COST STSM 2014

COST ACTION:

FPS COST Action FP1202. Strengthening conservation: a key issue for adaptation of marginal/peripheral populations of forest trees to climate change in Europe (MaP-FGR).

GRANT REFERENCE CODE: COST-STSM-ECOST-STSM-FP1202-290914-044284

STMS TYPE: Regular (from Spain to Scotland).

STSM APLICANT: Ruth C. Martín Sanz

INSTITUTION: CIFOR-INIA, Madrid (Spain).

CONTACT: martin.ruth@inia.es

+34 91 3473904

PERIOD OF STSM: 29/09/2014 to 20/10/2014

HOST: Dr. Stephen Cavers

Centre for Ecology & Hydrology (CEH), Edinburgh (Scotland, United Kingdom).

1. Purpose of the Short Term Scientific Mission.

The main purpose of the Short term Scientific Mission (STSM) was to address the joint analysis of adaptive geographic variation in 17 populations of Aleppo pine (from phenotypic data of different characters obtained in common gardens) with neutral molecular variation in cSSRs. To do this, first of all, we had to investigate the existence of population structure in our data and assign each individual to one of those populations.

2. Description of the work and main results.

First, I gave a talk in the Institute of Evolutionary Biology, University of Edinburgh on the entire PhD project to receive feedback from researchers of the hosting group. In it I presented the overall objectives of my thesis, the work I've done so far and the following issues to deal with. It was addressed to a group of researchers and PhD students belonging to the Centre for Ecology & Hydrology and the University of Edinburgh with whom Dr. Stephen Cavers works.

As a first step, we use GenAlEx to debug our molecular database searching for monomorphic SNPs. In addition, we used it to perform complementary analysis of the data like the calculation of Fst, finding rare SNPs, etc. GenAlEx (Genetic Analysis in Excel) is a cross platform package for population genetic analyses that runs within Microsoft Excel. GenAlEx offers analysis of diploid codominant, haploid and binary genetic loci and DNA sequences. Both frequency based (F statistics, heterozygosity, HWE, population assignment, relatedness) and distance-based (AMOVA, PCoA, Mantel tests, multivariate spatial autocorrelation) analyses are provided. New features include calculation of new estimators of population structure: G'(ST), G''(ST), Jost's D(est) and F'(ST) through AMOVA, Shannon Information analysis, linkage disequilibrium analysis for biallelic data and novel heterogeneity tests for spatial autocorrelation analysis.

Afterwards, we analyzed our molecular data (SNPs), for which I have learned to use the STRUCTURE and CLUMPP programs.
STRUCTURE is a free software package for using multi-locus genotype data to investigate population structure. Its uses include inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating population allele frequencies in situations where many individuals are migrants or admixed.
Once we had cleaned and prepared the SNP database, we ran STRUCTURE and used the results on STRUCTURE HARVESTER, a website and program for visualizing STRUCTURE output and implementing the Evanno method. Finally, we ran CLUMPP, a program that deals with label switching and multimodality problems in population-genetic cluster analyses. CLUMPP permutes the clusters output by independent runs of clustering programs such as structure, so that they match up as closely as possible.
Thus we find a structure in the distribution of Aleppo pine; fully differentiated the populations from eastern and western Europe. Our 17 populations were grouped into 5 different clusters (K = 5).

Finally, we started working with GENELAND to process our phenotypic and molecular data all together. I have begun analyzing molecular data only for

learning to use the program, thus we get the same results as with STRUCTURE, and now we keep in touch for proceeding with the analysis of all data.

GENELAND is a computer program, implemented in R, for statistical analysis of population genetics data. Its main goal is to detect population structure in form of systematic variation of allele frequency that can be detected from departure from Hardy-Weinberg and linkage equilibrium. It implements several models that can make use of both geographic and genetic information to estimate the number of populations in a dataset and delineate their spatial organization. It can also process phenotypic data and therefore any combination of genetic, phenotypic and geographic information.

3. Future collaborations.

The STSM has enhanced collaboration between two different institutions (Forest Research Centre (CIFOR-INIA) and Centre for Ecology & Hydrology (CEH)). This collaboration has been very fruitful for both sides and will go further by continuing to collaborate on the Geneland analysis, in order to establish models and develop a manuscript.

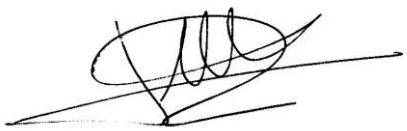
4. Foreseen publications.

The work developed during this STSM will be submitted to a scientific journal as soon as the analysis of the results is completely finished and the manuscript prepared. It will be also a chapter of my PhD thesis.

5. Acknowledgements

I would like to personally thank to COST ACTION FP 2012 for the great professional and personal experience they have enabled me to enjoy. I would also like to thank CEH for endorsing me during the STSM period and principally to Dr. Stephen Cavers as my host in the CEH for all the support and knowledge provided.

Signature



Ruth C. Martín Sanz

Madrid, 06/11/2014