Purpose of the Short Term Scientific Mission

Considering climate change and the associated shifts in species´ distribution tracing their optimum habitat, survival is strongly determined by adaptation and migration phenomena which are taking place within the distribution limits of the species. Climate change should thereby lead the conservation agenda towards establishing guidelines that ensure species survival.

The definition of conservation policies must be based on strict scientific criteria and Species Distribution Models (SDMs) provide a set of methods to illustrate in a spatially explicit framework the effects of climate change in species´ ranges. SDMs calibrate associations between environmental variables and known species´ occurrence records to identify environmental conditions under which viable populations can be maintained (Guisan & Thuiller, 2005). Furthermore, the incorporation of genetic information has been shown to improve models´ predictions (D'Amen, Zimmermann, & Pearman, 2013; Serra-Varela et al.,submitted) and enables to consider more accurately adaptation as a driver of the future range of the species.

In this framework, the Short Term Scientific Mission (STSM) aims to project to future climate scenarios SDMs of two Mediterranean pine species (*Pinus pinaster* and *Pinus halepensis*) integrating molecular information and to define priority conservation guidelines based on the distribution patterns predicted. This approach aims to collect already existing data bases in an effort to identify the dynamics of the genetic spatial structure and diversity in relation to adaptation to climate change

Description of the work carried out during the STSM;

As a first step during the STSM, we spatially defined the genetic groups of *Pinus halepensis* (already demarcated for *Pinus pinaster*). Genetic groups were defined based on the estimated membership coefficients for each individual within each cluster (Q). Q were calculated as a mean-over-population of individual coefficients of membership to each one of the 7 clusters defined by Structure clustering method (Pritchard *et al.*, 2000) on a 294 SNP dataset sampled in 1325 trees from 49 populations. We defined seven homogenous genetic clusters namely Moroccan-Southern Spain clade, Tunisia clade, Israel-Italy clade, Greek clade, North Italian-France clade, Central Spain clade and Northern Spain-Southern France clade. Nevertheless the definition of strict frontiers between the groups was not possible given the fast colonization history of the species from in a west-east direction (Grivet *et al.*, 2009) and some areas were left out of the classification due to the lack of enough sampled molecular information.

Secondly, we calibrated a SDM for each of the species based on individual models of each of their compounding genetic clades. SDMs of a species composed from the infraspecies level (*i.e.* genetic clades composing the whole species) obtain better results than those not integrating this information as we illustrated during the previously enjoyed COST ACTION during 2013 (Serra-Varela *et al.,* submitted). Therefore, we calibrated an individual model for each of the genetic clades (8 models for *P. pinaster* and 7 for *P. halepensis*) and we then built the species´ model superposing the predictions of its individual clades.

Each clade model combined five different statistical methods: as regression methods, we selected General Linear Models (GLM; McCullagh & Nelder, 1989) and Generalized Additive Models (GAM; Hastie & Tibshirani, 1986). Random Forest (RF; Breiman, 2001) represents machine-learning methods while Classification Tree Analysis (CTA; Breiman, 1984) stands for classification methods. Finally, we also used MaxEnt representing a maximum entropy model (Phillips *et al.*, 2006). All models were processed within BIOMOD (Thuiller *et al.*, 2009) using the package "biomod2" in the R statistical software environment (R

Development Core Team, 2013) .Model performance was assessed by True Skill Statistic (TSS) and Area Under the ROC Curve (AUC). We then created an ensemble model for each clade as the mean prediction of the five different algorithms.

As concerning to climate data we considered nineteen bioclimatic variables for the analysis (BIO1 to BIO19) which correspond to the available bioclimatic variables that can be directly downloaded from WORLDCLIM (Hijmans *et al.*, 2005). We selected BIO4 (Temperature Seasonality), BIO11 (Mean Temperature of Driest Quarter), BIO12 (Mean Precipitation) and BIO18 as those relevant and weakly correlated predictor variables.

Presence/absence records of the target species were obtained through a combination of two different datasets, EUFORGEN (http://www.euforgen.org/distribution_maps.html) and JRC - FOREST TREE SPECIES IN EUROPE (TSDE; Köble & Seufert, 2001) covering the whole natural range of the species. TSDE maps the percentage of occupancy of tree species within Europe at a 1 km grid, but it does not separate natural and non-natural populations. EUFORGEN delivers a spatially less accurate distribution shape, yet, it includes the whole species´ range within Europe and Northern Africa, and it specifically maps the natural species range. Therefore, by filtering TSDE occurrences with EUFORGEN we obtained a good approximation of both species natural range. To avoid computational issues due to large number of records, we selected a maximum of 100000 presences and as pseudoabsences we selected 5 times de number of presences.

Overall the models achieved very good performance (TSS>0.90 and AUC>0.92) although ensembles models generally outperformed slightly those based just on one algorithm.

Then, we used the relationship between environmental variables and occurrences detected by the current distribution models to predict the species´ distribution in different future climate scenarios. We did this individually for each genetic clade to get insights of the effects of climatic change on the different genetic clades spatial distribution and how we can benefit from this information in order to define priority conservation guidelines based on the distribution patterns predicted. We used the General Circulation Model *MPI-ESM-LR* and its three available scenarios downloadable from WORLDCLIM (rcp26, rcp45 rcp85) to do the first test. Several other GCMs must be incorporated to the analyses to obtain more robust results.

Description of the main results obtained;

- We have obtained fifteen ensemble clade models (wight for Pinus pinaster and seven for Pinus halepensis) based on the predictions of five different algorithms.
- We have obtained two SDMs for both Mediterranean pine species (*P. pinaster* and *P. halepensis*) based on their compounding genetic clades (eight clades and seven clades respectively).
- We have performed a first test of future climate projections of *Pinus pinaster* and *Pinus halepensis* in three future climate scenarios from the MPI-ESM-LR GCM. We will now perform further tests including new GCMs into the analyses.

Future collaboration with the host institution;

The collaboration between these institutions (Instituto Universitario de Investigación y Gestión Forestal Sostenible and Swiss Federal Institute for Forest, Snow and Landscape Research WSL) started in 2013 in the framework of a previous STSM. Again in 2014, this collaboration has been very fruitful for both sides, and will go further in a near future by investigating more deeply how to define conservation guidelines based on the predicted distribution of the species in different future climate scenarios.

Foreseen publications/articles resulting from the STSM;

The work developed during the STSM will be published as soon as the analysis of the results is completely finished, sometime in 2015.

Acknowledgements

In Zurich, 17/11/2014

Mª Jesús Serra Varela

References

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey.

D'Amen, M., Zimmermann, N.E. & Pearman, P.B. (2013) Conservation of phylogeographic lineages under climate change. *Global Ecology and Biogeography*, **22**, 93–104.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Grivet, D., Sebastiani, F., González-Martínez, S.C. & Vendramin, G.G. (2009) Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *The New phytologist*, **184**, 1016–28.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hastie, T.J. & Tibshirani, R. (1990) *Generalized additive models*, Chapman & Hall/CRC, Boca Raton - London - New York - Washington D.C.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Köble, R. & Seufert, G. (2001) *Novel maps for forest tree species in Europe. Proceedings of the 8th European Symposium on the Physico-Chemical Behaviour of Air Pollutants: "A Changing Atmosphere!,"* pp. 17–20. Data set available online at: http://afoludata.jrc.ec.europa.eu/index.php/dataset/detail/66, Torino.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*, (ed. by C. and Hall).

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2013) R: A Language and Environment for Statistical Computing.

Serra-Varela, M.J., Grivet, D., Vincenot, L., Broennimann, O., Gonzalo-Jiménez, J. & Zimmermann, N.E. Does phylogeographic structure relate to climatic niche divergence? A test using maritime pine (Pinus pinaster Ait.).

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.