# SCAR Fellowship Report

## Seascape genomics in the cold

**David L. J. Vendrami**
Postdoctoral researcher
Department of Animal Behaviour, Bielefeld University
Germany
david.vendrami@uni-bielefeld.de

**Home Institution:** Department of Animal Behaviour, Bielefeld University
Prof. Dr. Joseph I. Hoffman, joseph.hoffman@uni-bielefeld.de, Germany

**Host Institution:** British Antarctic Survey
Dr. Peter Fretwell, ptf@bas.ac.uk, UK

**Dates of Activity 01/05/2022 – 31/08/2022**

## Introduction

Antarctica is one of the areas of our planet that is most affected by ongoing climate change. One of the key aspects determining the persistence of species under such changing conditions is represented by the ability of a species to disperse, as this can mediate the potential for adaptive genetic variation to spread among populations. Understanding how organisms disperse and which factors affect their ability to do so is therefore of key importance, especially in the marine environment where unforeseen complexities may exist.

## Project Objectives

The objective of this project is to combine genomic and oceanographic data to better understand how marine invertebrates with direct development can disperse. This was carried out using the Antarctic sea snail *Margarella antarctica* as a "seascape" model species. This approach allows us to better understand which specific characteristics of the physical environment influence dispersal among populations. This provides key information for predicting the adaptive potential of marine invertebrates with reduced mobility under ongoing climate change.

## Methods, Execution and Results

To fulfill the scope of this project, we used over 600 samples of *Margarella antarctica* collected within four different geographic scales, ranging from a macrogeographic scale, consisting of samples collected along the entire Antarctic peninsula, to a nanogeographic scale, consisting of samples collected within grids with an area of 25 $m^2$. Genomic data, namely RAD sequencing data, for all of these samples were already available as part of one of my previous projects and were analyzed to quantify the extent of gene flow between sampling locations within the study area, separately for each geographic scale. Specifically, given the absence of a *M. antarctica* reference genome, our sequencing reads were assembled *de novo* into a set of reference RAD contigs, which were then used as a baseline for SNP calling. The resulting SNP genotypes were then filtered according to different criteria, including depth of coverage, minor allele frequency, rate of missing data, deviations from Hardy-Weinberg equilibrium expectations and linkage disequilibrium. We generated a number of different genotype datasets based on different thresholds, so that we could study the impact of different quality filtering procedures on the inference of gene flow. Subsequently, we aimed at integrating the resulting gene flow estimates into a geospatial context. To do so, I first had to extract a large amount of environmental and oceanographical data from available databases. These included, but were not limited to, data on bathymetry, substrate type, current circulation patterns and environmental data such as mean annual sea surface temperature. These data were then used in combination with the known geographical distances among sampling sites to build distance matrices that describe the cost associated with movement between any two specific sampling locations within our study area. The combination of these distance matrices with inferred measures of gene flow via geospatial modelling should allow inference of how different environmental and oceanographical features affect dispersal.

## Project Outcomes

The application of the bioinformatics workflow described above allowed me to obtain a dataset consisting of 2,388,010 raw SNPs. The application of different sets of filtering criteria resulted in 16 different quality filtered datasets that contained a number of SNPs ranging from 17,485 to 37,157. Estimates of gene flow were similar across all 16 datasets and we therefore decided to focus on the most stringently filtered dataset, as we believe this should minimize noise. Next, I started to extract relevant environmental and oceanographic data to build distance matrices. This procedure was more challenging than I had expected because it involved the extraction of a very large amount of information from different databases and an appreciable amount of manual curation of the resulting dataset. I have therefore only recently been able to start the process of practically integrating genomic estimates of gene flow in a geospatial context. Specifically, I have focused initially on two of the four geographic scales included in this project. By now, I feel comfortable with this procedure and I am currently in the process of analyzing the two remaining geographic scales. Thus, this project is still ongoing.

## Publications, Presentations and Products

Further analyses are required before this project can be translated into a scientific publication. We aim to produce a single manuscript for a leading field or international journal. Preliminary results of the current work will soon be presented at an international conference. We will also deposit the genomic data in a public repository at the time of publication. This includes RAD sequencing data for over 600 *M. antarctica* samples, which represents the very first RADseq dataset ever published for this species. Finally, all of the code utilized during my analysis will be made publicly available. In addition to making my analyses fully reproducible, this will provide a useful resource for other researchers aiming at reproducing our seascape genomics analysis.

## Capacity Building, Education and Outreach Activities

This collaboration allowed me to significantly expand my research network, especially in the context of Antarctic research. I look at this as being a remarkably successful outcome of this collaboration. Many interesting discussions took pace, which have contributed significantly to my development as an Antarctic scientist. Next, I had the opportunity to share my experience as a SCAR fellow with many students and colleagues, who all showed a great interest in my work. This included also peers who are not involved in Antarctic research and were not aware of SCAR. Finally, I am currently in the process of organizing a series of dissemination events during which I plan to “narrate this Antarctic story” to elementary school pupils. I plan to do so within my homeland, Italy, and I hope to stimulate the curiosity of many young students towards Antarctica and, more generally, towards science.

## Future Plans and Follow-ups

My host and I are still actively working on this project. Since I am still in the process of analyzing the data, we intend to continue our collaboration after the end of my SCAR fellowship. Given the positive experience we both had, we believe there may be scope

for additional collaborations in the future. Furthermore, the skills I acquired during this collaboration will be of great use for my next projects, so I am confident that I will have the opportunity to apply/transfer what I learnt during this collaboration to many new future projects.

## Personal Impact

I am a molecular egologist by education and I am proficient in generating and analyzing genetic and genomic data. However, I realized that this is not enough if I want to fulfill my goal of becoming an expert in seascape genomics. Through this collaboration, I acquired fundamental skills in geospatial analysis which are indispensable for a successful career in this field. This collaboration also benefitted my scientific network, which clearly expanded and now includes some of the maximum experts in Antarctic research.

## Financial Statement

The money provided by the fellowship were used exclusively to sustain my own leaving expenses, which includes costs of accommodation, health insurance and groceries.

## Acknowledgements and References

I would like to thank Prof. Dr. Joseph I. Hoffman for having informed me about the possibility to apply for this scholarship and for the help he provided during this project. I also would like to thank my host Dr. Peter Fretwell for having greatly supported me throughout our collaboration.