SUPPLEMENTARY MATERIALS

Detecting Mediterranean White Sharks with Environmental DNA

Jeremy F. Jenrette, Jennifer L. Jenrette, N. Kobun Truelove, Stefano Moro, Nick I. Dunn, Taylor K. Chapple, Austin J. Gallagher, Chiara Gambardella, Robert Schallert, Brendan D. Shea, David J. Curnick, Barbara A. Block, and Francesco Ferretti

https://doi.org/10.5670/oceanog.2023.s1.28

SAMPLE COLLECTION

Samples were collected in duplicates from each site using a 5L Niskin bottle. Whenever possible, we collected water from four different depths: 0 m, 10 m, 30 m, and 100 m (Truelove et al., 2019). Seawater was first contained in sterile 2L plastic bottles to be used for filtering (see Figure 2b).

FILTRATION AND DNA EXTRACTION

We filtered two to four liters of seawater using a hand pump, an 80 mm buchner funnel, and 75 mm diameter filter paper with 0.2 µm pore size. The filter paper was cut in half; one section was used for onboard analysis and the other section for analysis at the Virginia Tech Genomics Sequencing Center (VT-GSC). The amount of seawater collected per sample was selected based on available labor and processing time (see Figure S1). As the expedition continued, it became desirable to increase the concentration of extracted eDNA at the cost of labor and processing time. We did not detect white sharks in any samples where 4L of seawater was filtered.

We performed DNA extraction using an RNAGEM V extraction kit from MicroGem. Filter paper was treated with buffer solution (enough for a final volume of 1 mL) in addition to 1.5 µL of lysis enzyme before light vortexing. The solution was heated to 75°C for 15 minutes to lyse cells and extract both free-flowing and non-free-flowing DNA. We did not pool duplicates. Then we determined the concentration of DNA using a Qubit dsDNA high-sensitivity assay. We aliquoted 28 µL of the extract for the PCR step. We preserved the remaining filter paper in 500–700 µL DNA/RNA shield for safe transportation and analysis at the VT-GSC where they were stored at –25°C.

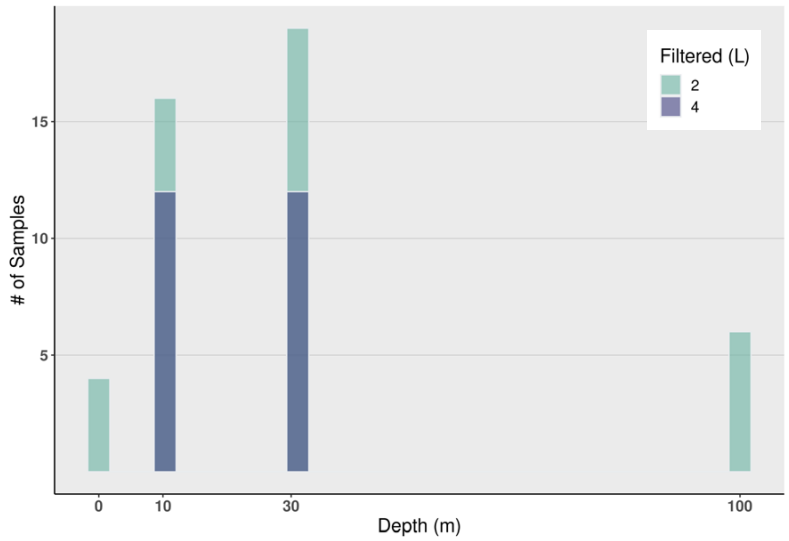

FIGURE S1. Sampling strategy for depth and amount of filtered seawater (2L: n = 45, 4L: n = 24).

LABORATORY ANALYSIS

For quasi real-time detection, analysis was performed in the field on board a Lagoon 42 Memi Catamaran vessel. On board, we filtered seawater, extracted eDNA, and performed PCR and gel electrophoresis protocols. Samples were also analyzed at the VT-GSC, where we purified DNA extract, performed PCR and library preparation, and sequenced samples using a standard Illumina MiSeq.

PCR

We used a portable mini16 thermal cycler from the miniPCR manufacturer to amplify white shark DNA onboard. At the VT-GSC, we performed two PCR reactions (one for initial amplification and a second for library preparation). We used white shark-specific PCR primers designed by Lafferty et al. (2018) to amplify white shark DNA and minimize false-positive detection with other Elasmobranch species. The forward primer, 5′ CGTCACCCCTCCACACATTA 3′, and the reverse primer, 5′ GGTGCTGCTACGTTGTTTGG 3′, amplify a unique fragment of the mitochondrial cytochrome B gene (CYTB) in white sharks.

Reactions were composed of 10 µL of Platinum SuperFi II PCR master mix, 0.5 µM of primers, 10 µL of nuclease-free water, and 28 µL of DNA extract. PCR conditions began with initial denaturation at 94°C for 3 minutes, followed by 40 cycles of 94°C for 30 seconds, annealing at 52°C for 30 seconds, and extension at 72°C for 30 seconds. Final extension was carried out at 72°C for 1 minute. For the library-preparation PCR reaction, we used Kapa HiFi hotStart Ready Mix. We subsequently quantified the DNA concentration in each sample with a Qubit fluorometer (Thermo Fisher Scientific). At the VT-GSC, we performed double bead purification to discharge DNA contaminates and inhibitors.

To minimize the possibility of contamination onboard, we washed water containers with 5% bleach (or we used new sterile containers) before collecting seawater. We washed filtering equipment with the same solution before filtering each sample. To prevent cross-contamination, there was no white shark DNA present on board. We sterilized the onboard lab area with 70% ethanol before performing assays. We used controls at the PCR step to validate the extraction of the white shark-specific CYTB gene fragment. The negative control consisted of 10 µL Platinum SuperFi II PCR master mix, 2 µL of forward and reverse primers, and 38 µL of nuclease-free water. The PCR conditions remained the same, and we used negative controls onboard as well as at the VT-GSC. Additionally at the VT-GSC, we used white shark genomic DNA at the PCR step as a positive control to show the correct amplification of the CYTB gene. The positive control consisted of 10 µL Platinum SuperFi II PCR master mix, 2 µL of primers, 10 µL of nuclease-free water, and 140 ng of white shark genomic DNA. We sterilized the work area with 70% ethanol and ultraviolet light for 20 minutes. There were no indications of contamination onboard or at the VT-GSC.

Onboard, we observed the amplified eDNA on a GELATO gel electrophoresis machine from miniPCR (see Figure 2c–d). We poured a 1.5% agarose gel, using agarose tablets, into a GELATO housing unit from miniPCR. We loaded 100 base-pair DNA ladders, samples, and one negative control for each assay. We applied electrophoresis at 75 volts for 45 minutes to pull the DNA. Then we checked the gel for any DNA bands. If a band was visualized at approximately 200–300 base-pairs, then the white shark-specific CYTB gene fragment was deemed qualitatively detected. At the VT-GSC, we used a TapeStation apparatus to visualize amplification product.

MISEQ SEQUENCING

At the VT-GSC, we prepared the amplification product for sequencing. The libraries were prepared from the amplified products using a KAPA HyperPrep kit. This kit allowed us to attach Illumina compatible adapters and barcodes to each sample for sequencing. The samples were pooled and loaded onto a standard MiSeq v2 300 Nano and sequenced in a 2 × 150 bp paired-end format. The MiSeq run was performed with an added 10% Phix spike to compensate for low base diversity and improve sequencing alignment. Discovering the sequences of the samples was accomplished by Illumina Real Time Analysis.

BIOINFORMATICS

We developed a custom pipeline to compare sequencing reads (FastQ data) with the reference sequence of the CYTB gene for *Lamnidae* species. We developed a list of sequence variants with the DADA2 pipeline in R (Callahan et al., 2016). This pipeline allowed us to distinguish sequence variation to a single nucleotide, increasing taxonomic identification accuracy. We implemented a command line basic local alignment search tool (BLAST) to compare our sequences with the reference sequence. We used a cutoff threshold of 95% for match percent identity to the white shark CYTB sequence to determine a positive detection.

Lastly, because shortfin mako sharks (*Isurus oxyrinchus*) are abundant in the region, have their own CYTB gene fragment that is 89% similar to the white shark fragment, and an individual was observed at a sampling site, we compared the target area of the amplified gene to determine whether sequencing could distinguish one species from the other using the designed primers (Figure S2). We did this by recording the amount of single nucleotide

polymorphisms (SNPs) in the target sequence of mako and comparing base mismatches in our samples to both species and the positive control. The alignment indicates a higher similarity with white shark than mako.

PREDICTING SPATIAL OCCURRENCE

A previous analysis confirmed the Sicilian Channel as a white shark hotspot in the Mediterranean Sea, though high-resolution patterns, environmental drivers, and seasonal dynamics remain largely unknown (Moro et al., 2020). In this light, we estimated high-resolution relative abundance to identify species hotspots and determine sampling sites within the study area.

Mediterranean white shark opportunistic occurrences were retrieved from the most updated version of the database built by Moro et al. (2020). Presence locations represent presence-only data (Pearce and Boyce, 2006), and a natural way to spatially analyze them is via Point Process Models (PPMs) (Renner et al., 2015). PPMs characterize the point pattern (i.e., shark occurrence locations) via an intensity function representing the relative abundance of points in the study area.

In our approach, we assumed that white shark occurrences recorded in May–June follow a Log-Gaussian Cox Process (LGCP; Renner et al., 2015). As in regressive approaches, the intensity of the process was also modeled as a function of a set of covariates measured in the same study area (Renner et al., 2015).

Environmental variables (i.e., sea surface temperature, salinity, chlorophyll *a* concentration, oxygen concentration) were retrieved from the Copernicus Marine Service repository (Clementi et al., 2021). These parameters are commonly used in Species Distribution Models of marine predators such as the white shark (Dambach and Rödder, 2011). The bottom slope was obtained from the GEBCO database (The General Bathymetric Chart of the Ocean, https://www.gebco.net/), while we estimated the bottom slope via the "terrain" function of the R package *terra*. Indeed, pelagic predators tend to aggregate around large canyons, seamounts, and steep underwater scarps in the Mediterranean Sea, given these areas' high productivity and their roles in recycling nutrients between shallow and deep waters (Vella and Vella, 2012).

Unlike systematic data, observation effort indices must be considered when analyzing opportunistic data (McPherson and Myers, 2009). Most historical white shark observations are contributed by commercial fisheries encounters (tuna traps, gillnets, and small-scale fishing) and happenchance sightings (strandings and predation events). Therefore, we considered three primary observation sources (i.e., fishing activities, tuna trap catches, and sightings) to be controlled with different observation effort proxies. Tuna traps highlighted a high catchability for this species in the Mediterranean Sea, with almost half of the catches related to this fixed fishing gear (Moro et al., 2020). Thus, we also chose to build an observation effort index for this process. The fishing effort, extracted from the Global Fishing Watch database (https://globalfishingwatch.org/), was used to control fishery-related occurrences. Tuna trap-related effort was estimated by geolocating 497 different Mediterranean tuna traps and calculating an Inverse Distance Weighting (IDW) considering a scaling factor of 38 km, representing the daily average distance traveled by a white shark in coastal areas (Jorgensen et al., 2010). Opportunistic sightings were, instead, standardized considering the density of pleasure boats, retrieved from the EMODnet Human Activities portal (https://www.emodnet-humanactivities.eu/). The relative indices of each proxy were summed up and included in the model as an offset parameter.

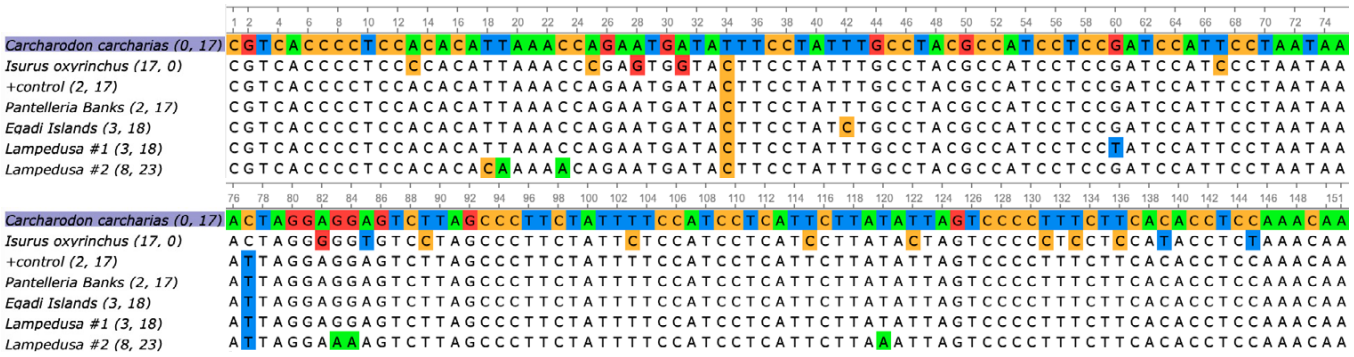

FIGURE S2. Sequence alignment of the white shark's (*Carcharodon carcharias*'s) 151 base-pair region of the CYTB gene. Colored bases represent either the main sequence alignment or mismatches of shortfin mako (*Isurus oxyrinchus*) and the samples to white shark. In the labels, (# of SNPs (single nucleotide polymorphisms) to white shark, # of SNPs to shortfin mako). *Figure created with the software Unipro UGENE (Okonechnikov et al., 2012)*

Finally, we predicted the relative intensity of white shark records within the Sicilian Channel, considering a fixed amount of observation effort.

PARTICLE TRACKING MODELING

Inferring past habitat preferences of white shark provides valuable information for predicting current and future presence of the animal. After detecting white shark eDNA, we produced an approximate estimate of the most likely geographical areas where the eDNA was shed by simulating movement of eDNA molecules backwards in time over 128 hours.

To model eDNA movement, we first downloaded water velocity data from June 5, 2021, to June 29, 2021, at ~10 m depth increments from 1 m to 128 m (n = 13) for the Sicilian Channel at a spatial resolution of 0.042° × 0.042° from the Mediterranean Sea Physics Analysis and Forecast model on the Copernicus Marine Service server (Clementi et al., 2021).

To create the simulations, we used the Lagrangian particle tracking model and the open-source software OpenDrift (Dagestad et al., 2018) following Andruszkiewicz et al. (2019) in Python 3 (Van Rossum and Drake, 2009) with the downloaded water current data. OpenDrift uses the Eulerian velocity fields obtained from the water velocity data to transport particles within a geographical range over set time steps. The PelagicEgg source code (which is established in the OpenDrift code to model the movement of fish eggs) was modified to remove the terminal velocity (settling rate) and buoyancy of particles, allowing the displacement of particles to be directed solely by water movement. Water current and wind uncertainty constants of 0.1 and 0.2, respectively, were added to introduce environmental randomness to the data following examples in the OpenDrift model code (Dagestad et al., 2018). One thousand particles were used in the simulations due to excessive computational costs for larger simulations, and robustness testing suggested that there was little difference in simulation variability if more particles were used. The particles were seeded from the location of each positive eDNA detection location at the time, date, and depth that the water sample was taken. The drift model was processed as a hindcast in 15-minute steps, showing where the particles may have been in the hours before detection. The simulations were run for a minimum of 4 timesteps (1 hour) and the time period was increased by doubling up to 512 steps (128 hours), providing a safe time period for encompassing the maximum time of eDNA persistence (eDNA may be difficult to detect after 48 hours in marine systems; Collins et al., 2018). As we were not able to directly degrade particles when simulating in hindcast, we used time of simulation as a proxy for degradation rate and assumed that eDNA is more likely to have originated in the points that represent shorter hindcast times. The resulting netCDF file was processed in R-studio (R Core Team, 2020) using the *ncdf4* package to recover the final location of each seeded particle at each time increment.

RESULTS AND DISCUSSION

We detected white shark eDNA from June 12 to June18, 2021, with >95% match identity at two sites in the Sicilian Channel—the Pantelleria Banks, and west of Lampedusa (see Figure 1). At The Egadi Islands, 92% match identity was observed for the CYTB gene fragment, and though it was not above the cut-off threshold, it was worth noting as a potential white shark detection. On board, white shark DNA was qualitatively detected once by gel electrophoresis at Lampedusa; thus, there were at least two instances of false-negative error. At the VT-GSC, PCR amplification product was observed in four samples, two of which were collected in the same duplicate (also at Lampedusa), on TapeStation.

A juvenile male shortfin mako was seen on a subsurface camera deployed at 10 m depth at the Pantelleria Banks during this survey. As a result of this sighting, we examined the target 151 base-pair region of the CYTB gene for mako sharks and identified 17 SNPs (compared to white shark CYTB), indicating an 11.3% difference. Though the primers used in this study were designed for California white shark populations, Gubili et al. (2011) and Leone et al. (2020) discovered that Mediterranean white sharks are most genetically similar to northeastern Pacific populations. Because of data paucity, the acute genetic variation is unclear between the two populations in regards to the mitochondrial CYTB gene.

Environmental DNA movement hindcasts showed the most likely position at which a detected white shark shed its DNA within five days of the detection. Five days represents an overestimate of eDNA persistence in marine systems, which is predicted to be around two days (Collins et al., 2018), and so our 128-hour predictions are likely overestimates of where the shark may have been when the eDNA was shed. From this spatiotemporal information, we recorded likely white shark presence for use in updating abundance maps so that supplemental surveys and conservation actions are more informed.

REFERENCES

Andruszkiewicz, E.A., J.R. Koseff, O.B. Fringer, N.T. Ouellette, A.B. Lowe, C.A. Edwards, and A.B. Boehm. 2019. Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Frontiers in Marine Science* 6:477, https://doi.org/10.3389/fmars.2019.00477.

Callahan, B., P. McMurdie, M. Rosen, A. Han, A. Johnson, and S. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581–583, https://doi.org/10.1038/nmeth.3869.

Clementi, E., A. Aydogdu, A.C. Goglio, J. Pistoia, R. Escudier, M. Drudi, A. Grandi, R. Lecci, S. Cretí, S. Ciliberti, and others. 2021. Mediterranean Sea physical analysis and forecast. (CMEMS MED-Currents 2016-2019) (Version 1) [s med-currents, eas6 system]. Copernicus Monitoring Environment Marine Service (CMEMS), https://doi.org/10.25423/cmcc/medsea_analysis_forecast_phy_006_013_eas4.

Collins, R.A., O.S. Wangensteen, E.J. O'Gorman, S. Mariani, D.W. Sims, and M.J. Genner. 2018. Persistence of environmental DNA in marine systems. *Communications Biology* 1(1):185, https://doi.org/10.1038/s42003-018-0192-6.

Dagestad, K.-F., J. Röhrs, Ø. Breivik, and B. Ådlandsvik. 2018. Opendrift v1.0: A generic framework for trajectory modelling. *Geoscientific Model Development* 11(4):1,405–1,420, https://doi.org/10.5194/gmd-11-1405-2018.

Dambach, J., and D. Rödder. 2011. Applications and future challenges in marine species distribution modeling. *Aquatic Conservation: Marine and Freshwater Ecosystems* 21(1):92–100, https://doi.org/10.1002/aqc.1160.

Gubili, C., R. Bilgin, E. Kalkan, S. Karhan, C.S. Jones, D.W. Sims, H. Kabasakal, A.P. Martin, and L.R. Noble. 2011. Antipodean white sharks on a Mediterranean walkabout? Historical dispersal leads to genetic discontinuity and an endangered anomalous population. *Proceedings of the Royal Society B: Biological Sciences* 278(1712):1,679–1,686, https://doi.org/10.1098/rspb.2010.1856.

Jorgensen, S.J., C.A. Reeb, T.K. Chapple, S. Anderson, C. Perle, S.R. Van Sommeran, C. Fritz-Cope, A.C. Brown, A.P. Klimley, and B.A. Block. 2010. Philopatry and migration of Pacific white sharks. *Proceedings of the Royal Society B: Biological Sciences* 277(1682):679–688, https://doi.org/10.1098/rspb.2009.1155.

Lafferty, K.D., K.C. Benesh, A.R. Mahon, C.L. Jerde, and C.G. Lowe. 2018. Detecting southern California's white sharks with environmental DNA. *Frontiers in Marine Science* 5:355, https://doi.org/10.3389/fmars.2018.00355.

Leone, A., G.N. Puncher, F. Ferretti, E. Sperone, S. Tripepi, P. Micarelli, A. Gambarelli, M. Sarà, M. Arculeo, G. Doria, and others. 2020. Pliocene colonization of the Mediterranean by great white shark inferred from fossil records, historical jaws, phylogeographic and divergence time analyses. *Journal of Biogeography* 47(5):1,119–1,129, https://doi.org/10.1111/jbi.13794.

McPherson, J.M., and R.A. Myers. 2009. How to infer population trends in sparse data: Examples with opportunistic sighting records for great white sharks. *Diversity and Distributions* 15(5):880–890, https://doi.org/10.1111/j.1472-4642.2009.00596.x.

Moro, S., G. Jona-Lasinio, B. Block, F. Micheli, G. De Leo, F. Serena, M. Bottaro, U. Scacco, and F. Ferretti. 2020. Abundance and distribution of the white shark in the Mediterranean Sea. *Fish and Fisheries* 21(2):338–349, https://doi.org/10.1111/faf.12432.

Okonechnikov, K., O. Golosova, M. Fursov, and the UGENE team. 2012. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28:1,166–1,167, https://doi.org/10.1093/bioinformatics/bts091.

Pearce, J.L., and M.S. Boyce. 2006. Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43(3):405–412, https://doi.org/10.1111/j.1365-2664.2005.01112.x.

R Core Team. 2020. R: A language and environment for statistical computing. Retrieved May 2, 2021, from http://www.R-project.org/.

Renner, I.W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S.J. Phillips, G. Popovic, and D.I. Warton. 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution* 6(4):366–379, https://doi.org/10.1111/2041-210X.12352.

Truelove, N.K., E.A. Andruszkiewicz, and B.A. Block. 2019. A rapid environmental DNA method for detecting white sharks in the open ocean. *Methods in Ecology and Evolution* 10(8):1,128–1,135, https://doi.org/10.1111/2041-210X.13201.

Van Rossum, G., and F.L. Drake. 2009. Python/C API Manual - Python 3. CreateSpace, Scotts Valley, CA, retrieved September 15, 2022, from https://dl.acm.org/doi/abs/10.5555/1593487.

Vella, A., and J.G. Vella. 2012. Central-southern Mediterranean submarine canyons and steep slopes: Role played in the distribution of cetaceans, bluefin tunas and elasmobranchs. Retrieved Oct 15, 2022, from https://www.um.edu.mt/library/oar/handle/123456789/93408.

ARTICLE DOI. https://doi.org/10.5670/oceanog.2023.s1.28