3D models related to the publication: Hide and seek shark teeth in Random Forests: Machine learning applied to *Scyliorhinus canicula*

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

[1] *Institut des Sciences de l'Evolution de Montpellier, UMR5554, Université Montpellier, CNRS, IRD, EPHE, c.c.064, place Eugène Bataillon, 34095 Montpellier Cedex 05, France*

[2] *Univ. Lyon, École Normale Supérieure de Lyon, Centre National de la Recherche Scientifique, Institut de Génomique Fonctionnelle de Lyon, UMR 5242, 46 Allée d'Italie, Lyon Cedex 07, France, F-69364*

[3] *Université de Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR5800, Talence, France, F-33400*

[4] *Zuse Institute Berlin, Department of Visual and Data-Centric Computing, Takustr. 7, Berlin, Germany, 14195*

***Corresponding author**: fidji.berio@gmail.com

Abstract

The present dataset contains the 3D models analyzed in Berio, F., Bayle, Y., Baum, D., Goudemand, N., and Debiais-Thibaud, M. 2022. Hide and seek shark teeth in Random Forests: Machine learning applied to *Scyliorhinus canicula*. It contains the head surfaces of 56 North Atlantic and Mediterranean small-spotted catsharks *Scyliorhinus canicula*, from which tooth surfaces were further extracted to perform geometric morphometrics and machine learning.

Keywords: geometric morphometrics, machine learning, *Scyliorhinus canicula*, sharks, tooth morphology

Submitted:2022-01-09, published online:2022-05-24. https://doi.org/10.18563/journal.m3.164

INTRODUCTION

Marine populations often display body size differences at sexual maturity that are distributed along a latitudinal gradient (Dittman et al., 1998; Lombardi-Carlson et al., 2003; Huret et al., 2019). This trend has been reported in shark populations, with bigger specimens inhabiting higher and colder latitudes (Lombardi-Carlson et al., 2003; Capapé et al., 2014). Differences in shark body length can also correlate with number of tooth files or number of vertebrae. The small-spotted catshark *Scyliorhinus canicula* is considered a model species among sharks and North Atlantic populations reach greater body length at maturity than their Mediterranean counterparts. Scyliorhinid teeth undergo sex-specific modifications of tooth shape, as well as ontogenetic form changes (Soares and Carvalho, 2019; Berio et al., 2020). The tooth forms of North Atlantic and Mediterranean *S. canicula* could therefore differ because of divergent ontogenetic and allometric patterns among populations. Geometric morphometrics combined with Linear Discriminant Analysis (LDA) are usually used to assess differences in form and shape between objects (e.g., teeth) (Mitteroecker and Bookstein, 2011; MacLeod, 2017). However, this framework requires data modification and reduction (Archer and Kimes, 2008) with a Principal Component Analysis before tooth classification. We published a proof of concept that Random Forests, a supervised machine learning algorithm, achieves better performances than traditional framework with LDA at classifying teeth from two shark populations. This contribution provides the head surfaces of the Atlantic and Mediterranean *S. canicula* specimens (Figure 1A, B, and C) used to extract 3D virtual teeth (Figure 1D) and set landmarks and semilandmarks.

METHODS

S. canicula specimens were sampled at Roscoff (Atlantic Ocean, France) and Banyuls-sur-mer (Mediterranean Sea, France), two localities separated by over 2,000 nautical miles. The heads of 56 *S. canicula* specimens (25 Mediterranean, 31 Atlantic, Table 1) sampled according to three ontogenetic stages (hatchling, juvenile, and sexually mature) were imaged using high-resolution microtomography (µCT) at the MRI platform of the Institut des Sciences de l'Évolution de Montpellier (ISEM) and at the platform SFR Biosciences of the ENS de Lyon. The 3D surfaces were extracted with AVIZO 2019.3 (FEI) software and are provided in .ply format. The voxel size ranges from 0.006 mm to 0.031 mm. A surface was generated for each tooth along the right side of the palatoquadrate (upper jaw) and Meckelian (lower jaw) cartilage. Seven landmarks and 31 semilandmarks were placed on the crown of each tooth with the Landmark software (v3.0.0.6) (Wiley et al., 2005), processed with Scyland3D (Berio and Bayle, 2020), and a Generalized Procrustes Analysis was performed to separate tooth shape and centroid size data. Shape and form (shape + centroid size) were classified into Atlantic and Mediterranean populations with traditional LDA framework and Random Forests. Classification performances were compared between both approaches using accuracy, precision, and recall metrics.

RESULTS

Visually, the main tooth differences between both *S. canicula* populations involve more accessory cusps in the Mediterranean population than in the Atlantic, as well as sharper cusps in juvenile and mature specimens. Furthermore, the teeth of hatchlings and juveniles display similar centroid size patterns along the jaws in both populations but centroid size differences by around

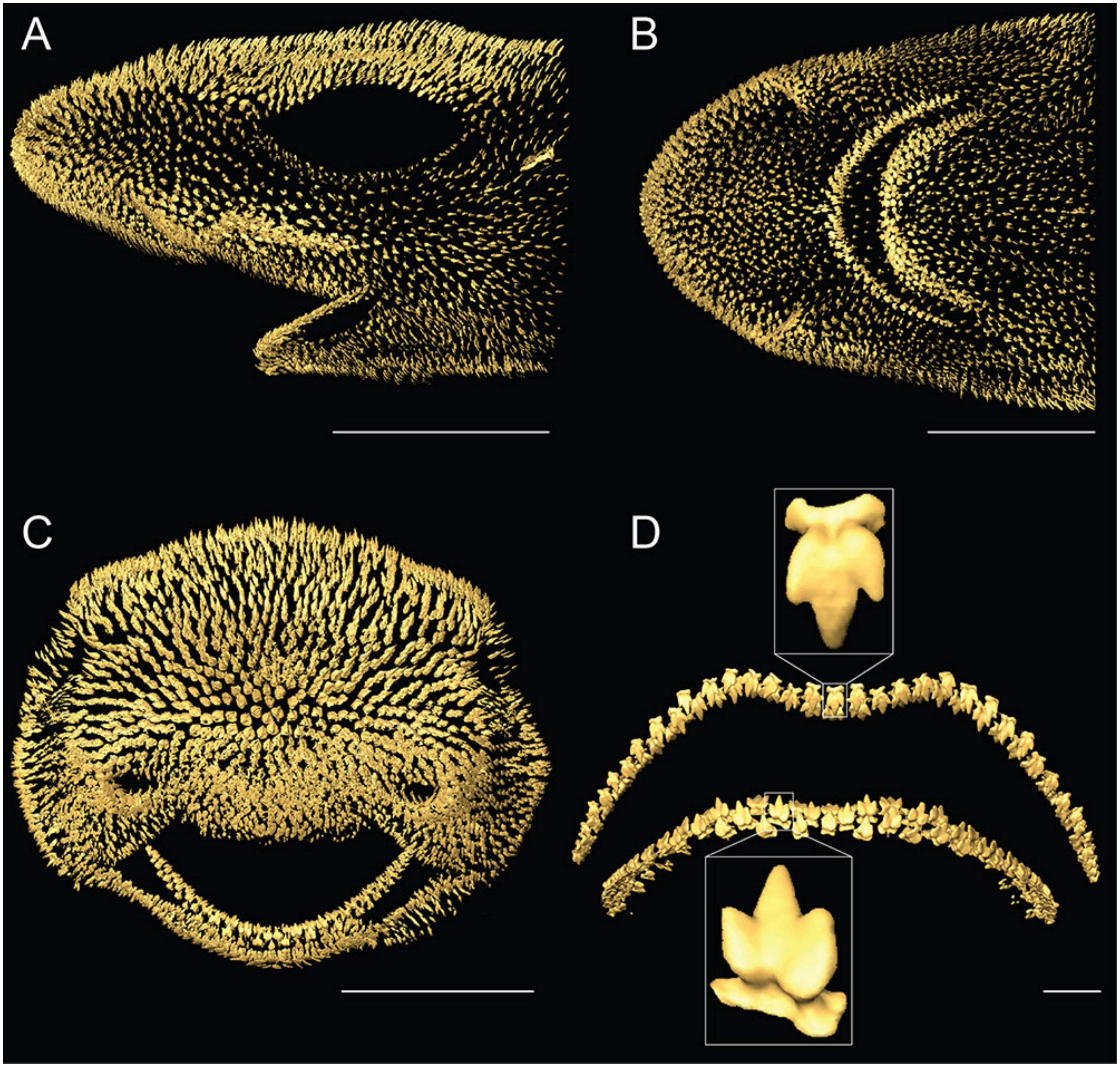

Figure 1. 3D surfaces of a *Scyliorhinus canicula* hatchling female. A) lateral view of the head; B) ventral view of the head; C) frontal view of the head. D) frontal view of palatoquadrate and meckelian teeth after processing the head surfaces. Scale bars are 5 mm in A, B, and C and 500 μm in D.

35% arise at sexually mature stage between populations, which is consistent with the differences observed in body length between sexually mature Atlantic and Mediterranean specimens. The PCA-LDA framework performed on 3D tooth coordinates to classify teeth between populations reaches lesser performances than Random Forests [LDA accuracies: 64.5±0.7 (tooth shapes); 74.6±1.2 (tooth forms); Random Forest accuracies: 81.7±1.7 (tooth shapes); 86.9±1.4 (tooth forms)]. Classification performances are, however, greater for tooth forms than tooth shapes. Moreover, Random Forests indicate which features contribute the most to population discrimination, which are the tooth centroid size and landmarks and semilandmarks located on the lateral edges of teeth, where accessory cusps develop. Different diets as well as genetic isolation between the Atlantic and Mediterranean populations sampled could explain the tooth shape differences observed (Barbieri et al., 2014; Kousteni et al., 2015; 2017) but further investigation is required to challenge these hypotheses. Our study is a proof of concept that geometric morphometrics combined with Random Forests can outperform traditional frameworks used to classify shark teeth. This method could further be used to improve or complete the automatic identification of isolated shark teeth (Naylor and Marcus, 1994; Soda et al., 2017) for example in the fossil record and help tracking illegal trade of sharks, as machine learning methods in this area already display promising results (Barone et al., 2021).

ACKNOWLEDGEMENTS

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 the contribution of SFR Biosciences (UMS3444/CNRS, US8/Inserm, ENS de Lyon, UCBL) AniRa-ImmOs facility. This work was supported by the ENS de Lyon « Attractivité nouveaux professeurs » fund.

BIBLIOGRAPHY

Archer, K. J. and Kimes, R. V., 2008. ‘Empirical characterization of random forest variable importance measures’, *Computational Statistics and Data Analysis*. North-Holland, 52(4), 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Barbieri, M. *et al.*, 2014. ‘Molecular contribution to stock identification in the small-spotted catshark, *Scyliorhinus canicula* (Chondrichthyes, Scyliorhinidae)’, *Fisheries Research*, 154, 11–16. https://doi.org/10.1016/J.FISHRES.2014.01.021

Barone, M. *et al.*, 2021. ‘Performance of iSharkFin in the identification of wet dorsal fins from priority shark species’, *Ecological Informatics*, 101514. https://doi.org/10.1016/j.ecoinf.2021.101514

Berio, F. and Bayle, Y., 2020. ‘Scyland3D: Processing 3D landmarks’, *Journal of Open Source Software*, 5(46), 1262. https://doi.org/10.21105/joss.01262

Berio, F. *et al.*, 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. https://doi.org/10.1111/joa.13257

Berio, F. *et al.*, 2022. ‘Hide and seek shark teeth in Random Forests: machine learning applied to *Scyliorhinus canicula* populations’. *PeerJ*. https://doi.org/10.7717/peerj.13575

Capapé, C. *et al.*, 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. https://doi.org/10.1134/S0032945214010020

Dittman, D. E. *et al.*, 1998. ‘Growth patterns in oysters, *Crassostrea virginica*, from different estuaries’, *Marine Biology*, 132, pp. 461–469. https://doi.org/10.1007/s002270050412

Huret, M. *et al.*, 2019. ‘Variation in life-history traits of European anchovy along a latitudinal gradient: a bioenergetics modelling approach’, *Marine Ecology Progress Series*, 617–618, 95–112. https://doi.org/10.3354/meps12574

Kousteni, V. *et al.*, 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. https://doi.org/10.1038/hdy.2014.107

Kousteni, V. *et al.*, 2017. ‘Diet of the small-spotted catshark *Scyliorhinus canicula* in the Aegean Sea (eastern Mediterranean)’, *Marine Biology Research*, 13(2), pp. 161–173. https://doi.org/10.1080/17451000.2016.1239019

Lombardi-Carlson, L. A. *et al.*, 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. https://doi.org/10.1071/MF03023

MacLeod, N., 2017. ‘On the use of machine learning in morphometric analysis’, in *Biological shape analysis: proceedings of the 4th international symposium*, 134–171. https://doi.org/10.1142/9789813225701_0008

Mitteroecker, P. and Bookstein, F., 2011. ‘Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics’, *Evolutionary Biology*, 38, 100–114. https://doi.org/10.1007/s11692-011-9109-8

Naylor, G. and Marcus, L., 1994. ‘Identifying isolated shark teeth of the genus *Carcharhinus* to species: relevance for tracking phyletic change through the fossil record’, *American Museum Novitates*, 3109, 1–53.

Soares, K. D. A. and de Carvalho, M. R., 2019. ‘The catshark genus *Scyliorhinus* (Chondrichthyes: Carcharhiniformes: Scyliorhinidae): taxonomy, morphology and distribution’, *Zootaxa*, 4601(1), 1–147. https://doi.org/10.1111/joa.13257

Soda, K. J. *et al.*, 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. https://doi.org/10.1002/jmor.20626

Wiley, D., 2006. ‘Landmark Editor 3.0’. Institute for Data Analysis and Visualization, University of California, Davis.

Inv nr.	Sex	Population	Stage	TL
081118A	female	Atlantic	Hatchling	10.6
081118B	female	Atlantic	Hatchling	11.0
081118C	female	Atlantic	Hatchling	11.2
081118D	female	Atlantic	Hatchling	10.2
081118E	male	Atlantic	Hatchling	12.0
081118F	male	Atlantic	Hatchling	10.7
081118G	male	Atlantic	Hatchling	10.8
030418A	female	Atlantic	Hatchling	13.9
030418B	female	Atlantic	Hatchling	13.6
030418C	male	Atlantic	Hatchling	13.4
030418D	male	Atlantic	Hatchling	13.2
071118A	female	Atlantic	Juvenile	36.0
071118B	female	Atlantic	Juvenile	33.0
071118C	female	Atlantic	Juvenile	32.0
071118D	male	Atlantic	Juvenile	35.0
071118E	male	Atlantic	Juvenile	35.0
071118F	male	Atlantic	Juvenile	33.0
121118G	female	Atlantic	Juvenile	36.0
121118H	female	Atlantic	Juvenile	35.0
121118I	male	Atlantic	Juvenile	33.0
121118J	male	Atlantic	Juvenile	36.0
180118A	female	Atlantic	Mature	57.0
180118B	female	Atlantic	Mature	58.0
180118C	female	Atlantic	Mature	58.5
180118D	male	Atlantic	Mature	56.0
180118E	male	Atlantic	Mature	58.0
180118F	male	Atlantic	Mature	59.0
270918A	male	Atlantic	Mature	56.0
270918B	male	Atlantic	Mature	59.5
270918C	female	Atlantic	Mature	63.0
270918D	female	Atlantic	Mature	64.0
12111931	male	Mediterranean	Hatchling	9.5
12111933	female	Mediterranean	Hatchling	9.5
190118A	female	Mediterranean	Hatchling	8.8
190118C	female	Mediterranean	Hatchling	9.0
190118D	male	Mediterranean	Hatchling	8.9
190118F	male	Mediterranean	Hatchling	9.1
060718A	male	Mediterranean	Juvenile	25.5
060718B	female	Mediterranean	Juvenile	23.0
060718C	male	Mediterranean	Juvenile	28.0
060718D	male	Mediterranean	Juvenile	21.0
060718E	male	Mediterranean	Juvenile	23.5
060718F	female	Mediterranean	Juvenile	22.5
121218A	female	Mediterranean	Juvenile	31.0
121218B	female	Mediterranean	Juvenile	31.0
121218C	female	Mediterranean	Juvenile	31.0
121218D	male	Mediterranean	Juvenile	31.0
200118A	male	Mediterranean	Mature	46.0
200118B	male	Mediterranean	Mature	44.0
200118C	male	Mediterranean	Mature	41.0
200118D	male	Mediterranean	Mature	42.0
200118E	female	Mediterranean	Mature	40.0
200118F	female	Mediterranean	Mature	41.5
200118G	female	Mediterranean	Mature	40.0
200118H	female	Mediterranean	Mature	47.0
200118I	female	Mediterranean	Mature	45.0

Table 1. List of *Scyliorhinus canicula* specimens included in the study. TL, total length in cm, from the tip of the snout to the tip of the tail.