SPRFMO
South Pacific Regional Fisheries Management Organisation

10th MEETING OF THE SCIENTIFIC COMMITTEE

26 to 30 September 2022, Seoul, Korea

SC10-SQ09_rev1

Preliminary results of the genetic population studies of jumbo flying squid collected in Peruvian jurisdictional waters

Peru

IMARPE

South Pacific Regional Fisheries Management Organisation
10th Meeting of the Scientific Committee
Seul, Republic of Korea (held hybrid), 26-30 September 2022

Preliminary results of the genetic population studies of jumbo flying squid collected in Peruvian jurisdictional waters

by

Giovanna Sotil & 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.

2022

SUMMARY

Preliminary results based on mtDNA genes and SNPs analysis with ddRAD-seq technique of *Dosidicus gigas* in Peruvian waters area is reported. Mature organisms (stage III and IV) from the three size phenotypes (small, medium and large), three latitudinal groups (north, central and south), and from two longitudinal distributions (coastal and oceanic groups) were considered. Two mtDNA genes, COI (658 bp) and ND2 (1084 bp), were analyzed. For COI, low genetic diversity (17 haplotypes, nucleotide diversity = 0.00066) and a star-like network was registered in 130 organisms analyzed. On the other hand, with ND2 gene, a higher genetic diversity (49 haplotypes identified in 123 individuals evaluated) was observed. The highest genetic diversity was observed in organisms in the large-size (among phenotypes), central (among latitudinal groups) and oceanic (among longitudinal distribution) groups. Haplotype network showed several rare haplotypes from oceanic organisms with more than two mutational changes. Under different hypothesis, groups comparisons (AMOVA) were done, observing a significant difference among coastal and oceanic groups. Pairwise Fst analysis showed significant differences between central oceanic and south coastal, as well as central oceanic and southern oceanic organisms. These differences were mainly related to the presence of large-size organisms from central oceanic zone. In addition, based on ddRAD-seq genotyping of 28 samples (representative from most of the groups), 310 polymorphic loci and 746 SNPs were retained and used for preliminary analysis. The central oceanic group showed the highest nucleotide diversity and differentiated from others according to PCA analysis.

TABLE OF CONTENTS

SUMMARY 2
TABLE OF CONTENTS 2
1. Introduction 3
2. Materials and methods 3
- Sampling 3
- mtDNA genes analysis 3
- ddRAD-seq analysis 4
3. Preliminary results 4
- mtDNA COI gene 4
- mtDNA ND2 gene 5
- ddRAD genotyping 7
4. Activities in progress 8
5. References 8

1. Introduction

Several reports have been done for the description of population genetic diversity of *D. gigas*, based on mtDNA cytochrome oxidase subunit I (COI) and NAD dehydrogenase subunit II (ND2) genes (Staaf et al. 2010).

As was mentioned in SC09 final report, a survey based on both mtDNA markers integrating information from different geographical zones will give a better description of the genetic diversity of the species and explain the possible differences in distribution and phenotypes better.

In this sense, we present advances in genetic analysis of 123 individuals, including representatives of the three phenotypic groups, from coastal and oceanic distribution, and from the north, central and southern of Peruvian waters. Parameters of population diversity based on mtDNA (ND2 and COI genes) were calculated. DNA sequences used for the analysis were registered in a nucleotide public database as recommended during SC9. Also, preliminary genome-wide SNPs information from some samples was obtained for the assessment of genomic population.

2. Materials and methods

- **Sampling**

Muscle tissue from the mantle of approximately 130 individuals was collected according to the protocol described by Sotil *et al.* (2019). Mature organisms (stage III and IV) were collected during 2018 to 2019 (medium and small phenotype-sizes) and 2021-2022 (large phenotype-size), in Peruvian jurisdictional waters. Samples were discriminated by northern, central and southern zones. Mantle length of the small-size phenotype group ranged from 19.3 to 34.0 cm, for medium-size from 50.0 to 66.5 cm, and large-size from 70 to 90.1 cm. Also, samples were grouped according to their longitudinal distribution (distance from the coast). Organisms collected from 38 to 95 nautical miles were considered as coastal groups, whereas the oceanic groups were from 156 to 200 nautical miles. All analysis were performed considering the three different grouping criteria: phenotype-sizes (large, medium, small), latitudinal (northern, central and southern), and longitudinal distribution (coastal and oceanic).

- **mtDNA genes analysis**

Total DNA was extracted using the CTAB 2X standard method based on Sambrook and Russell (2001), with modifications. COI (658 bp) and ND2 (1082 bp) genes were amplified and nucleotide sequences were obtained according to the protocol described in document SC9-SQ07. Genetic diversity indices (number of haplotypes H, number of polymorphic sites S, haplotype diversity Hp, and nucleotide diversity π) were calculated with DnaSP v6 (Rozas *et al.* 2017). AMOVA analysis was performed considering three groups' criteria. Pairwise Fst analysis was calculated between 15 groups formed

according to different combinations of size-latitudinal-longitudinal distribution, and tested the genetic differentiation among groups using Arlequin 3.5.2.2 (Excoffier 2010) with 1000 iterations (data not shown). The relationship between haplotypes was evaluated with haplotype network built using the Median-Joining algorithm (Bandelt 1999), in PopArt v1.7 (Leigh & Bryant 2015).

- **ddRAD-seq analysis**

Libraries were prepared as was previously reported (SC9-SQ07). Forty-eight samples were multiplexed in nine barcodes. The clone_filter module of STACKS (Catchen *et al*. 2013) was used to remove PCR duplicates from the raw samples. Later, the process_radstacks module demultiplexed all 48 samples with 95% of high-quality reads retained on average. All reads were barcode trimmed and truncated on 120 bp for downstream analysis. Stacks v2.6 was used to identify and extract single nucleotide polymorphism (SNPs) using the gstacks module in a *de novo* analysis, with previously optimized parameters following the protocol in Paris et al. (2017). A subset of the best 12 samples were selected for the optimization of the STACKS parameters (m, M and n). To filter samples with high percent of missing data, the protocol in Cerca et al. (2021) was followed. The basic population statistics and PCA analysis were performed using SNPs calling from 28 samples (representatives from the two groups of longitudinal and three of latitudinal distribution).

3. Preliminary results

- **mtDNA COI gene**

A total of 17 haplotypes, with 15 polymorphic sites, were identified in the 130 individuals evaluated. Low Hd (0.345 ± 0.054) and π (0.00066) values were registered. The haplotype network showed a star-like shape.

Table 1. Genetic diversity indices of gene sequences (658 bp) from D. gigas COI collected from Peruvian waters during 2018-2022. N = number of individuals, S= number of segregating sites, H = number of haplotypes, Hd = haplotype diversity, π=nucleotide diversity.

Groups		**n**	**S**	**H**	**Hd**	**π**
Total	-	130	15	17	0.345 ± 0.054	0.00066
Phenotype-size	Small	49	10	11	0.370 ± 0.089	0.00080
	Medium	45	7	7	**0.454 ± 0.088**	**0.00084**
	Large	36	2	3	0.160 ± 0.080	0.00025
	North	27	5	6	0.450 ± 0.114	0.00095

Latitudinal distribution	Central	50	7	7	0.228 ± 0.079	0.00043
	South	53	8	9	0.400 ± 0.085	0.00073
Longitudinal distribution	Coastal	75	9	9	0.309 ± 0.069	0.00055
	Oceanic	55	9	10	**0.396 ± 0.084**	**0.00081**

- **mtDNA ND2 gene**

A total of 49 haplotypes with 56 polymorphic sites, were identified in the 123 individuals evaluated. The Hd of the total sample was 0.826 ± 0.034, and π was 0.00195. Comparing groups, the highest haplotype and nucleotide diversities were observed in the large-size (among the three phenotypes), central (among three latitudinal distribution groups) and oceanic (among longitudinal distribution) organisms (Table 2).

The most common haplotype (Hap6) was registered in 50 organisms, followed by Hap9 represented by 10 individuals, and Hap2 by 7 organisms, while the other 53 haplotypes were present in one to three organisms. The Median-Joining haplotype network showed that the most common haplotype (Hap6), was present in all sizes, latitudinal and longitudinal groups. Several rare haplotypes from oceanic organisms with more than two mutational changes were observed.

Additionally, a low but significant Fst value showed genetic differentiation (Fst = 0.04298) between latitudinal distribution groups (Table 3). Specifically, when comparing latitudinal and longitudinal distribution groups, a significant Fst value was observed among the central oceanic and the southern coastal groups (Table 4). Another interesting comparison with significant pairwise differences was observed between (1) the large-size organisms from central oceanic and the medium-sizes from oceanic north and south groups, (2) medium-size organisms from central oceanic and south coastal, and (3) small-size from north oceanic and medium-size from north coast. It is important to note that large-size phenotypes were collected during 2021-2022.

Table 2. Genetic diversity indices of ND2 gene sequences (1082 bp) from D. gigas collected from Peruvian waters in 2018 and 2022. Haplotypes were analyzed as a total sample and between different groups criteria (phenotype-size and geographical distribution). N = number of individuals, S = polymorphic sites, H = number or haplotypes, Hd = haplotype diversity, π = nucleotide diversity.

Groups		n	S	H	Hd	π
Total		123	56	49	0.826 ± 0.034	0.00195
Phenotype	small-size	47	25	20	0.741 ± 0.071	0.00186
	medium-size	43	29	21	0.835 ± 0.054	0.00187
	large-size	33	23	17	**0.903 ± 0.037**	**0.00212**
Latitudinal distribution	North	28	13	13	0.815 ± 0.072	0.00143
	Central	49	32	24	**0.884 ± 0.036**	**0.00219**
	South	46	31	22	0.753 ± 0.070	0.00197
Longitudinal distribution	Coastal	69	30	30	0.769 ± 0.055	0.00166
	Oceanic	49	35	24	**0.871 ± 0.043**	**0.00227**

Table 3. AMOVA for different groups evaluation of *Dosidicus gigas*, based on mtDNA ND2 gene analysis. (* indicates values with p<0.05)

Source of variation	Percentage of variation		
	North vs Central vs South	Coastal vs Oceanic	Small vs Medium vs Large
Among groups	0.43	0.21	-0.05
Among population within groups	3.87	1.36	1.51
Within groups	95.7	98.43	95.54
Fixation indices Fsc	**0.03885***	0.01363	0.01505
Fst	**0.04298***	0.01571	0.0146
Fct	0.00429	0.0021	-0.00046

Table 4. Pairwise Fst analysis for different groups of *Dosidicus gigas*, based on mtDNA ND2 gene analysis. Groups were organized according their latitudinal and longitudinal distribution. Bold numbers indicate significant Fst.

Pairwise comparison		n	North		Central		South
			oceanic	coastal	oceanic	coastal	coastal
North	oceanic	10					
	coastal	18	0.00891				
Central	oceanic	16	0.05777	0.02273			
	coastal	31	0.00913	-0.01925	0.01432		
South	coastal	23	0.01961	-0.00636	**0.0701**	-0.00984	
	oceanic	23	0.0084	0.00165	**0.0407**	-0.00029	0.01209

- **ddRAD genotyping**

The total raw reads varied between 3.6 and 18.6 millions of reads and all samples possess an adequate Q30% score. Also, the GC content was similar (36%). The best combination of values was m3, M1 and n2, and considered for the *de novo* assembly with all samples. After removing the "bad apples", 28 samples were used for the catalog construction in STACKS. After applying strict MAF and maximum observed heterozygosity filters, a total of 310 polymorphic loci and 746 SNPs were retained. Central-Oceanic group showed the highest nucleotide diversity (Table 5). PCA (PC1 and 2 with 9.9% of variance explained) analysis using 746 SNPs showed all samples grouped, except from the central oceanic organisms (Fig. 1).

Table 5. Basic population statistics, using 746 SNPs obtained from northern, central and southern of Peru.

Sampled area		n	Private	Obs_Het	Exp_Het	Pi	Fis
North	*Coastal*	6	106	0.12172	0.21655	0.23917	0.29978
	Oceanic	6	93	0.12445	0.21479	0.2377	0.29302
Central	*Coastal*	5	35	0.11137	0.23927	0.27155	0.35277
	Oceanic	3	72	**0.15166**	**0.26314**	**0.31577**	0.30592
South	*Coastal*	8	71	0.14117	0.2415	0.25928	0.34752

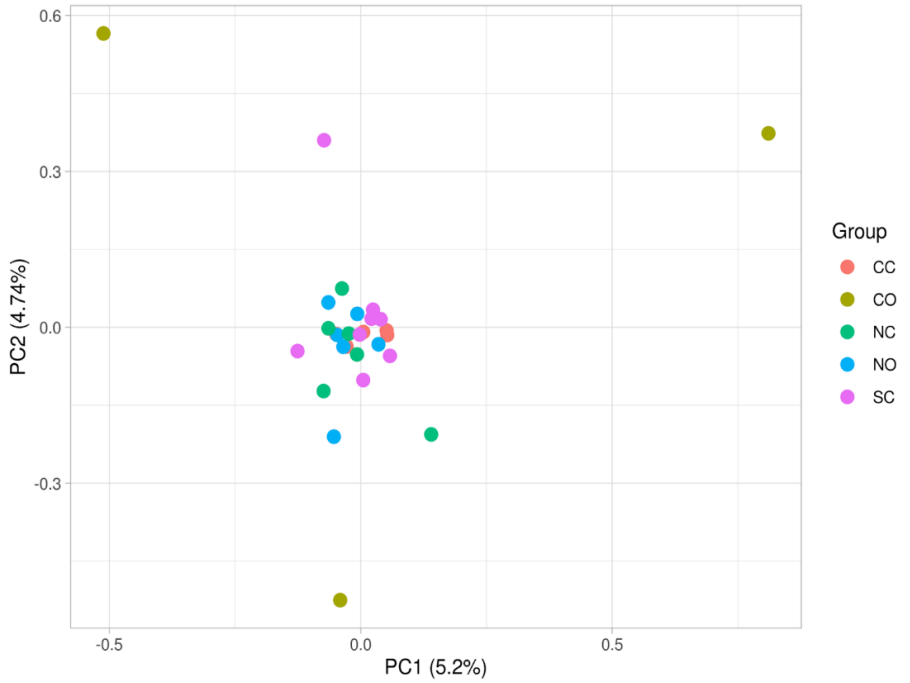

Figure 1. PCA using 746 SNPs obtained from central coastal (CO), central oceanic (CO), northern coastal (NC), northern oceanic (NO), southern coastal (SC) organisms.

4. Activities in progress

- Sampling activities are still ongoing in order to obtain more samples, mainly of large-size organisms from north of Peru.
- SNPs analysis with ddRAD-seq technique for all samples collected, are being evaluated.
- Environmental parameters will be considered for a better interpretation of possible genetic differentiations in various spatio- temporal scales.
- mtDNA COI and ND2 gene sequences are being registered in GenBank public database, according to the SPRFMO SC9 2021 meeting recommendation.

5. References

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

Catchen J,Hohenlohe P, Bassham S, Amores A, and Cresko W. *Stacks: an analysis tool set for population genomics*. Molecular Ecology. 2013.

Cerca, J., Maurstad, M. F., Rochette, N., Rivera-Colón, A., Rayamajhi, N., Catchen, J., & Struck, T. (2020, August 30). Removing the bad apples: a simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. https://doi.org/10.32942/osf.io/47tka

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.

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

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360-1373. https://doi.org/10.1111/2041-210X.12775

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.

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.