SPRFMO
South Pacific Regional Fisheries Management Organisation

7th MEETING OF THE SCIENTIFIC COMMITTEE

La Havana, Cuba, 7 to 12 October 2019

SC7-SQ11

RAD sequencing technology for the evaluation of the population genome variability of the Jumbo flying squid in Peruvian waters

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
2nd Squid Workshop and 7th Meeting of the Scientific Committee
La Havana, Cuba, 5-12 October 2019

RAD sequencing technology for the evaluation of the population genome variability of the jumbo flying squid *Dosidicus gigas* in Peruvian waters

by

Giovanna Sotil and Deivis Cueva
Instituto del Mar del Perú (IMARPE)
Esquina Gamarra y General Valle s/n, Callao, Perú

2019

SUMMARY

The jumbo flying squid *Dosidicus gigas* exhibits a complex structure along its distribution in the Pacific Ocean. Different groups of individuals showing contrasting size, growth rate and sexual maturation time have been reported, and great interest exists in knowing the spatial and temporal distribution of their genetic variability related to these phenotypic groups. Although genetic studies have been carried out on this species, only few loci have been successfully evaluated. Even more, the difficulty of generating genomic information in non-model species, like the jumbo flying squid, with large and complex genomes is well known; so, several markers should be considered to strengthen population structure inferences. Increasingly population genetic studies are using RAD sequencing because it is considered a more sensitive tool that could elucidate population patterns in organisms with dispersal potential and high connectivity among distant populations. In this sense, we have elaborated an instructive guide for tissue sampling and proposed the use of ddRAD-seq method to perform population genetic analysis of *D. gigas* across its latitudinal and longitudinal distribution in the Peruvian jurisdictional waters.

TABLE OF CCONTENT

1. Introduction

Jumbo flying squid *Dosidicus gigas* exhibits a complex population structure along its ample distribution in the eastern Pacific Ocean, which is expressed by the presence of three phenotypic groups or population units, occurring in somewhat different times and areas. There are rather consistent and noticeable differences in the maximum size, size at maturity, growth rate and general distribution areas of these three phenotypic groups, and all three can occur and are found alternating their dominance in the northern part of the Peruvian Current ecosystem. It is expected that the application of the methodology briefly described herewith will contribute to elucidate whether the various population units of jumbo flying squid that are so different in their growth, size at maturity, general distribution areas and environmental preferences, are genetically different or are just the result of a phenotypic differentiation in response to varying environmental conditions somewhere during the early life-history development stages.

Even when capacity of organisms to disperse contributes to their variability, sometimes the knowledge of their diversity is limited by the resolution of some traditional molecular markers. Since the introduction of molecular techniques in the study of population structures, like microsatellites and mitochondrial DNA (mtDNA) sequences, they has complemented information obtained with traditional methods and revealed other peculiarities of population dynamics not evidenced previously (Hauser and Carvalho 2008). Phylogeography, demographic history, population structure, and recently local adaptation are some of the most studied aspects using molecular markers, all of them of great importance in evolutionary biology, ecology, conservation and management of natural populations. However, as with any methodology, molecular markers have advantages and disadvantages when compared to each other. Experts often recommend the use of different molecular markers that provide different but complementary evolutionary information.

One of the limits for non-model species is the lack of genome information which is costly for some species due to their large genomes (Wang *et al.* 2012), so studies in population genetics have been limited to a small number of loci, which may not reflect the diversity. The developing molecular techniques from the first to the next generation sequencing (NGS) also called high throughput sequencing, has conducted to the possibility of doing analysis changing from the use of a small part of the DNA to identifying differentiation in single nucleotides across the whole genome, of a great number of individuals from any specie. This single-nucleotide polymorphisms (SNPs) are one of the genetic markers used for ecological and evolutionary studies.

2. NGS technology: RAD Sequencing for population genome analysis

In general, NGS technologies give several advantages for population genome studies in any species, increasing the sequencing throughput and in consequence information across genome, with low costs, but demanding development of bioinformatics' tools.

The Restriction-Site Associated DNA Sequencing (RAD-seq) has quickly become a cutting-edge method of NGS technology, applied for identification of polymorphic markers along all the genome, genotyping, the construction of genetic maps, population genomics analysis, and their possibility to be associated with phenotypic differences, among others. This technique involves the treatment of the DNA to restriction enzymes

(RE), to generate libraries of DNA fragments representative of the genome of an organism. Different RAD sequencing protocols have been developed like RAD-seq (Baird *et al.* 2008), ddRAD-seq (Peterson *et al.* 2012), 2b-RAD (Wang *et al.* 2012), and have been proven to be applicable on non-model species as they permit to register SNPs of multiplexed organisms, along their genome without knowing an allele-specific expression. DaCosta & Sorenson (2012) suggest that ddRAD-seq gives advantages due to the use of two enzymes that cut less frequently than others RAD-seq methods and the formation of larger fragments sizes of DNA, generating robust genotypic data for a relatively small fraction of the genome, obtaining thousands of loci, allowing increased multiplexing and reducing costs per sample, for analyses of population structure and gene flow.

Studies in population genomics is reveling previously unidentified structure of species and explore adaptive divergence, in wild populations (Larson et al 2013). For example, Rodriguez-Espeleta et al (2016) evaluated RAD-seq data obtained from Atlantic mackerel *Scomber scombrus* populations to determine the sensitivity of the technique to spatial population structure associated to hierarchical SNP selection, and they proposed its application for population structure analysis of species even with low intraspecific genetic differentiation. Larson *et al*. (2013) genotyped Chinook salmon *Oncorhynchus thsawytscha* using RAD sequencing, identifying 10000 SNPs in 270 individuals from five populations of western Alaska, and concluded that theses markers were good candidates for genetic monitoring and population assignment, reveling a fine-scale structure between regions, and also mentioned the possibility of its application in species with large population size and shallow structure. Pfaller et al (2019), comparing RAD-seq and traditional mitochondrial DNA sequencing (COI gene sequence analysis) of *Planes* crabs, demonstrated the limitations of single gene analysis and the value of genomic level resolution for estimating contemporary population structure in organisms with large and highly connected populations. Because these NGS tools are more sensitive, they could elucidate population patterns in organisms with dispersal potential and high connectivity among distant populations.

3. Molecular markers for the evaluation of population genetics of *Dosidicus gigas*

3.1. mtDNA and microsatellites

For the jumbo flying squid *D. gigas*, genetic studies have been carried out using mainly mtDNA (Sandoval-Castellanos *et al.* 2009; Sandoval-Castellanos et al. 2010, Staaf *et al.* 2010; Ibáñez *et al.* 2011; Ibáñez and Poulin 2014) and to a lesser extent microsatellite markers of the nuclear genome (Liu *et al.* 2014; Millán-Márquez *et al.* 2015; Sánchez *et al.* 2016). The main question addressed has been the population structure of the resource, motivated by the regional management required throughout the Pacific. Nevertheless, aspects such as demographic history have been described, suggesting a shallow population history defined by low levels of genetic diversity and sequence divergence (Sánchez *et al.* 2016).

Most markers used in *D. gigas* have evaluated the genetic variation harbored in different regions of the mitochondrial genome. On the other hand, microsatellites proposes a multilocus approach (different parts of the genome used to evaluate genetic variation); however, very few have been successfully achieved for *D. gigas*. Currently, there are few

published studies that have evaluated population genetic variation in *D. gigas* with microsatellites. So, Liu *et al.* (2014) and Millán-Márquez *et al.* (2015) isolated and characterized microsatellites for later use in population studies. Sánchez *et al.* (2016) used microsatellites to evaluate signs of genetic differentiation associated with a longitudinal and latitudinal distribution, and a segregation by size in front of Peru, however, a large part of the selected markers showed a high probability of the presence of null alleles and significant deviations from Hardy Weinberg's equilibrium.

3.2. ddRAD-Seq for the evaluation of population genomic variability of *Dosidicus gigas* in Peruvian waters

In spite of the great importance of the jumbo flying squid, still very few studies have been carried out, that can determine a significant number of useful molecular markers to be applied for monitoring fluctuations of the population genetic dynamics. Therefore, it is convenient to explore patterns of genetic variation with other markers in order to strengthen population structure inferences.

In this sense, IMARPE is implementing the ddRAD-Seq (double digest Restriction-Site Associated DNA sequencing) method to perform population genetic analysis of *D. gigas* across its latitudinal and longitudinal distribution in the Peruvian jurisdictional waters. During the Scientific Research Cruises of jumbo flying squid, specimens were capture and muscle tissues sampled, from three areas: north (from Paita to Pimentel), center (from Huacho to Pisco) and south (from Chala to Ilo) of Peru, between 50 and 200 mn, following the procedure detailed in document "Instructions for the collection of jumbo flying squid *Dosidicus gigas* muscle tissue for molecular analysis" (Sotil *et al.* 2019).

Once the DNA has been extracted and its characterization in degradation and purity has been carried out, DNA of all specimens will be processed following the methodology proposed by DaCosta & Sorenson (2014). In this sense, DNA will be fragmented through the use of two RE, producing a double RE digestion, with specific cutting sites throughout the entire genome of an organism. Following RE digestion, ligation of DNA fragments with double-stranded sequencing adapters (short and known DNA sequences of barcodes / indices that bind to DNA, and that permits further to discriminate each sample) will be done to create the libraries of DNA fragments. Actually, one of the advantages of ddRAD-seq is the simpler library preparation process comparing with other RAD techniques. After a size selection of the genomic fragments produced (300-400 bp approx), a library amplification via PCR, and the purification of the amplified libraries will be performed. Finally, due to the use of different sequences of barcodes / indices ligated for each DNA sample, all samples could be pooled into a single library for NGS sequencing.

Genomic fragments of 178-28 bp (after excluding adapter sequences) derived from thousands of regions genome-wide will be analyzed bioinformatically and used for the evaluation of population variability, searching for thousands of SNPs along different sequence fragments between organisms. Additionally, to each organism biological and biometric data could be related to the genetic variability, to evaluate a possible size structure association. Also our laboratory will evaluate microsatellites and compare with the information generated with NGS, in order to select a set of molecular markers that can be applied for traceability studies.

Even when genotyping with NGS technologies is increasing due to the easier and less expensive population genetic analysis demands, however it is very important to consider some aspects, like the development of bioinformatics tools for analysis of high-density information, and the quality of the archived samples and in consequence DNA fragmentation, which may affect the throughput in quality sequence reads generated in ddRAD datasets (Maroso *et al.* 2018). The information we can generate will allow the creation of genetic databases that can be used for monitoring studies and determination of the states of the resources under study, and the evaluation of the impact that may have the product of fishing, El Niño events, climate change, etc.

4. References

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.

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

Hauser L & Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries, 9*(4), 333-362.

Ibáñez CM, Cubillos LA, Tafur R, Argüelles J, Yamashiro C, Poulin E (2011) Genetic diversity and demographic history of Dosidicus gigas (Cephalopoda: Ommastrephidae) in the Humboldt Current System. Marine Ecology Progress Series, 431, 163-171.

Ibáñez CM, & Poulin E (2014) Genetic structure and diversity of squids with contrasting life histories in the Humboldt Current System. *Hidrobiologica, 24*(1), 1-10.

Larson W A, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol Appl, 7: 355-369.

Liu L, Chen X, Xu Q, Tian S, Liu B (2014) Development of 39 polymorphic microsatellite markers for the jumbo squid *Dosidicus gigas* using Illumina paired-end sequencing. *Conservation Genet Resour.* doi 10.1007/s12686-014-0179-7.

Maroso F, Hillen JEJ, Pardo BG, Gkagkavouzis K, Coscia I, Hermida M, Franch R, Hellemans B, Van Houdt J, Simionati B, Taggart JB, Nielsen EE, Maes G, Ciavaglia SA, Webster LMI, Volckaert FAM, Martinez P, Bargelloni L, Ogden R (2018) Performance and precision of double digestion RAD (ddRAD) genotyping in large multiplexed datasets of marine fish species. Marine Genomics,39, 64-72.

Millán-Márquez AM, Salinas-Zavala C, Cruz P, Pearse DE (2015) Isolation and characterization of microsatellites for jumbo squid Dosidicus gigas (Ommastrephidae). *Journal of genetics, 94*(1), 38-40.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135.

Pfaller JB, Payton AC, Bjorndal KA, Bolten AB, McDaniel SF(2019) Hitchhiking the high seas: Global genomics of rafting crabs. *Ecol Evol*. 9: 957– 974.

Rodríguez-Ezpeleta N, Bradbury IR, Mendibil I, Álvarez P, Cotano U, Irigoien X (2016) Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. Mol Ecol Resour, 16: 991-1001.

Sanchez G, Tomano S, Yamashiro C, Fujita R, Wakabayashi T, Sakai M, Umino T (2016) Population genetics of the jumbo squid Dosidicus gigas (Cephalopoda: Ommastrephidae) in the northern Humboldt Current system based on mitochondrial and microsatellite DNA markers. *Fisheries research, 175*, 1-9.

Sandoval-Castellanos E, Uribe-Alcocer M, Díaz-Jaimes P (2009) Lack of genetic differentiation among size groups of jumbo squid (Dosidicus gigas). Ciencias Marinas, 35(4), 419-428.

Sandoval-Castellanos E, Uribe-Alcocer M, Díaz-Jaimes P (2010) Population genetic structure of the Humboldt squid (Dosidicus gigas d'Orbigny, 1835) inferred by mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology, 385*(1-2), 73-78.

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-Doc. xxx(TBA): 7p

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.

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* **9**, 808.