Investigating the genetic diversity, population structure and demography of Baja and Pacific *Hemisquilla californiensis*

Ethan Holleman

Introduction

Mantis shrimp are a unique group of carnivorous stomatopods that are believed to have diverged from members of the class *Malacostraca* around 340 million years ago [1]. Mantis shrimp are notable for both their unique mechanically driven raptorial smashing or spearing claws, which are powerful enough to induce cavitation in the surrounding water [2, 3] and acutely sensitive visual system which is believed to be the most complex ever discovered in nature [4, 5]. Due to their extraordinary physiology, the biology of Mantis shrimp species has been the subject interdisciplinary study by researchers in the fields of optics, material science and biology [6, 7, 8]. However, comparatively few studies have examined on the genetic structure of Mantis shrimp populations and those that have focused on Asian Mantis shrimp species in the Yellow and East China Seas [9] where their study is motivated by the greater demand for the shrimp as a specialty seafood in Asian economies. *Hemisquilla californiensis*, the California Mantis shrimp, which ranges from Southern California to Panama [10], therefore represents a significant gap in population level scientific knowledge of these incredible organisms. *H. californiensis* and other Mantis shrimp species are known to play an important role in marine ecosystems by acting as efficient predators of other crustaceans and oxygenating ocean sediments [11]. Understanding the genetic history of these crustaceans in their Atlantic range is therefore key for accessing the overall health of important marine ecosystems, such as coral reefs which *H. californiensis* are known to inhabit.

Figure 1: Proposed *Hemisquilla californiensis* collection locations shown as colored dots. Image adapted from *Basch et al, 93*

Therefore, our main goal is to quantify genetic diversity of *H. californiensis* across their Atlantic range in order to access gene flow between populations and infer effects of recent antropogenic changes to the marine ecosystem by evaluating *H. californiensis* demographic and selective events. Towards this effort we will pursue three specific aims.

Aim 1: Characterize genetic variation in *H. californiensis* populations.

Hypothesis: All *H. californiensis* populations will deviate significantly from Hardy-Weinburg equilibrium and correlate with latitude. *H. californiensis* populations are mainly concentrated along the Southern coast of Baja California, with a fewer more Southerly populations identified along the West coast of Panama, although this may be a result of lower resolution reporting [10]. In order to access this we will genotype individuals from 12 Western Baja populations and 2 Western Panamanian populations using restriction site associated DNA sequencing (RAD-seq) [12] which will allow the low-cost genotyping of a large number of individuals. Using RAD-seq data we will identify a number of suitable sites for further genetic analysis based on sequence quality metrics. We will also identify sites my mapping RAD-seq reads to homologous regions in the closest high-quality related reference genome; Penaeid shrimp [13]. Any calculations made using RAD-seq data ultimately assume that individual reads are independent, an assumption that can be violated by PCR induced bias. To avoid erroneous calculations we will employ protocols that produce random fragment lengths in order to trim PCR duplicates *in silico*. We will then use this trimmed genotype data to calculate allele frequencies for all identified sites across the 14 sampled populations
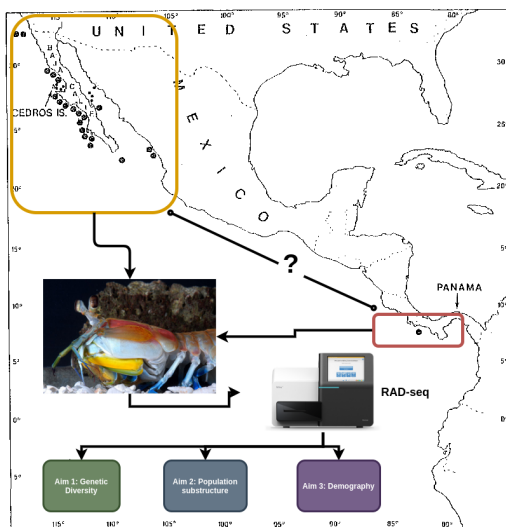
and determine genotypes of each individual using established methods [14]. Expected genotype frequencies can then be calculated for each site under the assumptions of Hardy-Weinberg Equilibrium (HWE); random mating and absence of migration between populations, genetic drift and mutation. HHE derived expectations will then be compared to observed genotypes utilizing Chi square tests. This test of significance assumes variables are mutually exclusive; this is appropriate as any two genotypes cannot occur simultaneously. The degree to which each site deviates from HHE expectation can then be compared across each population.

Aim 2: Quantify *H. californiensis* population sub-structure to evaluate gene flow

Hypothesis: *Odontodactylus scyllarus* will show significant gene flow within, but not between, Baja and Panamanian sampling locations. Previous studies examining genetic variation of *Oratosquilla oratoria*, a Mantis shrimp species closely related to the California Mantis shrimp, in the Yellow and East China seas found a surprising degree of gene flow between populations as determined by pairwise F_{st} calculations [9]. This represents occurrence of gene flow at a scale similar to that of the most Northern and Sourthern Baja *H. californiensis* populations but significantly less than distances between Baja and Panamanian populations between which no significant population centers have been identified [10]. However, these studies suffered from lack of high resolution sampling at mid latitudes and therefore gene flow could be occurring if significant numbers of *O. scyllarus* exist along the Western Mexican coast. To answer this question we will utilize the RAD-seq genotype data collected in aim 1 to access gene flow between the 14 sampled populations through pairwise comparison of inbreeding F statistics, and through a Bayesian approach to sub-population assignment. First, we will determine reductions in heterozygosity due to genetic differentiation between the sampled populations by calculating the pairwise F_{st} values between the 14 sampled populations. Ultimately, F_{st} is evaluated against the expected heterozygosity of the sub-populations if they were merged, had equal size and were randomly mating. We will evaluate error in our estimations through permutation testing, which repeatedly randomly assigns individuals to a sub-population and evaluates F_{st} to ultimately create a probability distribution which can then be used to evaluate the significance of our actual results. We expect pairwise F_{st} to deviate significantly from expectation between proximal Baja populations but to be non-significant for all pairwise comparisons that include the Panamanian populations. Additionally, using the pairwise F_{st} calculations between Baja populations, we will determine the mean decrease in F_{st} per unit distance between populations. If no major *O. scyllarus* exist between the Baja and Panamanian regions, this expectation should be predictive of the degree of decrease in pairwise F_{st} between Baja and Panamanian populations. A decrease that is significantly less than this expectation would indicate greater gene flow between these populations than would be expected, indicating the presence of mid-latitude *H. californiensis* populations. This however assumes that geographic distance is the main determinant of gene flow in *H. californiensis* which is reasonable considering their relatively stationary lifestyle [15]. Next, we will also preform an admixture analysis in order to estimate the proportion of ancestry from each subpopulation for all individuals. Similarly to pairwise F_{st} analysis we expect to find that significant admixture between but not within Baja and Panamanian populations. While out analysis will consider ancestry from multiple populations it ultimately relies on the same assumptions made under two-population maximum likelihood admixture model. Namely, that all sites are independent (not experiencing linkage disequilibrium) and mutually exclusive. We will quantify error in our estimations by calculating 95% confidence interval derived from maximum likelihood distributions that consider θ (percent ancestry) in steps of 0.01. Finally, we will visually evaluate sub-population structure using principal component analysis of all genotyped sites. We expect to find significant separation between Baja and Panamanian populations with some, but lower degree of sub-clustering of Baja populations that is reflects geographic separation.

Aim 3: Determine demographic history between Northern and Southern California Mantis shrimp populations.

Hypothesis: *H. californiensis* has undergone a recent global decrease in population size across its range in the last 200 years. It is estimated that ocean pH levels have decreased approximately 30% in the last 200 years [16]. This has been shown to be especially detrimental to species that rely on an calcium based exoskeleton, such as crustaceans including *H. californiensis* [17]. In order to access if recent decreases in ocean pH have impacted *H. californiensis* we will utilize approximate Bayesian computation (ABC) to infer both the most likely timing and magnitude of a potential population decrease based on allele frequency data generated in the course of aim 1. Since we expect an overall decline in *H. californiensis* populations but lack significant data regarding their population history we will utilize a bounded uniform prior which assumes equal probability of all parameters, time and strength of decline, within a discrete range. Our parameter space will be bounded between 10 and 50,000 generations in time and between a 0 and 100 fold change in population size (strength of decline). This will

allow for consideration of both recent (anthropogenic ocean acidification) and historical demographic events. The most probable parameters to explain *H. californiensis* observed allele frequency spectrum will be identified by measuring euclidean distances to the simulated allele frequency spectra. Due to how recently ocean acidification has occurred, we would not expect significant selection to have occurred. This would lead to changes in the allele frequency spectrum at specific loci opposed to genome wide changes expected by a demographic event. We will therefore compare the allele frequency spectrum of specific loci to genome-wide background to validate acidification as a demographic level perturbation. Finally, we will compare ABC results across all sampled populations. Since ocean acidification should effect all *H. californiensis* populations similarly, we would not expect to see significantly differentiated time and strength of selection parameters across populations. Finally, to directly connect our population level results to *H. californiensis* physiology, during sample collection we will measure calcium content of exoskeletons from a select number of individuals from all populations using Inductivity Coupled Plasma-Optical Emission Spectrometry Instruments (ICP-OES) and compare to measurements from historical Mantis shrimp samples. We expect to observe an overall decrease in calcium content.

References

[1] Cara Van Der Wal, Shane T. Ahyong, Simon Y.W. Ho, and Nathan Lo. The evolutionary history of stomatopoda (crustacea: Malacostraca) inferred from molecular data. *PeerJ*, 5:e3844, September 2017.

[2] S. N. Patek, W. L. Korff, and R. L. Caldwell. Deadly strike mechanism of a mantis shrimp. *Nature*, 428(6985):819–820, April 2004.

[3] S. N. Patek and R. L. Caldwell. Extreme impact and cavitation forces of a biological hammer: strike forces of the peacock mantis shrimp odontodactylus scyllarus. *Journal of Experimental Biology*, 208(19):3655–3664, October 2005.

[4] Thomas W. Cronin, Michael J. Bok, N. Justin Marshall, and Roy L. Caldwell. Filtering and polychromatic vision in mantis shrimps: themes in visible and ultraviolet vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1636):20130032, February 2014.

[5] Susan Milius. Life: Mantis shrimp flub color vision test: Results suggest alternate way of detecting wavelengths. *Science News*, 182(6):11–11, September 2012.

[6] Wei Huang, Mehdi Shishehbor, Nicolás Guarín-Zapata, Nathan D. Kirchhofer, Jason Li, Luz Cruz, Taifeng Wang, Sanjit Bhowmick, Douglas Stauffer, Praveena Manimunda, Krassimir N. Bozhilov, Roy Caldwell, Pablo Zavattieri, and David Kisailus. A natural impact-resistant bicontinuous composite nanoparticle coating. *Nature Materials*, 19(11):1236–1243, August 2020.

[7] Ali Altaqui, Pratik Sen, Harry Schrickx, Jeromy Rech, Jin-Woo Lee, Michael Escuti, Wei You, Bumjoon J. Kim, Robert Kolbas, Brendan T. O'Connor, and Michael Kudenov. Mantis shrimp–inspired organic photodetector for simultaneous hyperspectral and polarimetric imaging. *Science Advances*, 7(10):eabe3196, March 2021.

[8] Mary W. Donohue, Jonathan H. Cohen, and Thomas W. Cronin. Cerebral photoreception in mantis shrimp. *Scientific Reports*, 8(1), June 2018.

[9] Mei Yang and Xinzheng Li. Population genetic structure of the mantis shrimp oratosquilla oratoria (crustacea: Squillidae) in the yellow sea and east china sea. *Journal of Oceanology and Limnology*, 36(3):905–912, May 2018.

[10] L. V. Basch and J. Engle. Biogeography of hemisquilla ensigera californiensis (crustacea: Stomatopoda) with emphasis on southern california bight populations. 1993.

[11] P.J. Antony, S. Dhanya, P.S. Lyla, B.M. Kurup, and S. Ajmal Khan. Ecological role of stomatopods (mantis shrimps) and potential impacts of trawling in a marine ecosystem of the southeast coast of india. *Ecological Modelling*, 221(21):2604–2614, October 2010.

[12] RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 10(2):108–108, February 2011.

[13] Xiaojun Zhang, Jianbo Yuan, Yamin Sun, Shihao Li, Yi Gao, Yang Yu, Chengzhang Liu, Quanchao Wang, Xinjia Lv, Xiaoxi Zhang, Ka Yan Ma, Xiaobo Wang, Wenchao Lin, Long Wang, Xueli Zhu, Chengsong Zhang, Jiquan Zhang, Songjun Jin, Kuijie Yu, Jie Kong, Peng Xu, Jack Chen, Hongbin Zhang, Patrick Sorgeloos, Amir Sagi, Acacia Alcivar-Warren, Zhanjiang Liu, Lei Wang, Jue Ruan, Ka Hou Chu, Bin Liu, Fuhua Li, and Jianhai Xiang. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications*, 10(1), January 2019.

[14] Nicolas C Rochette and Julian M Catchen. Deriving genotypes from RAD-seq short-read data using stacks. *Nature Protocols*, 12(12):2640–2659, November 2017.

[15] Marjorie Lindquist Reaka. On learning and living in holes by mantis shrimp. *Animal Behaviour*, 28(1):111–115, February 1980.

[16] Ocean acidification | National Oceanic and Atmospheric Administration.

[17] Jennifer R. A. Taylor, Jasmine M. Gilleard, Michael C. Allen, and Dimitri D. Deheyn. Effects of CO2-induced pH reduction on the exoskeleton structure and biophotonic properties of the shrimp lysmata californica. *Scientific Reports*, 5(1), June 2015.