SPRFMO
South Pacific Regional Fisheries Management Organisation

9th MEETING OF THE SCIENTIFIC COMMITTEE

Held virtually, 27 September to 2 October 2021

SC9-SQ07

Genetic analysis of *Dosidicus gigas*

Peru

PO Box 3797, Wellington 6140, New Zealand
P: +64 4 499 9889 – F: +64 4 473 9579 – E: secretariat@sprfmo.int
www.sprfmo.int

IMARPE

South Pacific Regional Fisheries Management Organisation
9th Meeting of the Scientific Committee
Held virtually, 27 September - 2 October 2021

Progress report on the genetic analysis of jumbo flying squid *Dosidicus gigas* collected in different areas within the Peruvian jurisdictional waters during 2018 and 2019

by

Giovanna Sotil and Paul Guarnizo

Instituto del Mar del Peru (IMARPE)
Esquina Gamarra y General Valle s/n, Callao, Perú

This report contains information on the jumbo flying squid stock and fishery in Peruvian jurisdictional waters that, we reiterate, the delegation of Peru, in use of its discretionary powers, voluntarily provides for the purpose of information and support to the scientific research work within the Scientific Committee of the SPRFMO. In doing so, while referring to Article 5 of the Convention on the Conservation and Management of High Seas Fishery Resources in the South Pacific Ocean and reiterating that Peru has not given the express consent contemplated in Article 20 (4) (a) (iii) of the Convention, Peru reaffirms that the decisions and conservation and management measures adopted by the SPRFMO Commission are not applicable within Peruvian jurisdictional waters.

2021

SUMMARY

Some population genetic studies tend to use single gene analysis, as the mitochondrial DNA markers. However, its limitations are well known compared with the resolution of genomic molecular markers using NGS technology, when they are used for estimating contemporary population structure in organisms with large and highly connected populations. Therefore, it is important to note that several molecular markers can provide diverse but complementary evolutionary information. In this sense, we report advances in the genetic population analysis of jumbo squid collected in different areas within the Peruvian jurisdictional waters. The mtDNA (ND2 and COI) genes and SNPs analysis with ddRAD-seq technique are being evaluated. Preliminary results of ND2 gene analysis are presented, comparing the genetic diversity between groups established according to different phenotypes (small and medium sizes), latitudinal (north, central and south) and longitudinal (coastal and oceanic) distribution of samples. From the ND2 gene analysis, no genetic differences were observed between small and medium phenotypes, while the highest haplotype diversity was observed in oceanic organisms, and the highest nucleotide diversity from organisms collected in the central part of Peru marine area. In addition, the optimization of ddRAD-Seq protocol for the *D. gigas* library preparation using the EcoRI-HF and SbfI-HF restriction enzymes is presented, and further analysis are being performed for the evaluation of SNPs.

TABLE OF CONTENTS

1. Introduction

Jumbo flying squid *Dosidicus gigas* is widely distributed and inhabits highly variable environments in the eastern Pacific Ocean, exhibiting a complex structure with the presence of three size-phenotypic groups at sexual maturity (Nigmatullin 2001). Like other cephalopods, *D. gigas* has a short life span and fast somatic growth, generally composed by individuals of a single annual class. These characteristics make squid populations sensitive to changes in environment conditions affecting their abundance and distribution patterns. These three groups tend to segregate their principal nuclei of abundance, and exist as separate dominant groups that tend to alternate sporadically, from one year to the next and in longer decadal time-scales which could be interpreted as regime changes (Csirke *et al.* 2018, Arguelles *et al.* 2019). The nature of these alterations and possible regime changes requires further investigation in order to elucidate the influence of the genetic variation in population structure.

Mitochondrial DNA (mtDNA) is one of the most widely genetic tools used for population genetic analysis in animals, with a relatively high variability and its nearly neutral model of evolution (Avise *et al.* 1987), but they have limited resolution for elucidating complex relations in populations, compared with genomic approaches. Due to next generation sequencing (NGS), it is possible to assess genetic diversity, scanning the entire genome rather than a single portion of it, looking for single nucleotide polymorphisms (SNPs) at a large set of loci. The ddRADseq technique enables high throughput simultaneous discovery of sequence polymorphisms without a reference genome. And, compared to other genotyping by sequencing (GBS) techniques, allows for greater flexibility, robustness, and a substantial decrease in cost.

In this report, we present the results of the initial progress made in the population genetic analysis of *D. gigas* using mtDNA (ND2 and COI genes) and SNPs markers (based on ddRAD-seq). We evaluated mature organisms from the north, central and southern part of Peruvian waters, collected during 2018 and 2019. We selected small and medium-size phenotypes and extracted DNA from muscle tissue of the mantle. No squids of the large-size phenotype were available during 2018 and 2019. This situation persisted in 2020 and first part of 2021 when, in addition, the field sampling possibilities and the access to laboratory facilities were severely limited due to Covid-19 related restrictions.

2. Materials and methods

- **Sampling**

Muscle tissue from the mantle of 94 individuals was collected, according to protocol indicated by Sotil *et al.* (2019), during 2018 and 2019. At the moment, only two phenotype sizes of mature organisms (stages 3 and 4) were collected, from the north, central and southern part of Peruvian waters. The mantle length of the small-size phenotype group (n=48) ranged from 19.3 to 34.0 cm, while for medium-size was from 50.0 to 66.5 cm mantle length. Also, samples were selected according to their longitudinal distribution, considering as of the coastal group those organisms collected from 38 to 95 nautical miles, and the oceanic group from 156 to 200 nautical miles.

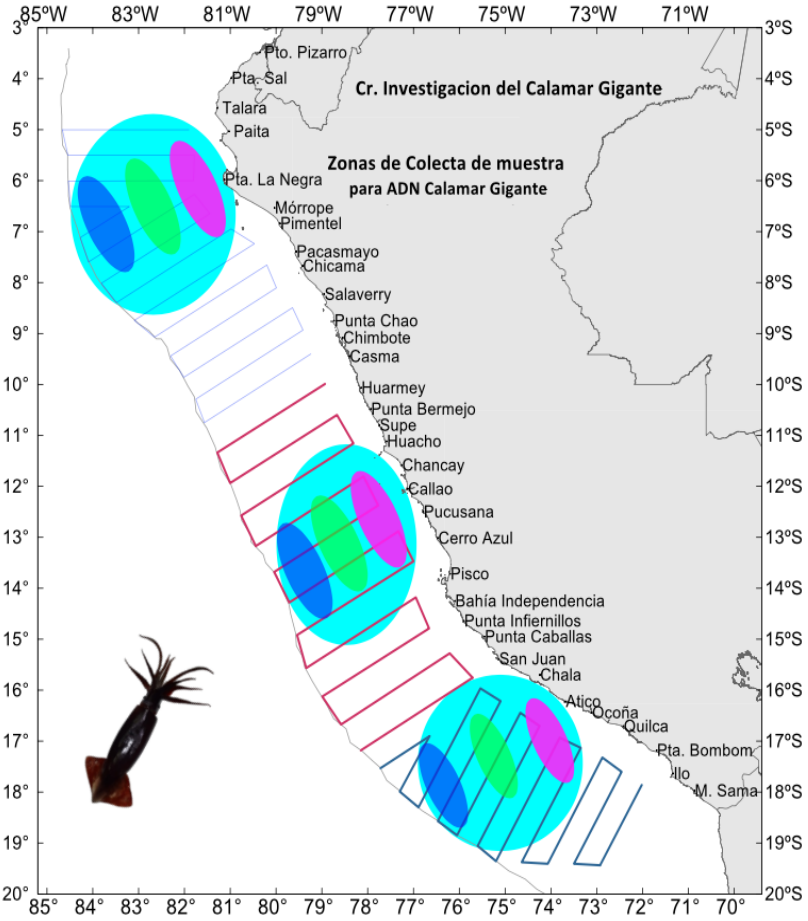


Figure 1. Peruvian sampling areas for genetic studies of the jumbo flying squid (from Sotil *et al.* 2019)

- **DNA extraction**

Total DNA was extracted using the CTAB 2X standard method based on Sambrook and Russell (2001) method, with modifications. All samples were treated with RNAse A (Thermo Scientific) and were resuspended in nuclease free water (NFW). DNA was quantified in a spectrophotometer Nanodrop OneC, and diluted to final concentrations of 200 ng/µL, selecting those with the best purity indices (between 1.8 and 1.6 for A260/280 nm and A260/230 nm, respectively). Also, DNA integrity was verified by electrophoresis in agarose gel 1% with TBE 1X.

- **Mitochondrial DNA genes analysis**

The mitochondrial NAD dehydrogenase subunit II gene was amplified using the SqSF and DCOR2 primers reported in Staff *et al.* (2010). PCR reactions were performed in a Proflex PCR system (Applied Biosystems) thermal cycler, using the kit HotStart Taq Plus Master Mix (Qiagen), with 1-2 ng of total DNA, 0.2 µM of each primer, 1.75 mM of $MgCl_2$ as final concentrations, in a total volume of 10 µL. PCR thermal conditions were an initial denaturation of 95°C for 5 min, followed by 39 cycles of 94°C for 1 min, 47°C for 1 min and 72°C for 1 min, and a final extension of 72°C for 5 min. Amplified products with the

expected size were sequenced on an ABI 3500 Genetic Analyzer (Applied Biosystems) using the same primers for previous PCR reactions. Obtained electropherograms were manually edited using Chromas Lite version 2.6.5 (Technelysium Pty Ltd, South Brisbane, Australia); the consensus sequences were aligned using MUSCLE (Edgar 2004) implemented in Geneious Prime® 2021.1.1 (https://www.geneious.com), and finally trimmed to 1082 bp length.

The genetic diversity indices were calculated in DnaSP v6 (Rozas *et al.* 2017), including the number of haplotypes (H), number of polymorphic sites (S), haplotype diversity (Hp) and nucleotide diversity (π). Analysis was performed considering the total samples and also three different group criteria (phenotype size, latitudinal and longitudinal distribution). The genetic differentiation among groups was evaluated in Arlequin 3.5.2.2 (Excoffier 2010) for the Fst calculation with 1000 iterations. The relationship between haplotypes was evaluated with a haplotype network built using the Median-Joining algorithm (Bandelt 1999), in PopArt v1.7 (Leigh & Bryant 2015). Additionally, to evaluate the isolation by distance (IBD) hypothesis, the Mantel test was performed in Arlequin, with 1000 permutations.

We also amplified the mtDNA cytochrome oxidase subunit I gene (COI), using the primers ZplankF1 and ZplankR1 described by Prosser *et al.* (2013). PCR products (~700 bp) were sequenced for both strands, and the consensus sequences are being analyzed, considering the same criteria for the ND2 gene analysis, previously mentioned.

- **ddRAD-Seq library preparation**

Libraries were prepared according to DaCosta and Sorensen (2014) protocol, with some modifications optimized for our samples, like total genomic DNA, conditions of digestion, ligase concentration, and PCR reaction. Briefly, total DNA was digested with EcoRI-HF and SbfI-HF restriction enzymes (New England Biolabs) during 4 h at 37°C. Digestion effectiveness was verified in agarose gel 2%. P1 and P2 adapters were ligated with digested DNA, using ATP (New England Biolabs) and T4 DNA Ligase (New England Biolabs), in a final volume of 70 µL. Ligated samples were run in 2% LMP agarose electrophoresis, and DNA sizes between 300 to 450 bp were selected by cutting a gel area. GelGreen dye and blue light exposure was used to reduce DNA damage during gel evaluation. Selected DNA from gels were purified using MinElute Gel Extraction Kit (Qiagen), and total volume was used for the PCR amplification using 0.5 µM of RAD.1F and RAD.1R primers. PCR was performed using a Phusion High Fidelity Hot Start Flex 2X master mix (New England Biolabs), and thermal cycler conditions were: an initial denaturation of 98°C for 30 s, followed by 25 cycles of 98°C for 10s, 58°C for 30 s and 72°C for 40 s, with a final extension of 72°C for 5 min. Products were verified in 2% agarose gel, purified using AMPure XP (Beckman Coulter), and eluted in NFW.

For library normalization, two independent 1:1000 library dilutions were quantified, in duplicates. Quantification was performed with KAPA Library Quant Kit (Illumina) Universal qPCR mix (ROCHE), in a QuantStudio™ 5 Real-Time PCR System. Libraries will be sequenced in one lane of Illumina 2500 platform, 150 bp PE sequencing.

3. Preliminary results

- **mtDNA ND2 gene**

We obtained longer mtDNA ND2 (1082 bp) sequences than previously reported in the literature on genetic population studies in this species. A total of 38 haplotypes were identified in the 90 individuals evaluated. Haplotypes were differentiated by one to three mutation steps. The most common haplotype (Hap1) was registered in 41 organisms, which means that nearly 50% of the samples collected shared the same haplotype. Haplotype Hap11 was represented by 7 individuals, Hap7 and Hap38 by 3 individuals each, Hap6 and Hap24 with 2 individuals, while the other haplotypes were represented only by one organism. The median-joining haplotype network showed that the common haplotype Hap1, was present in small and medium-size organisms, from the north, central and southern part of Peruvian waters, and from coastal and oceanic zones. The haplotype diversity (Hd) of the total sample was 0.788 ± 0.046, nucleotide diversity (π) was 0.00187. There were 44 polymorphic sites, from which 31 were singleton variable sites, and 13 parsimony informative sites 13.

For this mtDNA gene, no genetic diversity differences were observed between small and medium-size phenotypes, while the highest haplotype diversity was observed within oceanic organisms, and the highest nucleotide diversity was from organisms collected in the central part of Peruvian waters (Table 1).

Low and non-significant Fst values showed no genetic differentiation of individuals grouped by sizes (FST = 0.0006). Similar percentages of variation and fixation indices were found when they were grouped by location, of coastal vs. oceanic and of north vs. center vs. south, with FST = 0.0017 and FST = 0.0133, respectively (Table 2). This non-structuration was also seen in the haplotypes networks suggesting the similar genetic background across all samples. The cytochrome c oxidase I (COI) gene and a smaller portion of the ND2 gene also revealed no population structure when comparing different points along the Pacific Coast including Peruvian and Chilean populations (Sanchez 2016).

The correlation coefficient (rY1 = 0.1361, p value = 0.157) suggested no correlation between geographical and genetic distances matrices when the 12 sub-groups were compared in the Mantel test. This result suggests that there is not a distribution pattern known as isolation by distance. IBD on squid populations is poorly studied but was found when comparing northern and southern hemisphere populations (Sandoval-Castellanos *et al.* 2007).

However, Teske *et al.* (2018) conclude that mtDNA is not suitable to address the IBD hypothesis for being a single-locus marker and for the probable selection-driven reduction in the genetic diversity that hide spatial genetic differentiation. In contrast, SNP markers were proved to detect it, hence an advantage of genomic data over genetic markers.

Table 1. Genetic diversity indices of jumbo squid ND2 gene sequences (1082 bp) collected from Peruvian waters in 2018 and 2019. Haplotypes were analyzed as a total sample and between different groups criteria (phenotype-size and geographical distribution). n = number of individuals, H = number or haplotypes, Hd = haplotype diversity, π = nucleotide diversity.

Groups		n	H	Hd	π
Total	-	90	38	0.788	0.00187
Phenotype-size	Small size	47	20	0.74098	0.00186
	Medium size	43	21	0.83499	0.00187
Latitudinal distribution	North	28	13	0.81481	0.00143
	Central	24	13	0.84783	**0.00233**
	South	38	17	0.72546	0.00188
Longitudinal distribution	Coastal	46	20	0.68502	0.00149
	Oceanic	44	22	**0.87315**	0.00186

Table 2. AMOVA for different groups evaluation of *Dosidicus gigas*, based on mtDNA ND2

Sources of variation		Percentage of variation		
		Small vs. Medium size	Coastal vs. Oceanic	North vs. Central vs. South
Among groups		0.04	0.36	0.31
Among population within groups		0.03	-0.19	1.02
Within groups		99.94	99.83	98.67
Fixation indices:	Fsc	0.0002	- 0.0019	0.0103
	Fst	0.0006	0.0017	0.0133
	Fct	0.0004	0.0036	0.0031

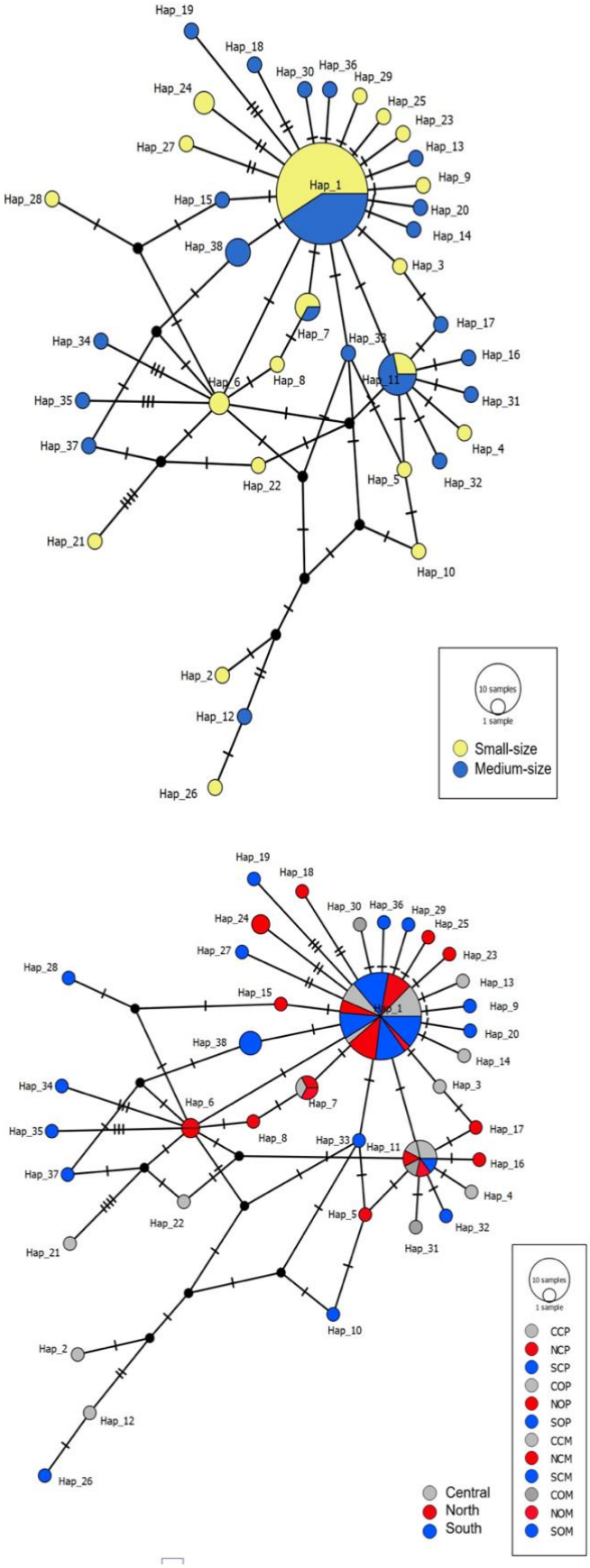


Figure 2. Haplotype networks based on *Dosidicus gigas* mtDNA ND2, discriminating size-phenotypes (top) and sampling zones (bottom) groups. Each circle represents one haplotype. The size of the circles is proportional to the number of haplotype copies (samples). Colors represent groups. Each dash represents one mutation step. Small black points represent absent haplotypes

- **Optimization of ddRAD-seq libraries preparation**

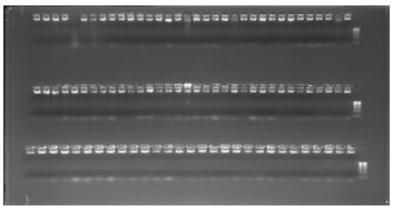
After several evaluations, some modifications to the DaCosta and Sorenson protocol (2014) have been incorporated into our ddRAD-Seq library final protocol for the *D. gigas* DNA sample analysis. The laboratory steps and results are illustrated in the flowchart below.

First, high DNA quality samples (according to parameters of integrity and purity) were obtained using the CTAB 2X extraction method, due also to the sampling protocol used and proposed previously (Sotil *et al.* 2019). Also, the 300 bp size and yield of PCR products from library amplification done using RAD primers, after the ligation step, were obtained as was expected. Moreover, the standard curve obtained from the qPCR analysis showed a perfect fit of points (R^2=1.0 and Efficiency=96%) for the calculation of the libraries' quantification. Additionally, the melting curves demonstrated the presence of one unique peak, suggesting the absence of undesired regions. In this sense, we expect to obtain adequate numbers of loci for the evaluation of SNPs in this species.

Flowchart of ddRAD-Seq Library preparation for the evaluation of SNPs in *Dosidicus gigas*

High quality of genomic DNA extractions

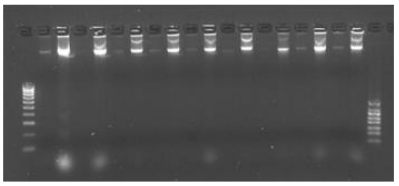
- CTAB 2X method with RNAse treatment
- Agarose gel evaluation
- Spectrophotometer quantification
- Purity indices control [1.6-1.8, for A260nm/A230nm and A260nm/A280nm, respectively], and selection of best samples



Agarose gel with DNA extractions

Double digestion of genomic DNA

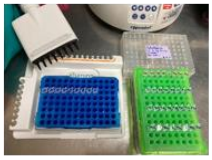
- Restriction enzymes: EcoRI HF and SbfI-HF (NEB)
- Incubation: 37°C for 4 h in NEB 4 buffer
- DNA: 0.8 µg
- Verification of digestion in 2% agarose gel



Non-digested and digested (intercalated in wells) genomic DNA evaluated in agarose gel

DNA Ligation

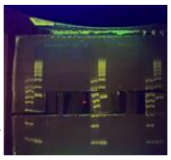
- 1400 U T4 ligase per reaction



P1 and P2 barcodes ligated to digested DNA

DNA Size selection

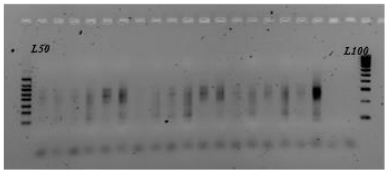
- LMP agarose electrophoresis
- GelGreen and blue light gel evaluation for selecting area cut of DNA in LMP agarose gel



LMP agarose gel area cut, selected according molecular marker size references

PCR amplification

- Primers: RAD.1F and RAD.2R
- Product verification in 2% agarose gel



Amplified products with expected sizes, evaluated in 2% agarose gel. L50=50bp size ladder, L100=100 bp size ladder

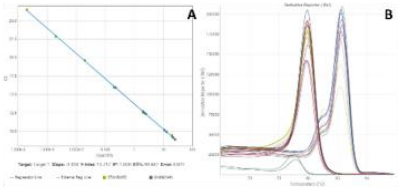
DNA Purification

- AmpureXP beads and Nanodrop quantification
- Check of purity indices

DNA purification with beads

DNA Libraries Normalization

Library quantification from a standard curve (A) using the KAPA kit. Verification of correct products in Melting Curves (B)



Sample pooling for NGS sequencing

4. References

Arguelles J, Csirke J, Grados D, Tafur R & Mendoza J. 2019. Changes in the predominance of phenotypic groups of jumbo flying squid *Dosidicus gigas* and other indicators of a possible regime change in Peruvian waters. Meeting document of the 7th meeting of the SPRFMO Scientific Committee, La Havana, Cuba, 7-12 October 2019. Doc. SPRFMO SC7-SQ03: 19p. http://www.sprfmo.int/assets/2019-SC7/Meeting-Docs/SC7-SQ03-Changes-in-predominance-of-phenotypic-groups-of-Jumbo-flying-squid-in-Peruvian-waters.pdf

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, *et al.* 1987. Intraspecific phylogeography: the mitochondrial-DNA bridge between population-genetics and systematics. Annu. Rev. Ecol. Syst. 18: 489–522.

Teske PR, Golla TR, Sandoval-Castillo J, *et al.* 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. Sci Rep 8, 8448. https://doi.org/10.1038/s41598-018-25138-9

Leigh, JW, Bryant D. 2015. PopART: Full-feature software for haplotype network construction. Methods Ecol Evol 6(9): 1110–1116.

Bandelt H, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16(1):37–48.

Csirke J, Argüelles J, Alegre A, Ayón P, Bouchon M, Castillo G, Castillo R, Cisneros R, Guevara-Carrasco R, Lau L, Mariátegui L, Salazar C, Tafur R, Torrejón J, Yamashiro C. 2018. Biology, population structure and fishery of jumbo flying squid (*Dosidicus gigas*) in Peru. Bol Inst Mar Peru. 33(2): 302-364. https://revistas.imarpe.gob.pe/index.php/boletin/article/view/46/43

DaCosta JM, Sorenson MD. 2014. Amplification Biases and Consistent Recovery of Loci in a Double-Digest RAD-seq Protocol. PLoS ONE 9(9): e106713.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797. https://doi.org/10.1093/nar/gkh340.

Nigmatullin CM, Nesis KN, Arkhipkin A. 2001. A review of the biology of the jumbo squid *Dosidicus gigas* (Cephalopoda: Ommastrephidae). Fisheries Research. 54: 9-19.

Prosser S, Martínez-Arce A and Elías-Gutiérrez M. 2013. A new set of primers for COI amplification from freshwater microcrustaceans. Mol Ecol Resour, 13: 1151-1155. https://doi.org/10.1111/1755-0998.12132

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol. Biol. Evol. 34: 3299-3302. DOI: 10.1093/molbev/msx248

Sambrook J, Russell DW. 2001. Molecular cloning: A laboratory manual (3rd ed.). Cold Spring Harbor, NY, USA: Cold Spring Harbor Press.

Sandoval-Castellanos E, Uribel-Alcocer M, Díaz-Jaimes P. 2007. Population genetic structure of jumbo squid (*Dosidicus gigas*) evaluated by RAPD analysis. Fish. Res. 83: 113–118.

Sotil G, Cueva D, Espinel M and Mariátegui L. 2019. Protocol for the collection of jumbo flying squid *Dosidicus gigas* muscle tissue for molecular analysis used by the Instituto del Mar del Peru (IMARPE). Working paper presented to the 7th meeting of the Scientific Committee of the SPRFMO, La Havana, Cuba, 7-12 October 2019. SPRFMO document SC7-SQ10: 8p. http://www.sprfmo.int/assets/2019-SC7/Meeting-Docs/SC7-SQ10-Protocol-for-collecting-jumbo-flying-squid-muscle-tissue-for-molecular-analysis.pdf

Staaf DJ, Ruiz-Cooley RI, Elliger C, Lebaric Z, Campos B, Markaida U, Gilly WF. 2010. Ommastrephid squids *Sthenoteuthis oualaniensis* and *Dosidicus gigas* in the eastern Pacific show convergent biogeographic breaks but contrasting population structures. Marine Ecology Progress Series, 418, 165-178.