

Report on the outcomes of a Short-Term Scientific Mission¹

Action number: CA18201

Grantee name: Nikolaos Tourvas

Details of the STSM

Title: Use of genomic tools for the management of forest reproductive material under climate change

Start and end date: 02/05/2022 to 30/06/2022

Description of the work carried out during the STSM

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

(max. 500 words)

During this STSM I received training with regards to analysing genomic data sets. Results from an ongoing experiment were kindly provided by the host institution (BOKU), and I was supported into developing a software pipeline for their analysis. In more detail, the work carried out was focused on pooled exome sequencing data. This is a particular type of DNA sequencing experiment in which the DNA of multiple individuals from the same population is used to create one population sample that is afterwards sequenced. This experimental design was chosen as it provides an economically viable way to obtain high statistical power for the exploration of signatures of selection for populations of non-model species. It is therefore suitable as a genomic tool for applications regarding the management of forest reproductive material and the management and conservation of forest genetic resources.

Prior to carrying out this project, I received documentation from similar experiments of the research group to guide me through, while throughout this process members of the group were always accessible to assist me. The software pipeline I developed achieves the following: Firstly, a quality assessment of the raw data is performed with software fastqc. Then a trimming step for the removal of low-quality DNA reads with software trimmomatic is utilized, followed by a further quality assessment of the remaining data. These remaining reads are then mapped to a provided reference genome with the bwa-mem algorithm and the output of this function is used for the detection of variant sites (single nucleotide polymorphisms, insertions, deletions) between samples and the reference genome

¹ This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

(software VarScan). A data file in the VCF format containing all detected variations is produced after filtering steps to ensure their integrity. This VCF file is one of the major outputs of the pipeline as it is a widely used format which can be used as input for many downstream software. In the context of the current analysis two scripts in the programming language R were produced to accompany the rest of the pipeline for the exploration of adaptive genetic diversity. In the first one, the poolstat package is utilized to ingest data in R, perform LD pruning (optionally), calculate Fst and allelic frequencies. The second script is dedicated to the study of loci under selection using the pcadapt package.

To tackle the challenges presented by the use of high performance clusters instead of desktop based data analysis, the full code for the work described above is available online through a github repository https://github.com/nikostourvas/acorn_poolseq_pipeline. Information on how to create a computational environment (necessary software and databases) for using this pipeline is also provided to enhance the reproducibility and reusability of this work.

Finally, based on the results obtained from this exercise, a number of discussions were held within the research group of BOKU regarding proper interpretation, the challenges of these genomic based results and their connection with guidelines for the use of FRM in the context of climate change.

Description of the STSM main achievements and planned follow-up activities

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

(max. 500 words)

After the completion of my training at BOKU, I acquired valuable experience in several aspects concerning experiments to study local adaptation. I had the chance to work with real world data from ongoing experiments and discuss about their conception and implementation. The research group of BOKU provided me with the computational resources which helped me in my transition from desktop-based work to high performance computing data analysis. Thanks to this work, I now have a stronger background in the use of software tools commonly employed for the analysis sequencing data sets. Furthermore, I became acquainted with databases of resources for sequencing experiments.

It was recognized that the work achieved during the STSM could be useful in future projects. To that end, the BOKU research group and I agreed to continue collaboration for the development of this pipeline to improve its efficiency and usability. Furthermore, a container-based version of the pipeline's computational environment will be developed using the Singularity platform to make this work better adhere to the FAIR (Findability, Accessibility, Interoperability, Reusability) principles.