachikawa, 1997; Yamaguchi et al., 2000; Horie and Tanaka, 2002; Lombardi-Carlson et al., 2003; Capapé et al., 2014; Kousteni and Megalofonou, 2019). The warmer temperatures at lower latitudes (Blackburn et al., 2008) are thought to limit the energy allowed for somatic growth by inducing increased energy expenditure (Parsons, 1993; Carlson and Parsons, 1997) and also trigger early sexual maturity (Parsons, 1993; Yamaguchi et al., 2000; Goren, 2014).

Scyliorhinus canicula is an abundant benthic species in the eastern Atlantic Ocean (from Senegal to the UK) and Mediterranean Sea that inhabits depths from a few meters to 500 m (most commonly found around 110 m) (Ellis and Shackley, 1997; Compagno, 1984; Rodríguez-Cabello et al., 2004). Support for population differentiation was raised by morphometric and genetic diversity analyses within the distribution range of this species (Barbieri et al., 2014; Capapé et al., 2014; Kousteni et al., 2015). Population genetic structures have been attributed to the philopatric behaviour of *S. canicula* and to its low dispersal ability across basins (Leloup and Olivereau, 1951; Mellinger et al., 1984; Rodríguez-Cabello et al., 2004; Barbieri et al., 2014; Capapé et al., 2014; Kousteni et al., 2015; Anastasopoulou et al., 2016; Kousteni and Megalofonou, 2019).

Despite genetic structuration, Mediterranean *S. canicula* populations exhibit very slight body size differences at sexual maturity (Barbieri et al., 2014; Capapé et al., 2014; Kousteni et al., 2015; Kousteni and Megalofonou, 2019). Conversely, populations differ greatly in body size at sexual maturity between the North Atlantic Ocean (Bristol Channel, UK) and the Mediterranean Sea (Leloup and Olivereau, 1951; Mellinger et al., 1984; Rodríguez-Cabello et al., 1998; Kousteni et al., 2010; Capapé et al., 2014; Kousteni and Megalofonou, 2019). Furthermore, there is currently no recognition of unequivocal

morphological differences between *S. canicula* or any shark populations, other than body length at sexual maturity which differs along vast latitudinal gradient and separated marine environments. This is an important gap in our ability to follow population dynamics, notably for sharks that have long generation time and struggle withstanding or recovering from accidental catches and fishing (Smith et al., 1998; Cortés, 2000; Dulvy et al., 2014, 2021). Commercial frauds are regularly checked, but the identification of species mainly relies on a framework based on DNA barcoding (Barbuto et al., 2010; Melo Palmeira et al., 2013; Almerón-Souza et al., 2018), whose cost, time, and dependence to a wet lab are limitations to the current inspection process. Therefore, we investigated the possibility to identify differences in tooth morphology between catshark populations, considering that shark tooth shape undergoes morphological changes with ontogeny, sexual maturation and diet (Powter et al., 2010; Moyer and Bemis, 2016; Tomita et al., 2017; Cullen and Marshall, 2019; Berio et al., 2020) and that life history traits differ between Atlantic and Mediterranean *S. canicula* populations (Leloup and Olivereau, 1951; Ivory et al., 2004; Bendiab et al., 2012; Capapé et al., 2014).

This work tests a feature-based framework to reliably discriminate shark populations based on their tooth morphological characteristics. The assessment of shark tooth form differences is successfully achieved with geometric morphometrics (Whitenack and Gottfried, 2010; Soda et al., 2017; Cullen and Marshall, 2019; Berio et al., 2020). Linear Discriminant Analysis (LDA) is a frequently used machine learning algorithm to discriminate between groups based on geometric morphometric data (Mitteroecker and Bookstein, 2011; MacLeod, 2017; Doyle et al., 2018). The performances of LDA depend on a much higher number of items (e.g., teeth) as compared to the number of features (e.g., aligned coordinates). Yet, these conditions are difficult to meet in biological datasets, which often implies a feature reduction step through a Principal Component Analysis (PCA) (Fort and Lambert-Lacroix, 2005; Pechenizkiy et al., 2006; Sheets et al., 2006; Archer and Kimes, 2008). Moreover, subtle discriminant patterns can also be missed when reducing data dimensionality to the first PCA axes prior to a classification task (MacLeod, 2018). Traditional and geometric morphometric studies use several machine learning algorithms for classification, but often focus on the classification performances –rather than on interpretable features (Santos et al., 2014; Navega et al., 2015; Doyle et al., 2018; Courtenay et al., 2019). As opposed to LDA, Random Forests (Breiman, 2001) usually outperform other machine learning algorithms for supervised classification (Caruana and Niculescu-Mizil, 2006; Domínguez-Rodrigo and Baquedano, 2018; Doyle et al., 2018; Püschel et al., 2018; Courtenay et al., 2019). Furthermore, Random Forests provide intuitive and interpretable importance measures of feature contribution to classification, is highly resistant to overfitting, and do not require feature reduction prior to the analysis to achieve good performances (Díaz-Uriarte and Alvarez de Andrés, 2006; Archer and Kimes, 2008; Doyle et al., 2018).

This work is a proof of concept that shark populations can be discriminated based on tooth morphology. We take advantage of the geometric morphometric and machine learning methods to test for differences between *S. canicula* teeth from northeast Atlantic and Mediterranean populations that exhibit clear body size differences at sexual maturity (Capapé et al., 2014; Kousteni and Megalofonou, 2019). By using Random Forests to classify the teeth of *S. canicula* population samples of different geographic origin, we aim not only to challenge the traditional geometric morphometrics workflow, but also to

130 provide indications of discriminant tooth form features between *S. canicula* populations.
131 Our work aims to provide an alternative tool for discrimination of shark species and
132 populations based on dental features and to set the basis for future research.

MATERIALS AND METHODS

Sampling

135 *S. canicula* populations were sampled in 2018 and 2019 in two localities separated by
136 over 2,000 nautical miles: off Roscoff (France, northeast Atlantic Ocean) and Banyuls-
137 sur-mer (France, western Mediterranean Sea). The Atlantic specimens were sampled
138 at the Roscoff fishmarket and were provided by the Station Biologique de Roscoff
139 and the University of Montpellier. The Mediterranean specimens were incidentally
140 caught during experimental surveys of the Observatoire Océanologique de Banyuls-sur-
141 Mer or were formerly euthanized for independent experiments led by the Observatoire
142 Océanologique de Banyuls-sur-Mer and the University of Montpellier. Biological
143 samples were preserved in 70% ethanol. Specimens were selected based on their
144 total body length (TL, from the tip of the snout to the tip of the tail) to represent
145 three ontogenetic stages within each population, hereafter referred to as "hatchling",
146 "juvenile", and "mature". The TL of mature Atlantic and Mediterranean specimens
147 was selected according to Ivory et al. (2004) (>53.5 cm TL for males, >57 cm TL
148 for females; reference lengths for 50% maturity) and Leloup and Olivereau (1951)
149 (≥ 40 cm TL for both sexes; reference length for "frequently" observed maturity),
150 respectively. Hatchling specimens were euthanized just after hatching. Mediterranean
151 juveniles were selected between hatchling [>9 cm TL for both sexes (Leloup and
152 Olivereau, 1951)] and sexually mature [<37.5 cm TL for both sexes; reference length
153 for first sexual maturity (Leloup and Olivereau, 1951)] stages. Atlantic juveniles were
154 selected between hatching [>10.5 cm TL for both sexes (Ellis and Shackley, 1997)]
155 and first sexual maturity [<49 cm and <52 cm in males and females, respectively (Ellis
156 and Shackley, 1997; Ivory et al., 2004)]. The Mediterranean sample is composed of
157 six hatchling (3 females, 3 males; 9.1 ± 0.3 cm TL; 8.8 to 9.5 cm TL), 10 juvenile
158 (5 females, 5 males; 26.8 ± 4.1 cm TL; 21 to 31 cm TL), and nine mature (5 females,
159 4 males, 42.9 ± 2.7 cm TL; 40 to 47 cm TL) specimens. The Atlantic sample includes
160 11 hatchling (6 females, 5 males, 11.9 ± 1.4 cm TL; 10.2 to 13.9 cm TL), 10 juvenile
161 (5 females, 5 males, 34.4 ± 1.5 cm TL; 32 to 36 cm TL), and 10 mature (5 females,
162 5 males, 58.9 ± 2.7 cm TL; 56 to 64 cm TL) specimens. The maturity assessment of
163 specimens was not conducted because only the heads were collected for the juvenile and
164 sexually mature specimens after body length was recorded in marine stations. Attempts
165 were made to equally sample specimens from both sexes (F, female; M, male) within
166 each category. In addition, we estimated the approximate age of the specimens using
167 the von Bertalanffy growth parameters for the Atlantic and Mediterranean populations
168 provided by Ivory et al. (2004) and Bendiab et al. (2012) respectively because these
169 studies provide sex-specific growth curves. Ivory et al. (2004) estimated von Bertalanffy
170 parameters based on vertebral growth increment counts, whereas Bendiab et al. (2012)
171 estimated these parameters by analyzing length-frequency distributions.
172   We used the growth parameters (K, coefficient of growth; $L\infty$, asymptotic length; Lt,
173 length at age t; $t0$, theoretical age at which the size is zero) from literature data (Ivory

174 et al., 2004; Bendiab et al., 2012) (Figure 1) to compute von Bertalanffy growth curves
175 with the following growth equation: $Lt = L\infty[1 - e^{-K(t-t0)}]$.
176 The ages were estimated for each specimen according to its sex and population and the
177 age estimates were subsequently averaged per sex, ontogenetic stage, and population.
178 We estimated the age of Mediterranean hatchlings to be 0 year. Female and male Atlantic
179 hatchlings are 0.2 and 0.1 year old, respectively. Juvenile females in the Mediterranean
180 population are 0.8 year old and juvenile males are estimated to be 1.1 year old. Mean
181 age of juveniles from the Atlantic is 3.1 years for both sexes. Within the Mediterranean
182 population, sexually mature females and males are 3.7 and 3.9 years old, respectively.
183 The sexually mature specimens from the Atlantic were estimated to be 9.9 years old
184 for females and 7.9 years old for males. However, the parameters for the Atlantic and
185 Mediterranean specimens were estimated with different methods, the results of which
186 depend either on calcification degree (Natanson et al., 2018) and periodicity, or on
187 sample size and reproducibility of the results (Schwamborn et al., 2019). The estimated
188 ages of the specimens sampled in this study are therefore not strictly comparable and
189 should be interpreted in light of the method limitations, but are useful indicators of
190 differences between the populations and ontogenetic stages considered.
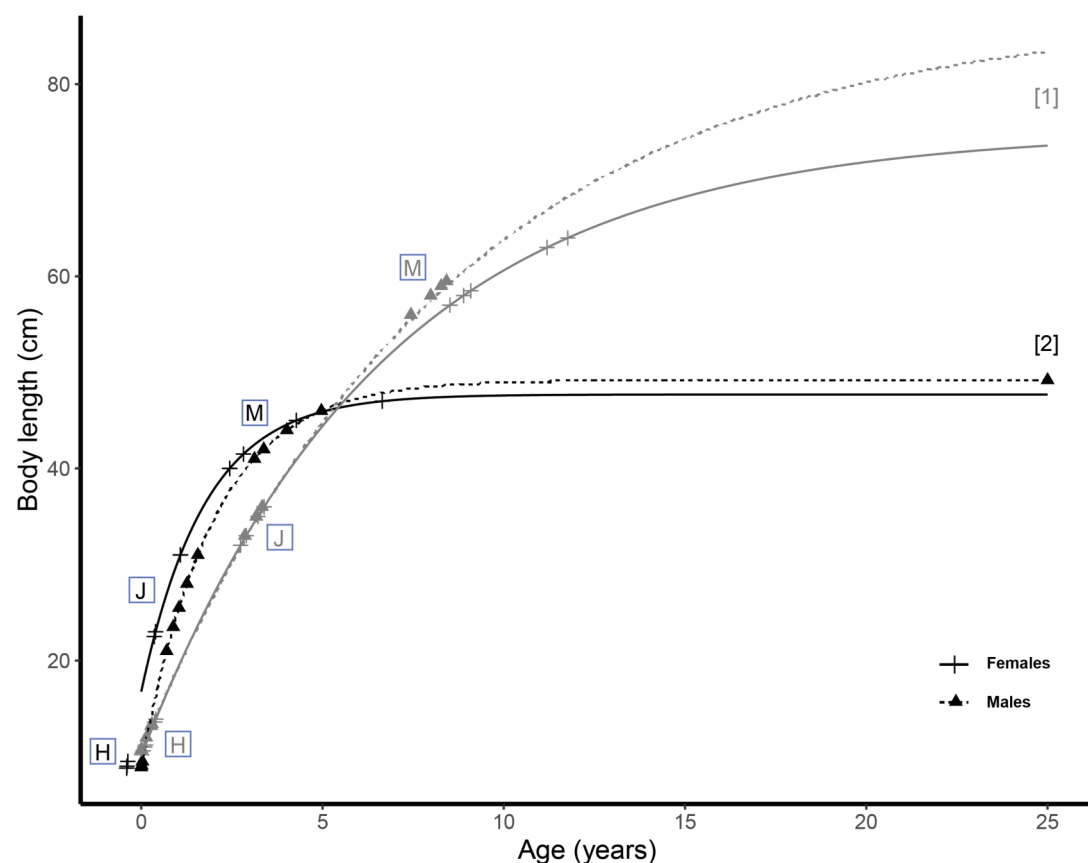
Figure 1. Growth curves of *Scyliorhinus canicula* from northeast Atlantic Ocean and
Mediterranean Sea. Von Bertalanffy growth parameters retrieved from [1] Ivory et al.
(2004) and [2] Bendiab et al. (2012). Markers represent the specimens used in the
current study. Black, Mediterranean specimens; grey, Atlantic specimens. H, hatchling;
J, juvenile; M, mature.

Data acquisition

The jaws were microCT scanned using a Phoenix Nanotom S or an EasyTom 150 with voxel sizes ranging from 6.0 μm to 30.8 μm and 3D volumes were reconstructed using the phoenix datos x2 (v2.3.0) reconstruction or xact softwares (v11025). The surfaces are available in Berio et al. (2022). 3D surfaces of right palatoquadrate (upper) and Meckelian (lower) teeth were generated with the AmiraZIBEdition software (v6.5) (Stalling et al., 2005), extracted using the ContourTreeSegmentation module, and were labelled according to their mesio-distal position along the jaw ($N_{Atlantic}$ = 1,757 and $N_{Mediterranean}$ = 1,542, Figure 2A). Seven landmarks (respectively numbered 1, 13, 15, 17, 19, 21, and 33 in Figure 2, Supp. Mat. 1, and Supp. Mat. 2) and 31 semilandmarks were placed on each 3D tooth surface. The semilandmarks were made denser on the lateral sides of the teeth, where modifications of accessory cusp number are reported during *S. canicula* ontogeny (Debiais-Thibaud et al., 2015) (Figure 2B and see Berio et al. (2020) for similar trend in *S. stellaris*).
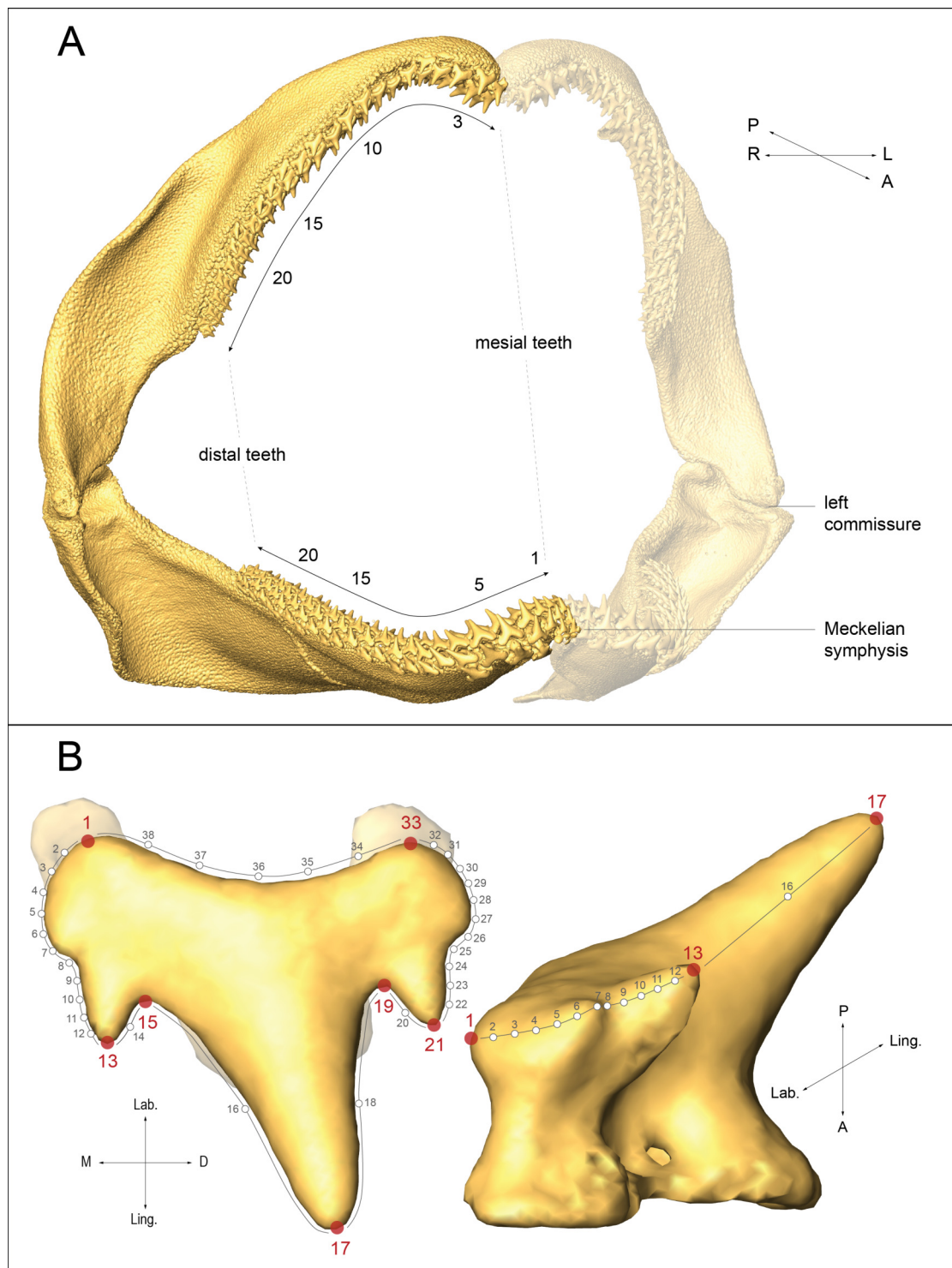
Figure 2. Labelling and landmarking *Scyliorhinus canicula* teeth. A) Mesio-distal numbering of right palatoquadrate and Meckelian teeth in a mature male from the Atlantic population (59 cm TL); B) Right Meckelian lateral tooth of the specimen in A with numbering of landmarks (red dots) and semilandmarks (small circles on grey lines). Dorsal (left) and mesial (right) views. A, anterior; D, distal; L, left; Lab., labial; Ling., lingual; P, posterior; R, right.

Data analyses

A Generalised Procrustes Analysis was performed on 3D coordinates that were formerly preprocessed following Berio and Bayle (2020). Semilandmarks were allowed to slide based on minimised bending energy (Bookstein, 1991). The structure of the dataset was first investigated through a PCA and centroid size patterns, and tooth centroid sizes were used as a proxy for tooth size. The centroid size is computed as the square root of the sum of squared distances between all landmarks and semilandmarks and the centroid of a form (Webster and Sheets, 2010; Klingenberg, 2016). The slopes of allometric patterns between populations of same sex was assessed with ANCOVAs and the interaction between shape data and size was tested using linear regression models. The shape data used to test for allometry were PC axes, whose number was determined following the procedure described by Evin et al. (2013). The relationship between shape data and size was tested using centroid size values of teeth and the specimens TL. Allometric vectors were computed with linear models and pairwise comparisons allowed to compare lengths and angles between the vectors of populations of same sex.

Supervised classification of teeth from Atlantic and Mediterranean populations was first achieved with LDA on the same PC axes used for testing allometry, without and with tooth centroid sizes [respectively tooth shapes and forms (Klingenberg, 2016)]. The results from the LDA model were compared with those obtained with Random Forests on tooth shape (Procrustes coordinates) and form (Procrustes coordinates + tooth centroid sizes) (Breiman, 2001). A five-fold cross-validation was performed in both methods (train set=80% and test set=20%).

In Random Forest models, the number of trees was set to 500 and the minimal node size was set to 1. The models were allowed to sample among 114 (shape) to 115 (form) variables to split each internal node. A good fit of the models was determined through the comparison between the accuracies reached on the train and test sets, and for both models (on shape and form), the difference between these metric values was 0.2%. The feature importance for the classification was assessed with a measure of Mean Decrease Accuracy (MDA) and features with importance values $\geq 1.0\%$ were commented (Breiman, 2001).

The classification performances reached by the models were compared based on three metrics: the accuracy, precision, and recall. The accuracy is an average of precision values for the Atlantic and Mediterranean populations. The precision is the number of items correctly assigned to a group (e.g., Atlantic), as compared to all items (Atlantic and Mediterranean) classified in this same group (Atlantic). The recall is interpreted as the number of specimens correctly assigned to a group (e.g., Atlantic), as compared to the total number of specimens actually belonging to this class (Atlantic). Additionally, a detailed confusion matrix is proposed to identify which teeth of males, females and hatchling, juvenile, and mature specimens were classified the best.

The classification results on shape and form were compared to evaluate the contribution of the centroid size to population discrimination. The overall results of LDA and Random Forest models on tooth shape and form were further compared.

The geometric morphometric analyses were performed using the geomorph package (v3.1.1) and supervised classification was computed using the MASS package (v7.3.53) for LDA and the randomForest package (v4.6.14) for Random Forests with R software (v4.0.3) (Liaw and Wiener, 2002; Venables and Ripley, 2002; Adams et al., 2019).

RESULTS

Visual description

The tooth diversity presented in Figure 3 is a selection of examples among the whole variation observed in the dataset. It provides a broad overview of the main morphological differences and associated factors in *S. canicula* tooth forms. Differences in mesio-distal location of a tooth are usually linked with an addition of accessory cusps (Figure 3A-B) and an increase of the main cusp bending. Gynandric heterodonty at sexually mature stage is characterised by more accessory cusps in females as compared to males (Figure 3C-D). Along the ontogeny, new teeth undergo size increase, as well as the addition of accessory cusps, except after sexual maturation (Figure 3E-F), when males experience a decrease in number of accessory cusps, as opposed to females. The interpopulational tooth form differences involve less accessory cusps and blunter teeth in the Atlantic population than in the Mediterranean population (Figure 3G-H).
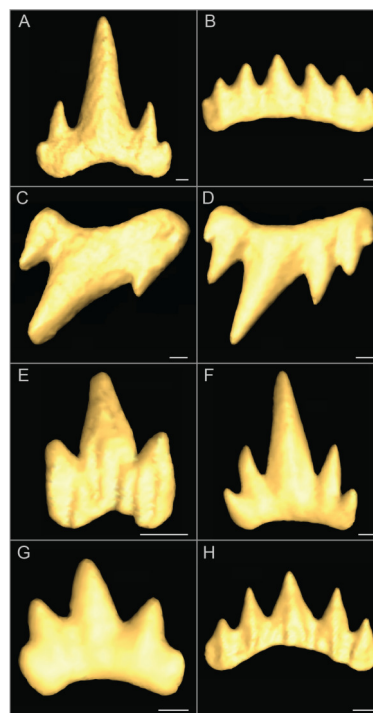
Figure 3. Examples of tooth morphological differences in *Scyliorhinus canicula*. (A-B) Form differences between tooth files: Meckelian teeth from file 5 (A) and 20 (B) of an Atlantic mature female; (C-D) Form differences between sexes: Palatoquadrate teeth from file 15 of Mediterranean mature male (C) and female (D); (E-F) Form differences between ontogenetic stages: Meckelian teeth from file 5 of Atlantic hatchling (E) and juvenile (F) males; (G-H) Form differences between populations: Meckelian teeth from file 15 of Atlantic (G) and Mediterranean (H) juvenile males. Scale bars are 100 μm.

We report significantly more Meckelian tooth files in Atlantic juvenile females (23 ± 2) as compared to the Mediterranean ones (20 ± 1) (permutation t test, t=3.29, p-val=0.04), all other interpopulation tests being not statistically significant (permutation

267 t tests, p-vals> 0.11).

Geometric morphometrics

269 The first two principal components gather 61% of the total variation in the dataset.
270 The Atlantic and Mediterranean populations are not visually discriminated in the mor-
271 phospace (Figure 4). Extreme shapes for PC1 and PC2 suggest that most variation of
272 the dataset relates to the number of accessory cusps, to their relative size compared to
273 the main cusp, and to the mesio-distal bending of the main cusp (Figure 4). In addition,
274 shape variation along PC1 might also involve the relative width of the crown base, as
275 compared to the main cusp height (Figure 4).
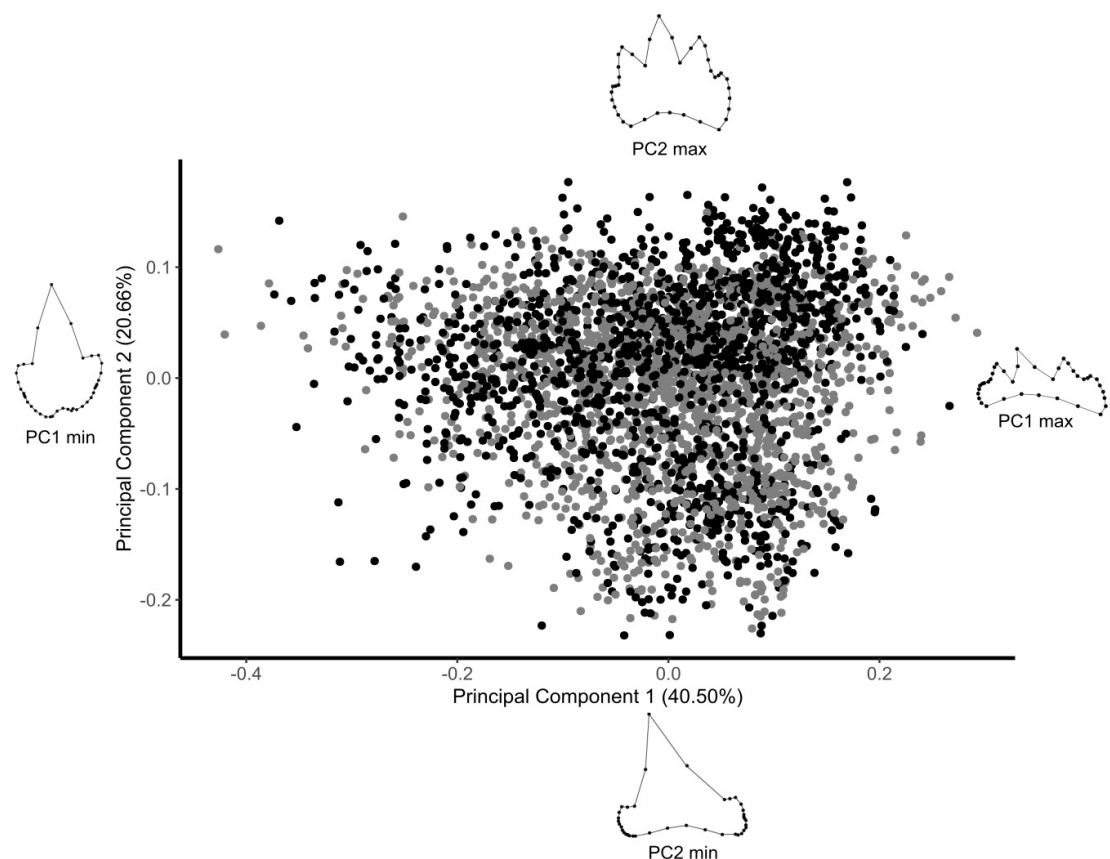
Figure 4. PCA with all *Scyliorhinus canicula* teeth contained in the dataset. Black, teeth from Mediterranean specimens; grey, teeth from Atlantic specimens. Wireframes are the extreme shapes for PC1 and PC2.

276 The tooth centroid size patterns in the *S. canicula* dataset are exemplified for male
277 palatoquadrate teeth (Figure 5), but similar trends are observed for both jaws and
278 sexes. In hatchlings, the tooth centroid size patterns display no variation along the jaw
279 nor visual differences between populations (Figure 5). The patterns overlap between
280 juveniles of both populations, while tooth centroid sizes of mature Atlantic males are
281 35% bigger than those of mature Mediterranean males (Figure 5). In mature specimens
282 of both populations, the tooth centroid sizes are overall higher for the Atlantic population,
283 whose body length is also higher as compared to Mediterranean specimens (Figure 1 and

Figure 5). However, no such trend is observed for juveniles, whose differences in body
length are not reflected by their tooth centroid size patterns (Figure 1 and Figure 5).
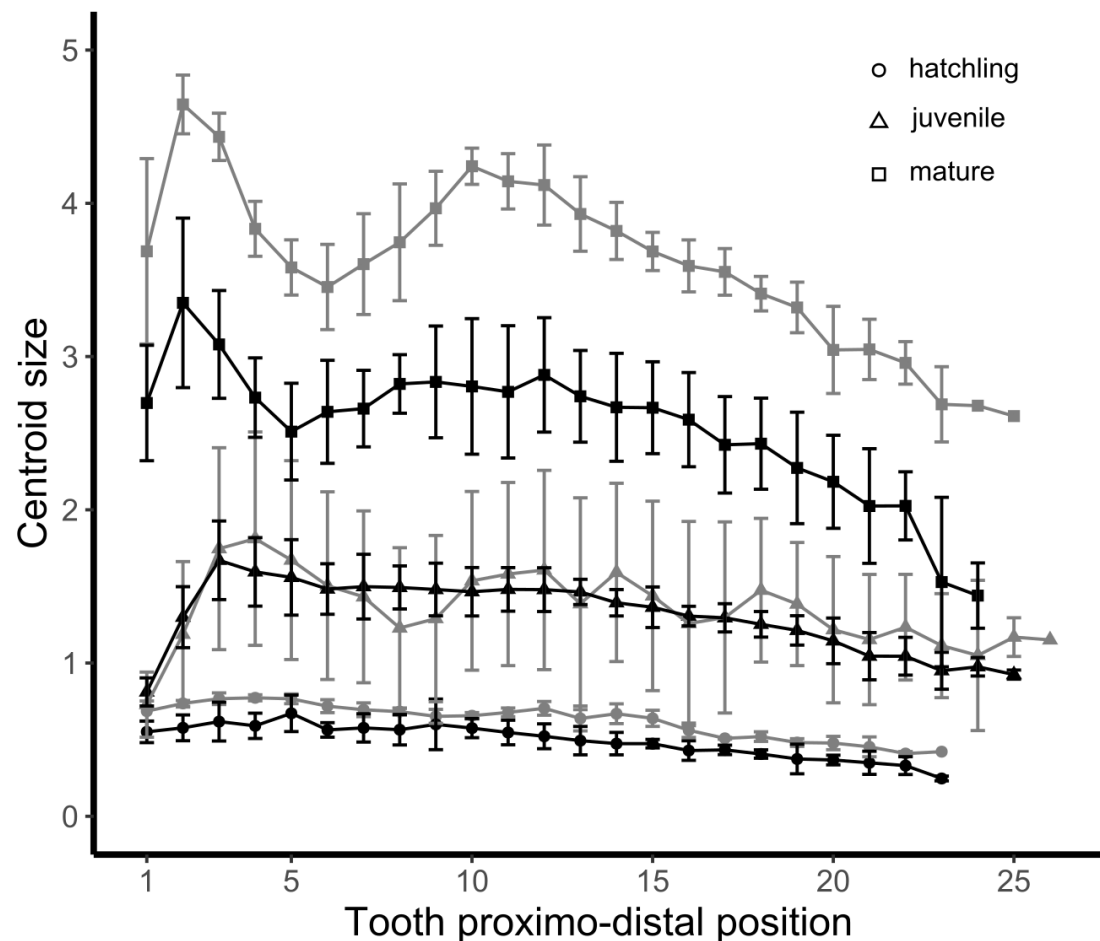
Figure 5. Tooth centroid size patterns of North Atlantic and Mediterranean
populations of *Scyliorhinus canicula*: Example of male palatoquadrate teeth. Black,
teeth from Mediterranean specimens; grey, teeth from Atlantic specimens. The centroid
size is in mm.

The first 12 PCs (93.43% of the total variation) are selected to represent the tooth
shape in the following statistical tests and are used as LDA features. The tooth centroid
size and the TL of specimens significantly impact the tooth shape in all subgroups
(e.g., Mediterranean x females) (One-Way MANOVAs, p-vals< $2.20e^{-16}$), meaning
for example that bigger teeth in larger specimens constrain the global shape of their
teeth. The slopes of allometric patterns significantly differ between females and between
males of both populations, indicating that the relationship between tooth size and shape
is not equivalent between populations (Two-Way ANCOVAs, p-vals< $2.20e^{-16}$). The
strength of relationship between shape and tooth centroid size is similar between females
from Atlantic and Mediterranean populations (linear regressions, adjusted R-squared
respectively of 0.83 and 0.81, p-vals< $2.20e^{-16}$), as well as between Atlantic and
Mediterranean males (linear regressions, adjusted R-squared respectively of 0.86 and

298 0.84, p-vals$< 2.20e^{-16}$). The allometric relationship between shape and TL is also
299 similar between Atlantic and Mediterranean populations (linear regressions, adjusted
300 R-squared respectively of 0.73 and 0.69 in females and of 0.82 and 0.81 in males,
301 p-vals$< 2.20e^{-16}$), which indicates that the tooth shapes are similarly modified over the
302 ontogeny in both populations.

Supervised classification

General performances

305 The LDA classification reached an accuracy of $64.5\% \pm 0.7\%$ with tooth shapes and
306 $74.6\% \pm 1.2\%$ with tooth forms. The precision is similar for both populations, whereas
307 the recall is much higher for the Atlantic population as compared to the Mediterranean
308 one (Figure 6). Better performances are also achieved on tooth forms as compared to
309 tooth shapes (Figure 6).

310 The classification task performed by the Random Forests reaches an accuracy of
311 $81.7\% \pm 1.7\%$ with tooth shapes and $86.9\% \pm 1.4\%$ with tooth forms, indicating a
312 significant contribution of the centroid size information at improving the discrimination
313 between populations. This means our protocol allows differentiating Mediterranean
314 from Atlantic *S. canicula* tooth forms with 85.5% to 88.3% accuracy. The precision
315 values are similar between both populations, but the recall values are better for the
316 Atlantic population than for the Mediterranean one (Figure 6).
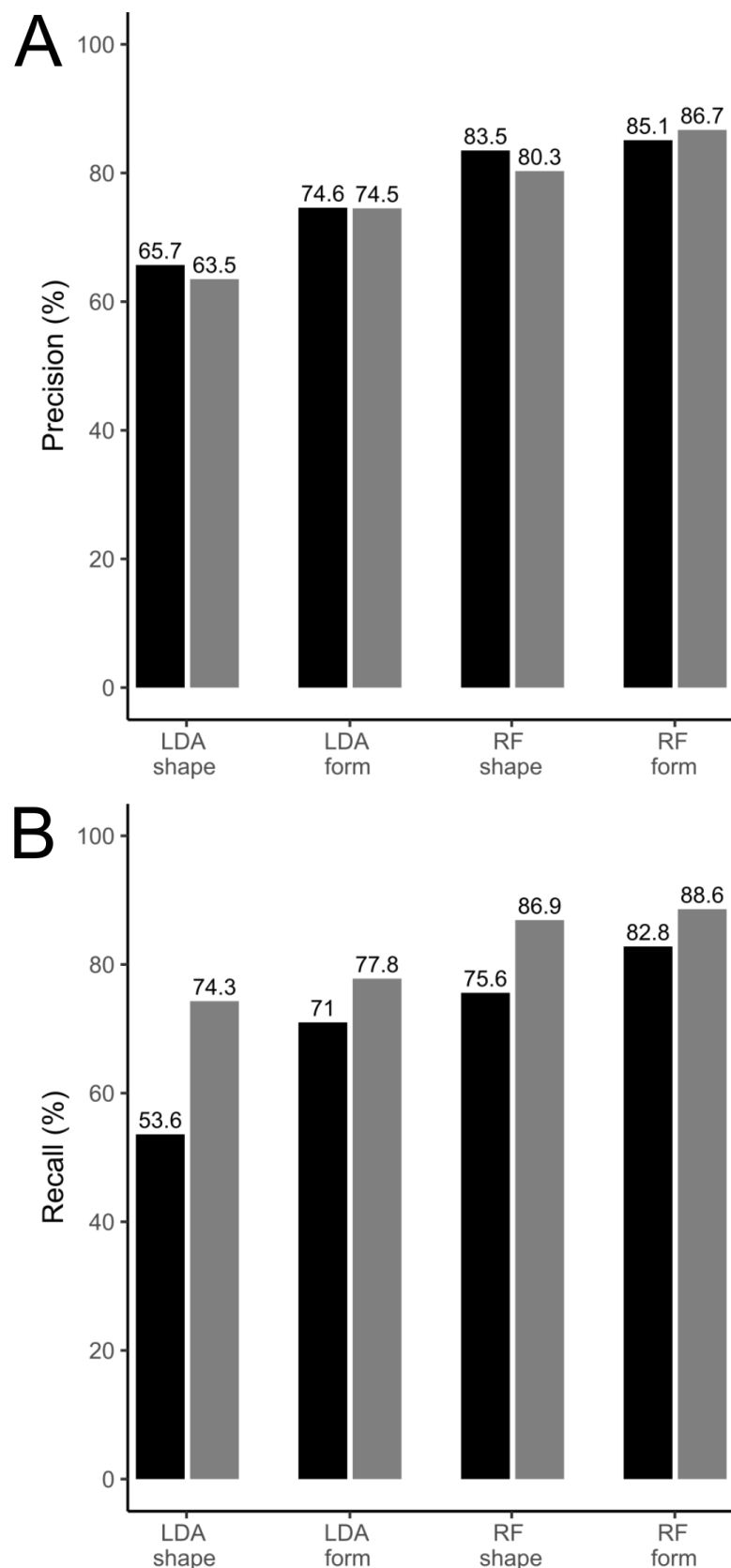
Figure 6. Classification performances of LDA and RF algorithms on *Scyliorhinus canicula* teeth from an Atlantic and a Mediterranean population. (A) Precision values; (B) Recall values. Black, teeth from Mediterranean specimens; grey, teeth from Atlantic specimens. LDA, Linear Discriminant Analysis; RF, Random Forests.

Subclass results

To get further indications of the classification performances, we detail the confusion matrix for each sex-stage subclass in the dataset (Table 1). We remind, however, that these values are still computed based on the population class (Atlantic and Mediterranean) only and that the performances for subclasses (e.g., Atlantic female hatchling) are detailed after the classification process.

Table 1. LDA and Random Forest performances of tooth shape and form classification from Atlantic and Mediterranean *Scyliorhinus canicula* populations. Atl., Atlantic; Med., Mediterranean.

	LDA				Random Forests			
	Shape		Form		Shape		Form	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Atl. *(female hatchling)*	57.1%	63.8%	63.5%	68.1%	64.4%	80.6%	82.8%	80.0%
Atl. *(female juvenile)*	74.5%	75.0%	88.5%	69.4%	81.4%	81.4%	88.5%	83.1%
Atl. *(female mature)*	73.6%	61.3%	78.8%	84.0%	81.5%	84.3%	85.7%	84.6%
Atl. *(male hatchling)*	49.3%	59.6%	60.7%	59.6%	70.4%	79.2%	83.3%	83.3%
Atl. *(male juvenile)*	78.4%	79.0%	86.1%	73.4%	94.3%	97.1%	91.9%	89.5%
Atl. *(male mature)*	50.9%	92.2%	65.2%	94.8%	77.9%	92.3%	84.9%	100%
Med. *(female hatchling)*	41.9%	35.3%	52.2%	47.1%	69.6%	50.0%	66.7%	70.6%
Med. *(female juvenile)*	65.7%	65.1%	67.9%	87.7%	76.8%	76.8%	79.2%	85.7%
Med. *(female mature)*	72.8%	82.4%	86.5%	81.9%	83.1%	80.2%	84.6%	85.7%
Med. *(male hatchling)*	45.2%	35.2%	58.2%	59.3%	80.0%	71.4%	72.2%	72.2%
Med. *(male juvenile)*	71.7%	71.0%	70.3%	84.1%	95.7%	91.7%	87.7%	90.5%
Med. *(male mature)*	57.1%	10.5%	90.4%	49.0%	88.9%	70.2%	100%	75.8%

With LDA, the lesser precision values are achieved for hatchlings and mature males with shape data and for hatchlings only with form data (Table 1). The less complex dental morphologies of the dataset are also visually identified in these groups that display one or three tooth cusps. Recall values are low for Mediterranean hatchlings and mature males for which the recall reaches 10.5% with shape data (Table 1). This means that amongst all Mediterranean mature males only 10.5% are detected as such by the model.

The Random Forests models on shape and form data achieve better classification performances for juvenile and mature specimens than for hatchling ones (Table 1). The model with form data also reaches 100% precision for Mediterranean mature males and 100% recall for Atlantic mature males (Table 1).

In most subclasses, better performances are obtained with form data as compared with shape data (Table 1). In some cases, however, the classification of form data confuses the models, which reach identical or lesser performance values than with shape data. The tooth centroid size for example does not improve the recall of Atlantic hatchling males after a LDA (Table 1). Lesser performances with form rather than shape data are also reported with LDA performed on teeth of Atlantic juvenile males and females (recall values) and with Random Forests performed on teeth of Atlantic and Mediterranean juvenile males (precision and recall) (Table 1).

With shape data, the most important feature lies in semilandmark 11 (4.5% in x), followed by semilandmark 24 (1.5% in x and 1.4 in y) and landmarks 1, 13, and 19 (1.5% in y, y, and x respectively) (Supp. Mat. 1). Semilandmarks 12, 22, and 23 also account for more than 1.0% in accuracy (1.0% in x and 1.3% in y, 1.0% in y, and 1.1%

in y, respectively) (Supp. Mat. 1). With form data, however, the feature contributing the most to the classification is centroid size (12.3%) (Supp. Mat 2). The following most important features with form data are semilandmark 11 (5.7% in x and 1.6% in y), semilandmark 24 (2.1% in x), landmark 19 (1.1% in x and 1.8% in y), semilandmark 23 (1.7% in x), semilandmark 12 (1.2% in y), landmarks 1 and 13 (1.0% in y), and semilandmark 38 (1.0% in z) (Supp. Mat. 2).

DISCUSSION

Visual descriptions and GM struggle to discriminate between populations

This work highlights that inter-population tooth differences lie in the lateral cusps. We first visually examined *S. canicula* teeth from both populations and, except in hatchlings, the teeth in the Mediterranean population appear sharper than in the Atlantic at all locations and usually display more accessory cusps for equivalent mesio-distal positions along the jaw.

On centroid size patterns, very few elements discriminate between the two populations: hatchling and juvenile specimens of both populations display similar tooth centroid size patterns and amplitude along the jaw. The main difference between populations arises between juvenile and mature stages because mature Atlantic specimens have teeth whose centroid size is about 35% higher than those of mature Mediterranean *S. canicula*. The similarities between the tooth centroid size patterns of Atlantic and Mediterranean hatchlings are consistent with their very close estimated age and TL (Figure 1). In mature specimens also, the amplitude delta in tooth centroid size patterns can be easily interpreted in light of their TL and age differences (Figure 1). Yet, the age estimations we provide remain approximative because we chose among several studies on *S. canicula* growth parameters. The similar amplitude of tooth centroid size patterns between juveniles of both populations could also be a consequence of the very close ranges of size of our specimens (Figure 1). Regarding allometric patterns, differences in slopes were detected between populations, meaning that the difference of increase between tooth centroid size and body size is not the same between the Atlantic and Mediterranean populations considered.

In summary, the reported differences between the populations' tooth centroid size patterns seem to be related to the TL of mature specimens, but they do not discriminate between both populations among hatchlings and juveniles.

The contribution of Random Forests to decipher inter-population differences

Machine learning models have already improved the understanding of subtle structures in geometric morphometric data (Lorenz et al., 2015; Soda et al., 2017; Doyle et al., 2018; Courtenay et al., 2019; Quenu et al., 2020; Barone et al., 2021), especially when the shapes between groups share common quadrants in a morphospace, as it is the case for the *S. canicula* populations considered in this study. However, even though the algorithms perform well at classifying geometric morphometric data, the choices made by these models to make groups are usually unknown and deprived of biological meaning (Lorenz et al., 2015; Quenu et al., 2020).

In traditional geometric morphometrics, supervised classification is performed with

an LDA on several PC axes based on raw shape data. We aimed to compare the results from this traditional workflow with the classification performances obtained with Random Forests on raw shape data. Doyle et al. (2018) already compared the classification performances of LDA and Random Forests on the shells of populations of common periwinkles (*Littorina littorea*) distributed in different niches. They recommended the use of Random Forests over LDA because the former is more straightforward and robust, does not make assumptions about the data nor necessitate the user to check the violation of LDA assumptions (Doyle et al., 2018). However, the models of Doyle et al. (2018) do overfit and the results obtained might not be optimal.

Overall, LDA models achieve lower performances than Random Forests with our dataset, which can be due to the information reduction. We used LDA for classification as in the majority of geometric morphometric articles, e.g., without considering the data distribution nor the homogeneity of variance. That such criteria for optimality are not met does not prevent from performing an LDA. The LDA algorithm is robust to such violations and still achieves good performances when assumptions of normality and common covariance matrix among groups are not met (Lachenbruch and Goldstein, 1979; Li et al., 2006). Furthermore, Doyle et al. (2018) showed similar performances between LDA and Random Forests on geometric morphometric data and conclude that LDA is robust enough to the abovementioned violated conditions to achieve good classification performances.

The contribution of Random Forest models to our work is four-fold. It requires less preprocessing steps than LDA, performs better than LDA at the classification task, achieves good performances at classifying teeth from two *S. canicula* populations based on raw data, and determines the most discriminant features to perform this task. Among landmark and semilandmark features, it is clear that the most discriminant information is contained in two dimensions (x and y), whereas the third dimension brings less information. This implies that running the models with 2D landmark data would have probably achieved similar classification performances as those obtained here, as this has also been evidenced in some traditional geometric morphometric studies (Cardini, 2014; Buser et al., 2018; Wasiljew et al., 2020). Nevertheless, 3D landmarking avoids parallax biases (Cardini, 2014) and none of the landmarks and semilandmarks caused a decrease of the metrics, indicating that even though the z dimension contains little information, it still cannot be considered as noise. Furthermore, the landmarks (1, 13, 19) and semilandmarks (11, 12, 23, and 24) contributing the most to the classification are located at geometrical extrema, at extreme mesial and distal locations of the teeth and on the lateral edges, where accessory cusps emerge. It is also likely that the spacing of these points makes them useful to the algorithm and that spatially close points would bring less information. Overall, the results show that a few points, especially on the lateral sides of the teeth, provide enough information to represent most differences between the tooth shapes of the two *S. canicula* populations considered. This is consistent with our visual inspection on the variation of accessory cusps number.

The use of form features instead of shape greatly improves the overall classification. We expected such impact in the classification of mature specimens due to the amplitude differences of their centroid size patterns (likely caused by TL and putative age discrepancies, Figure 1). However, the reason for the significant contribution of tooth centroid size to the classification improvement of the teeth of juveniles and hatchlings

is less intuitive. We first assume that the very slight differences in tooth centroid size patterns we visually interpret as part of the inter-populational variability might actually be considered useful information for the model to discriminate between populations. The centroid size is theoretically independent of shape, however, the placement and density of landmarks and semilandmarks modify the contribution of certain parts of a tooth (e.g., the lateral sides of the crown as compared to the main cusp) to the centroid size value. Thus, we assume that slight changes in centroid size values between populations might also convey relevant shape information for the Random Forest model, which is cryptic to the observer. For a minority of subclasses, however, the addition of centroid size to shape features does not improve the classification performances or diminish them. In the first case, the centroid size is probably too similar between two subgroups (e.g., Atlantic and Mediterranean hatchling males) to allow the model to discriminate against the populations. In the second case, the centroid size information confuses the model, which might indicate that some form data contain more noise than in other groups, which could be overcome by increasing the sampling effort for the subclasses considered (e.g., Atlantic and Mediterranean juvenile males).

Ecological origins of anatomical divergence between populations

The inter- and intraspecific diversity of shark tooth shapes correlates with their feeding behaviour. Molariform teeth, for example, help crushing hard-bodied preys, while cutting teeth allow to remove pieces from larger items (Cappetta, 1986).

S. canicula is considered a generalist predator with opportunistic behaviour, whose favourite preys are teleosts, cephalopods, and crustaceans but it also occasionally feeds on macroalgae and echinoderms (Olaso, 1998; Mnasri et al., 2012; Kousteni et al., 2017, 2018). Within the same population, however, slight differences in diet composition can occur between specimens of different sex and ontogenetic stages (Lyle, 1983; Olaso, 1998; Rodríguez-Cabello et al., 2007; Mnasri et al., 2012; Kousteni et al., 2017, 2018). *S. canicula* also displays seasonal diet shifts that differ according to the geographic area: more cephalopods are eaten in the winter than in autumn by a Cantabrian Sea (North Atlantic Ocean) population, while a population from eastern Mediterranean Sea feeds most on teleosts in spring and on molluscs in autumn (Olaso, 1998; Kousteni et al., 2017). Molluscs are far less important in the diet composition of Atlantic populations than in Mediterranean ones, whereas Atlantic specimens feed more on teleosts (Kousteni et al., 2017). However, the diet differences between several Atlantic and Mediterranean *S. canicula* populations are probably due to the variability of available prey items in contrasting habitats, as suggested by the opportunistic behaviour of this species (Lyle, 1983; Kousteni et al., 2017). Additionally, we show in this work that *S. canicula* tooth shapes differ between one Atlantic and one Mediterreanean populations. If this variation is under functional selection, diet differences would correlate with distinct tooth forms: sharper teeth of the Mediterranean population may enhance grasping molluscs such as cephalopods, whereas teeth of Atlantic specimens might perform better at catching benthic teleosts. However, the specific diet of the specimens sampled in this study is not known and the relationship between the tooth morphologies depicted here and broad dietary trends in all Atlantic and Mediterranean *S. canicula* remains hypothetical. Overall, several studies already suggested that the diet composition of *S. canicula*, as well as of other elasmobranchs, is correlated to the body size (Lyle, 1983; Bethea et al.,

2006; Ellis and Musick, 2007; Borrell et al., 2011; Šantić, M. and Rađa, B. and Pallaoro, A., 2012; Kousteni et al., 2017). Furthermore, our results support the correlation between tooth shape and ontogeny, and thus body length in *S. canicula*, which could suggest that there is an association between diet shifts and tooth shape modifications over a specimen lifetime.

S. canicula has a philopatric behaviour and most specimens do not move further than 30 km over the years (Rodríguez-Cabello et al., 2004). Additionnally, shared haplotypes between specimens caught off Portugal and specimens from the Mediterranean Sea demonstrate past communication occurrences between the populations, probably resulting from colonisation events from the Atlantic or retention of ancestral polymorphism (Kousteni et al., 2015; Ramírez-Amaro et al., 2018). There is evidence of multiple genetic stocks of *S. canicula* within the Mediterranean Sea (Barbieri et al., 2014; Kousteni et al., 2015). Yet, molecular studies suggest that the Siculo-Tunisian Strait may allow gene flow between the eastern and western Mediterranean populations, whereas genetic exchanges are currently very low between the Atlantic and Mediterranean populations of *S. canicula* (Barbieri et al., 2014; Kousteni et al., 2015). The communication between these Atlantic and Mediterranean populations would indeed only be permitted through the Strait of Gibraltar, yet bottom topography may limit the migration between populations to a few specimens (Ramírez-Amaro et al., 2018). The differences in tooth morphology between the Atlantic and Mediterranean population samples of *S. canicula* is probably related to the species genetic differentiation recorded between these locations (Barbieri et al., 2014). This assumption could be further tested by examining the differentiation at both molecular and tooth morphology level between these stocks.

CONCLUSIONS

We combined geometric morphometrics with a machine learning approach to discriminate between teeth of population samples of *S. canicula* from the northeastern Atlantic Ocean and the Mediterranean Sea. The traditional framework used in geometric morphometrics reached lesser performances at distinguishing the tooth shapes of *S. canicula* specimens from the two populations. Nevertheless, these shape data combined with centroid sizes allow a Random Forest model to classify the *S. canicula* teeth with up to 100% precision. This framework should be further tested in more *S. canicula* populations to decipher tooth morphological differences between spatially and genetically close-related populations. We hope this emergent framework to be further tuned by ichthyologists by including geographical parameters and life history traits to discriminate between subtle tooth morphologies and to provide a basis for facilitating the identification procedures of fish stocks and improving fisheries management.

ACKNOWLEDGMENTS

We acknowledge the MRI platform member of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04, "Investments for the future"), the labex CEMEB (ANR-10-LABX-0004) and NUMEV (ANR-10-LABX-0020) and thank Renaud Lebrun for his help with microCT imaging. We acknowledge the contribution of SFR Biosciences (UMS3444/CNRS, US8/Inserm, ENS de Lyon, UCBL) AniRa-ImmOs facility and thank Mathilde Bouchet-Combe for

her help with microCT imaging. We thank Sylvie Agret for her help generating 3D surfaces.

REFERENCES

Adams, D., Collyer, M., and Kaliontzopoulou, A. (2019). Geomorph: Software for geometric morphometric analyses. R package version 3.1.0.

Almerón-Souza, F., Sperb, C., Castilho, C. L., Figueiredo, P. I. C. C., Gonçalves, L. T., Machado, R., Oliveira, L. R., Valiati, V. H., and Fagundes, N. J. R. (2018). Molecular Identification of Shark Meat From Local Markets in Southern Brazil Based on DNA Barcoding: Evidence for Mislabeling and Trade of Endangered Species. *Frontiers in Genetics*, 9:138.

Anastasopoulou, A., Biandolino, F., Chatzispyrou, A., Hemida, F., Guijarro, B., Kousteni, V., Mytilineou, C., Pattoura, P., and Prato, E. (2016). New Fisheries-related data from the Mediterranean Sea (November, 2016). *Mediterranean Marine Science*, 17(3):822–827.

Archer, K. J. and Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics and Data Analysis*, 52(4):2249–2260.

Barbieri, M., Maltagliati, F., Roldán, M., and Castelli, A. (2014). Molecular contribution to stock identification in the small-spotted catshark, *Scyliorhinus canicula* (Chondrichthyes, Scyliorhinidae). *Fisheries Research*, 154:11–16.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., and Casiraghi, M. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo" (*Mustelus* spp.). *Food Research International*, 43(1):376–381.

Barone, M., Mollen, F. H., Giles, J. L., Marshall, L. J., Villate-Moreno, M., Mazzoldi, C., Pérez-Costas, E., Heine, J., and Guisande, C. (2021). Performance of *iSharkFin* in the identification of wet dorsal fins from priority shark species. *Ecological Informatics*, page 101514.

Bendiab, A. A. T., Mouffok, S., and Boutiba, Z. (2012). Reproductive biology and growth of lesser spotted dogfish *Scyliorhinus canicula* (Linnaeus, 1758) in western Algerian coasts (Chondrichthyes, Scyliorhinidae). *Biodiversity Journal*, 3(1):41–48.

Berio, F. and Bayle, Y. (2020). Scyland3D: Processing 3D landmarks. *Journal of Open Source Software*, 5(46):1262.

Berio, F., Bayle, Y., Agret, S., Baum, D., Goudemand, N., and Debiais-Thibaud, M. (2022). 3D models related to the publication: Hide and seek shark teeth in Random Forests: Machine learning applied to *Scyliorhinus canicula*. *MorphoMuseuM*.

Berio, F., Evin, A., Goudemand, N., and Debiais-Thibaud, M. (2020). The intraspecific diversity of tooth morphology in the large-spotted catshark *Scyliorhinus stellaris*: insights into the ontogenetic cues driving sexual dimorphism. *Journal of Anatomy*, 237(5):960–978.

Bethea, D. M., Carlson, J. K., Buckel, J. A., and Satterwhite, M. (2006). Ontogenetic and site-related trends in the diet of the Atlantic sharpnose shark *Rhizoprionodon terraenovae* from the northeast Gulf of Mexico. *Bulletin of Marine Science*, 78(2):287–307.

Blackburn, T. M., Gaston, K. J., and Loder, N. (2008). Geographic gradients in body size: a clarification of Bergmann's rule. *Diversity and distributions*, 5(4):165–174.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge, 1 edition.

Borrell, A., Aguilar, A., Gazo, M., Kumarran, R. P., and Cardona, L. (2011). Stable isotope profiles in whale shark (*Rhincodon typus*) suggest segregation and dissimilarities in the diet depending on sex and size. *Environmental Biology of Fishes*, 92(4):559–567.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Buser, T. J., Sidlauskas, B. L., and Summers, A. P. (2018). 2D or Not 2D? Testing the Utility of 2D Vs. 3D Landmark Data in Geometric Morphometrics of the Sculpin Subfamily Oligocottinae (Pisces; Cottoidea). *The Anatomical Record*, 301(5):806–818.

Capapé, C., Mnasri-Sioudi, N., El Kamel-Moutalibi, O., Boumaïza, M., Amor, M. B., and Reynaud, C. (2014). Production, maturity, reproductive cycle and fecundity of small-spotted catshark, *Scyliorhinus canicula* (Chondrichthyes: Scyliorhinidae) from the northern coast of Tunisia (Central Mediterranean). *Journal of Ichthyology*, 54(1):111–126.

Cappetta, H. (1986). Types dentaires adaptatifs chez les sélaciens actuels et post-paléozoïques. *Palaeovertebrata*, 16(2):57–76.

Cardini, A. L. (2014). Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures? *Hystrix*, 25:73–81.

Carlson, J. K. and Parsons, G. R. (1997). Age and growth of the bonnethead shark, *Sphyrna tiburo*, from northwest Florida, with comments on clinal variation. *Environmental Biology of Fishes*, 50(3):331–341.

Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168.

Compagno, L. J. V. (1984). FAO species catalogue: vol 4 Sharks of the world an annotated and illustrated catalogue of shark species known to date Part 2-Carcharhiniformes. *FAO Fisheries Synopsis*, 4:251–265.

Cortés, E. (2000). Life History Patterns and Correlations in Sharks. *Reviews in Fisheries Science*, 8(4):299–344.

Courtenay, L. A., Yravedra, J., Huguet, R., Aramendi, J., Maté-González, M. Á., González-Aguilera, D., and Arriaza, M. C. (2019). Combining machine learning algorithms and geometric morphometrics: A study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 522:28–39.

Cullen, J. A. and Marshall, C. D. (2019). Do sharks exhibit heterodonty by tooth position and over ontogeny? A comparison using elliptic Fourier analysis. *Journal of Morphology*, 280(5):687–700.

Debiais-Thibaud, M., Chiori, R., Enault, S., Oulion, S., Germon, I., Martinand-Mari, C., Casane, D., and Borday-Birraux, V. (2015). Tooth and scale morphogenesis in shark: an alternative process to the mammalian enamel knot system. *BMC Evolutionary Biology*, 15(1):292.

Díaz-Uriarte, R. and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1):1–13.

Domínguez-Rodrigo, M. and Baquedano, E. (2018). Distinguishing butchery cut marks from crocodile bite marks through machine learning methods. *Scientific Reports*, 8(1):1–8.

Doyle, D., Gammell, M. P., and Nash, R. (2018). Morphometric methods for the analysis and classification of gastropods: A comparison using *Littorina littorea*. *Journal of Molluscan Studies*, 84(2):190–197.

Dulvy, N. K., Fowler, S. L., Musick, J. A., Cavanagh, R. D., Kyne, P. M., Harrison, L. R., Carlson, J. K., Davidson, L. N. K., Fordham, S. V., Francis, M. P., Pollock, C. M., Simpfendorfer, C. A., Burgess, G. H., Carpenter, K. E., Compagno, L. J. V., Ebert, D. A., Gibson, C., Heupel, M. R., Livingstone, S. R., Sanciangco, J. C., Stevens, J. D., Valenti, S., and White, W. T. (2014). Extinction risk and conservation of the world's sharks and rays. *eLife*, 3:e00590.

Dulvy, N. K., Pacoureau, N., Rigby, C. L., Pollom, R. A., Jabado, R. W., Ebert, D. A., Finucci, B., Pollock, C. M., Cheok, J., Derrick, D. H., Herman, K. B., Sherman, C. S., VanderWright, W. J., Lawson, J. M., Walls, R. H. L., Carlson, J. K., Charvet, P., Bineesh, K. K., Fernando, D., Ralph, G. M., Matsushiba, J. H., Hilton-Taylor, C., Fordham, S. V., and Simpfendorfer, C. A. (2021). Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. *Current Biology*, 31(22):5118–5119.

Ellis, J. K. and Musick, J. A. (2007). Ontogenetic changes in the diet of the sandbar shark, *Carcharhinus plumbeus*, in lower Chesapeake Bay and Virginia (USA) coastal waters. *Environmental Biology of Fishes*, 80:51–67.

Ellis, J. R. and Shackley, S. E. (1997). The reproductive biology of *Scyliorhinus canicula* in the Bristol Channel, U.K. *Journal of Fish Biology*, 51(2):361–372.

Evin, A., Cucchi, T., Cardini, A., Vidarsdottir, U. S., Larson, G., and Dobney, K. (2013). The long and winding road: identifying pig domestication through molar size and shape. *Journal of Archaeological Science*, 40:735–743.

Fort, G. and Lambert-Lacroix, S. (2005). Classification using partial least squares with penalized logistic regression. *Bioinformatics*, 21(7):1104–1111.

Goren, M. (2014). The Fishes of the Mediterranean: A Biota Under Siege. In *The Mediterranean Sea*, pages 385–400. Springer Netherlands, Dordrecht.

Gruber, S. H. and Compagno, L. J. V. (1981). Taxonomic status and biology of the bigeye thresher *Alopias superciliosus* (Mediterranean Sea, New Zealand). *Fishery Bulletin*, 79(4):617–640.

Horie, T. and Tanaka, S. (2002). Geographic variation of maturity size of the cloudy catshark, *Scyliorhinus torazame*, in Japan. *Journal of the Faculty of Marine Science and Technology-Tokai University*.

Ivory, P., Jeal, F., Nolan, C. P., and Jeal, P. F. (2004). Age determination, growth and reproduction in the lesser-spotted dogfish, *Scyliorhinus canicula* (L.). *Journal of Northwest Atlantic Fishery Science*, 35:89–106.

Karl, S. A., Castro, A. L. F., and Garla, R. C. (2012). Population genetics of the nurse shark (*Ginglymostoma cirratum*) in the western Atlantic. *Marine Biology*, 159(3):489–498.

Klingenberg, C. P. (2016). Size, shape, and form: concepts of allometry in geometric morphometrics. *Development genes and evolution*, 226(3):113–137.

Kousteni, V., Karachle, P. K., and Megalofonou, P. (2017). Diet of the small-spotted

catshark *Scyliorhinus canicula* in the Aegean Sea (eastern Mediterranean). *Marine Biology Research*, 13(2):161–173.

Kousteni, V., Karachle, P. K., Megalofonou, P., and Lefkaditou, E. (2018). Cephalopod prey of two demersal sharks caught in the Aegean Sea (eastern Mediterranean). *Journal of the Marine Biological Association of the United Kingdom*, 98(1):81–88.

Kousteni, V., Kasapidis, P., Kotoulas, G., and Megalofonou, P. (2015). Strong population genetic structure and contrasting demographic histories for the small-spotted catshark (*Scyliorhinus canicula*) in the Mediterranean Sea. *Heredity*, 114(3):333–343.

Kousteni, V., Kontopoulou, M., and Megalofonou, P. (2010). Sexual maturity and fecundity of *Scyliorhinus canicula* (Linnaeus, 1758) in the Aegean Sea. *Marine Biology Research*, 6(4):390–398.

Kousteni, V. and Megalofonou, P. (2019). Reproductive strategy of *Scyliorhinus canicula* (L., 1758): a holistic approach based on macroscopic measurements and microscopic observations of the reproductive organs. *Marine and Freshwater Research*, 71(6):596–616.

Lachenbruch, P. A. and Goldstein, M. (1979). Discriminant analysis. *Biometrics*, 35(1):85.

Leloup, J. and Olivereau, M. (1951). Données biométriques comparatives sur la Roussette (*Scyllium canicula* L.) de la Manche et de la Méditerranée. *Vie et milieu*, 2(2):182–206.

Li, T., Zhu, S., and Ogihara, M. (2006). Using discriminant analysis for multi-class classification: An experimental investigation. *Knowledge and Information Systems*, 10(4):453–472.

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3):18–22.

Lombardi-Carlson, L. A., Cortés, E., Parsons, G. R., and Manire, C. A. (2003). Latitudinal variation in life-history traits of bonnethead sharks, *Sphyrna tiburo*, (Carcharhiniformes: Sphyrnidae) from the eastern Gulf of Mexico. *Marine and Freshwater Research*, 54(7):875–883.

Lorenz, C., Ferraudo, A. S., and Suesdek, L. (2015). Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Tropica*, 152:165–169.

Lucifora, L. O., Cione, A. L., Menni, R. C., and Escalante, A. H. (2003). Tooth row counts, vicariance, and the distribution of the sand tiger shark *Carcharias taurus*. *Ecography*, 26(5):567–572.

Lyle, J. M. (1983). Food and feeding habits of the lesser spotted dogfish, *Scyliorhinus canicula* (L.), in Isle of Man waters. *Journal of Fish Biology*, 23(6):725–737.

MacLeod, N. (2017). On the use of machine learning in morphometric analysis. In *Biological shape analysis: proceedings of the 4th international symposium*, pages 134–171.

MacLeod, N. (2018). The quantitative assessment of archaeological artifact groups: Beyond geometric morphometrics. *Quaternary Science Reviews*, 201:319–348.

McEachran, J. D. and Martin, C. O. (1977). Possible occurrence of character displacement in the sympatric skates *Raja erinacea* and *R. ocellata* (Pisces: Rajidae). *Environmental Biology of Fishes*, 2(2):121–130.

Mellinger, J., Wrisez, F., and Alluchon-Gérard, M.-J. (1984). Caractères biométriques distinctifs de l'embryon et de ses annexes chez la roussette (*Scyliorhinus canicula*) de

la Manche, comparée à celle de la Méditerranée, et détermination précise du stade d'éclosion. *Cahiers de biologie marine*, 25(3):305–317.

Melo Palmeira, C. A., da Silva Rodrigues-Filho, L. F., de Luna Sales, J. B., Vallinoto, M., Schneider, H., and Sampaio, I. (2013). Commercialization of a critically endangered species (largetooth sawfish, *Pristis perotteti*) in fish markets of northern Brazil: Authenticity by DNA analysis. *Food Control*, 34(1):249–252.

Mitteroecker, P. and Bookstein, F. (2011). Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, 38:100–114.

Mnasri, N., El-Kamel, O., Boumaïza, M., Reynaud, C., and Capapé, C. (2012). Food and feeding habits of the small-spotted catshark, *Scyliorhinus canicula* (Chondrichthyes: Scyliorhinidae) from the northern coast of Tunisia (central Mediterranean). *Cahiers de Biologie Marine*, 53(1):139–150.

Moyer, J. K. and Bemis, W. E. (2016). Tooth Microstructure and Replacement in the Gulper Shark, *Centrophorus granulosus* (Squaliformes: Centrophoridae). *Copeia*, 104(2):529–538.

Natanson, L. J., Skomal, G. B., Hoffmann, S. L., Porter, M. E., Goldman, K. J., and Serra, D. (2018). Age and growth of sharks: do vertebral band pairs record age? *Marine and Freshwater Research*, 69(9):1440–1452.

Navega, D., Vicente, R., Vieira, D. N., Ross, A. H., and Cunha, E. (2015). Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. *International Journal of Legal Medicine*, 129(3):651–659.

Olaso, I. (1998). Importance of discarded blue whiting (*Micromesistius poutassou*) in the diet of lesser spotted dogfish (*Scyliorhinus canicula*) in the Cantabrian Sea. *ICES Journal of Marine Science*, 55(3):331–341.

Parsons, G. R. (1993). Geographic variation in reproduction between two populations of the bonnethead shark, *Sphyrna tiburo*. In Demski, L. S. and Wourms, J. P., editors, *The reproduction and development of sharks, skates, rays and ratfishes*, chapter 3, pages 25–35. Springer.

Pechenizkiy, M., Puuronen, S., and Tsymbal, A. (2006). The impact of sample reduction on PCA-based feature extraction for supervised learning. In *Proceedings of the ACM Symposium on Applied Computing*, volume 1, pages 553–558, New York, New York, USA. ACM Press.

Portnoy, D. S., McDowell, J. R., Heist, E. J., Musick, J. A., and Graves, J. E. (2010). World phylogeography and male-mediated gene flow in the sandbar shark, *Carcharhinus plumbeus*. *Molecular Ecology*, 19(10):1994–2010.

Powter, D. M., Gladstone, W., and Platell, M. (2010). The influence of sex and maturity on the diet, mouth morphology and dentition of the Port Jackson shark, *Heterodontus portusjacksoni*. *Marine and Freshwater Research*, 61(1):74.

Püschel, T. A., Marcé-Nogué, J., Gladman, J. T., Bobe, R., and Sellers, W. I. (2018). Inferring locomotor behaviours in Miocene New World monkeys using finite element analysis, geometric morphometrics and machine-learning classification techniques applied to talar morphology. *Journal of the Royal Society Interface*, 15(146).

Quenu, M., Trewick, S. A., Brescia, F., and Morgan-Richards, M. (2020). Geometric morphometrics and machine learning challenge currently accepted species limits of the land snail *Placostylus* (Pulmonata: Bothriembryontidae) on the Isle of Pines, New

Caledonia. *Journal of Molluscan Studies*.

Ramírez-Amaro, S., Picornell, A., Arenas, M., Castro, J. A., Massutí, E., Ramon, M. M., and Terrasa, B. (2018). Contrasting evolutionary patterns in populations of demersal sharks throughout the western Mediterranean. *Marine Biology*, 165(1):1–16.

Rodríguez-Cabello, C., Cabello, F., Olaso, I., and Olaso, I. (1998). Reproductive biology of lesser spotted dogfish *Scyliorhinus canicula* (L., 1758) in the Cantabrian Sea. *Scientia Marina*, 62(3):187–191.

Rodríguez-Cabello, C., Sánchez, F., Fernández, A., and Olaso, I. (2004). Is the lesser spotted dogfish (*Scyliorhinus canicula*) from the Cantabrian Sea, a unique stock? *Fisheries research*, 69(1):57–71.

Rodríguez-Cabello, C., Sánchez, F., and Olaso, I. (2007). Distribution patterns and sexual segregations of *Scyliorhinus canicula* (L.) in the Cantabrian Sea. *Journal of Fish Biology*, 70(5):1568–1586.

Šantić, M. and Rađa, B. and Pallaoro, A. (2012). Feeding habits of small-spotted catshark (*Scyliorhinus canicula* Linnaeus, 1758) from the eastern central Adriatic Sea. *Marine Biology Research*, 8(10):1003–1011.

Santos, F., Guyomarc'h, P., and Bruzek, J. (2014). Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic Science International*, 245:204.e1–204.e8.

Schwamborn, R., Mildenberger, T. K., and Taylor, M. H. (2019). Assessing sources of uncertainty in length-based estimates of body growth in populations of fishes and macroinvertebrates with bootstrapped ELEFAN. *Ecological Modelling*, 393:37–51.

Sheets, H. D., Covino, K. M., Panasiewicz, J. M., and Morris, S. R. (2006). Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. *Frontiers in Zoology*, 3(1):15.

Smith, S. E., Au, D. W., and Show, C. (1998). Intrinsic rebound potentials of 26 species of Pacific sharks. *Marine and Freshwater Research*, 49(7):663–678.

Soda, K., Slice, D., and Naylor, G. (2017). Artificial neural networks and geometric morphometric methods as a means for classification: A case-study using teeth from *Carcharhinus sp*. (Carcharhinidae). *Journal of Morphology*, 278(1):131–141.

Stalling, D., Westerhoff, M., and Hege, H.-C. (2005). Amira: a highly interactive system for visual data analysis. In *The Visualization Handbook*, pages 749–767. Elsevier.

Taniuchi, T. and Tachikawa, H. (1997). Geographical variation in age and growth of *Squalus mitsukurii* (Elasmobranchii: Squalidae) in north Pacific. In *Proceedings of 5th Indo-Pacific Fish Conference, Nouméa*, pages 321–328.

Templeman, W. (1984). Variations in numbers of median dorsal thorns and rows of teeth in thorny skate (*Raja radiata*) of the Northwest Atlantic. *Journal of Northwest Atlantic Fishery Science*, 5(2):171–179.

Tomita, T., Miyamoto, K., Kawaguchi, A., Toda, M., Oka, S.-I., Nozu, R., and Sato, K. (2017). Dental ontogeny of a white shark embryo. *Journal of Morphology*, 278(2):215–227.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer New York, fourth edition.

Veríssimo, A., McDowell, J. R., and Graves, J. E. (2010). Global population structure of the spiny dogfish *Squalus acanthias*, a temperate shark with an antitropical distribution. *Molecular Ecology*, 19(8):1651–1662.

Wasiljew, B. D., Pfaender, J., Wipfler, B., Utama, I. V., and Herder, F. (2020). Do we need the third dimension? Quantifying the effect of the *z*-axis in 3D geometric morphometrics based on sailfin silversides (Telmatherinidae). *Journal of Fish Biology*, 97(2):537–545.

Webster, M. and Sheets, H. D. (2010). A practical introduction to landmark-based geometric morphometrics. *The Paleontological Society Papers*, 16:163–188.

Whitenack, L. B. and Gottfried, M. D. (2010). A morphometric approach for addressing tooth-based species delimitation in fossil mako sharks, *Isurus* (Elasmobranchii: Lamniformes). *Journal of Vertebrate Paleontology*, 30(1):17–25.

Yamaguchi, A., Taniuchi, T., and Shimizu, M. (1998). Geographic variation in growth of the starspotted dogfish *Mustelus manazo* from five localities in Japan and Taiwan. *Fisheries Science*, 64(5):732–739.

Yamaguchi, A., Taniuchi, T., and Shimizu, M. (2000). Geographic Variations in Reproductive Parameters of the Starspotted Dogfish, *Mustelus manazo*, from Five Localities in Japan and in Taiwan. *Environmental Biology of Fishes*, 57(2):221–233.