Research Networking Programmes

Short Visit Scientific Report

Proposal Title: *Treatment of Rad seq data and discovery of SNPs for seascape genomics*

Reference 6747

Applicant: Adriana Villamor adriana.villamor@unibo.it

Host Institution: Hellenic Center of Marine Research, Crete Greece

Host Researcher: Tereza Manousaki tereza@hcmr.gr

1- Purpose of the visit:

The aim of this short visit was to learn in a practical way some principles on the necessary bioinformatic tools, necessary to properly analize large amount of genetic data. These skills became needed in the frame of a 3-year project aimed to identify new polymorphic markers in several marine invertebrate species for connectivity studies accros marine areas at different spatial scales. In particular I focus my efforts in three Mediterranean cnidarians, widespread and with high interest as they are found in key habitats: *Corallium rubrum*, the Mediterranean red coral, is a widespread species, with very specific habitat requirements, but endangered due to its commercial value; *Parazoanthus axinellae*, found in dense aggregations in vertical walls and caves, is a common epiphyte of several sponge species, and *Eunicella singularis* a gorgonian found in a wide range of depths and substrates. All of them have shown very low sequence variation in the commonly used markers, both mitochondrial and nuclear, so the discovery of new polymorphic markers is needed to properly assess their status of genetic diversity and connectivity between populations of these species, which are important representatives of

mesophotic Mediterranean reefs, a singular and fragile ecosystem of the Mediterranean Sea.

2- Description of the work carried out during the visit:

During the two weeks at the HCMR working with Dr Tereza Manousaki I have learn basic skills of Unix environment, creating, handling files, as well as some basic programming, useful and easy functions to work with large genetic data sets. Using the resulting 2bRAD libraries, prepared for several individuals of each of the three species mentioned, we have investigated the best ways of analyzing these data, how to get the most trustful results, as well as the optimal number of SNPs. Specifically we first focused in the widely used Stacks software. Apart from learning basic usage of the software we had to modify scripts and create some corrections for our sequences, given that 2bRAD procedures are not included in the standard pipeline of Stacks.

After that we focused in a second pipeline, proposed by Eli Meyer, researcher at the Oregon State University who developed a full pipeline specifically for 2bRAD data from organisms for which no reference genome exists. This pipeline required a longer preparation, as is formed by a large set of scripts, for which parameters for the optimization of quality filters and SNPs calling. Furthermore, additional modifications had to be included given that the enzyme used by Meyer is different from the enzyme used in our libraries.

3- Description of the main results obtained:

In these two weeks of work we got to run both pipelines in several ways, with different parameters of quality filtering, SNP calling, minimum depth of reads, etc. for all three species. Thanks to these tests I believe now we are aware of the limits of these new techniques, how results can dramatically change depending on small modifications of criteria. We got to set a final pipeline, mixture of both procedures in which we got the most of both, from one side we used the very restrictive quality filters of Meyer's pipeline, plus the capacity to prepare our 2bRAD sequences for Stacks procedure. After that we chose to run Stacks *denovo_map* procedure, which allows the creation of a denovo reference genome, determination of a catalogue of SNPs, ready to use for

several downstream applications, all this in a more user friendly interface, and obtaining very similar results in terms of numbers of SNPs to those obtained with the more complex scripts-based pipeline.

4- Future collaboration with host institution:

During this visit I have known a fantastic institute with friendly, helpful and enthusiastic people working in it. Tereza Manousaki's expertise and clarity of ideas has been incredibly inspiring, we have worked long hours with great enthusiasm and given this good personal communication between us we have decided to continue our interaction from now on, not only for the present project, but also several ideas for near future projects have arised, involving our institutions. In the next year my institution will invite Dr Manousaki to give some seminar or talk for students and /or researchers.

5- Projected articles to result from the grant:

We expect to submit the first publication resulting from this visit by the end of 2015, however I expect to publish at least another work, hopefully two, related to the specific connectivity patterns and particularities of each of the species. ESF will be properly acknowledged and informed when this works are published and available.