REPORT OF THE COST SHORT TERM SCIENTIFIC MISSION (STSM):
Development of historical colonization scenarios of *Pinus sylvestris* L. in Scotland

Cost Action Code: COST-STSM-ECOST-STSM-FP1202-131014-050048

1. Purpose of the STSM;

During the last 20,000 years, climate cycles have oscillated from the coldest and driest glacial period centred around 18,000 years ago, to the warmer and more humid period around 9,000 years ago. These climatic oscillations have played an important role in changing the geographical dynamics of plant populations, varying from migration, changes in population size or some degree of adaptation, leaving signatures in their genetic structures (Comes & Kadereit 1998; Hewitt 1996).

Scots pine (*Pinus sylvestris* L.) in Scotland is a keystone species of the remnant Caledonian pine forests. Based on fossil data, this species reached its maximum extent in Scotland around 8000 years ago, covering 1.5 million ha (Bennett 1984; Birks 1989; Froyd 2006). Nowadays, 84 fragments of the ancient native pinewoods remain in this area, scattered over a total area of 17.882 ha and distributed across sites. Today, Scots pine forest in Scotland represents the North West limit of the species' distribution and a globally important marginal population. This raises the possibility that these forests might harbour local adaptations and constitute valuable forest genetic resources.

Estimating the extent of the postglacial dynamics and past genetic processes that contribute to observed patterns in Scots pine is essential for characterizing this ecosystem's dynamics and to understand how future changes are likely to affect biodiversity. Such information would be valuable to inform native pinewood conservation and management in Scotland.

The **purpose** of the Short Term Scientific Mission carried out was to receive training in Approximate Bayesian Computation (ABC) analyses and other bayesian clustering methods from experienced researchers in order to test past demographical scenarios and contemporary structure of *Pinus sylvestris* L. in Scotland.

Approximate Bayesian Computation (ABC) is a powerful simulation method with the potential to handle virtually any complex models, provided that simulation of data under the model is feasible (Beaumont et al. 2002). ABC can allow us to estimate demographic history and local adaptation events by investigating the genetic consequences of demographic changes detected in contemporary populations (Csilléry et al. 2010) (Lander et al. 2011). Other bayesian model-based clustering method to infer population structure, have the ability to provide a description of clusters by making use of multilocus genotypes obtained from a sample of individuals. The simulation of historical demographic scenarios of *Pinus sylvestris* through the use of ABC analysis and the inference of contemporary structure could promote a more comprehensive understanding of demographic history of Scots pine in Scotland and how genetic variability has developed and been maintained.

Specific objectives were (1) to estimate the contemporary genetic structure of the Scots pine populations; (2) to estimate the genetic processes after the last glaciation maximum; (3) to understand the consequences of historical population size change and structure for contemporary patterns of genetic diversity and gain detailed insights into the evolutionary dynamics of Scots pine in Scotland.

2. Description of the work carried out during the STSM;

To achieve the objectives describe above, the main scientific tasks carried out during the STSM were:

1. Consolidation of existing genetic data for Scots pine in Scotland
2. Training of software packages to study the contemporary genetic structure of the Scottish populations of Scots pine and to study the evolutionary dynamics of Scots pine in Scotland.

Existing genetic data from a total of 20 populations from Scotland were used for this study (Figure 1). They included 6 nuclear SSR and 5 chloroplast SSR for 18 population, and 236 SNPs from 12 candidate genes from 12 population. Dataset were consolidated and converted for the use of the different package software used.
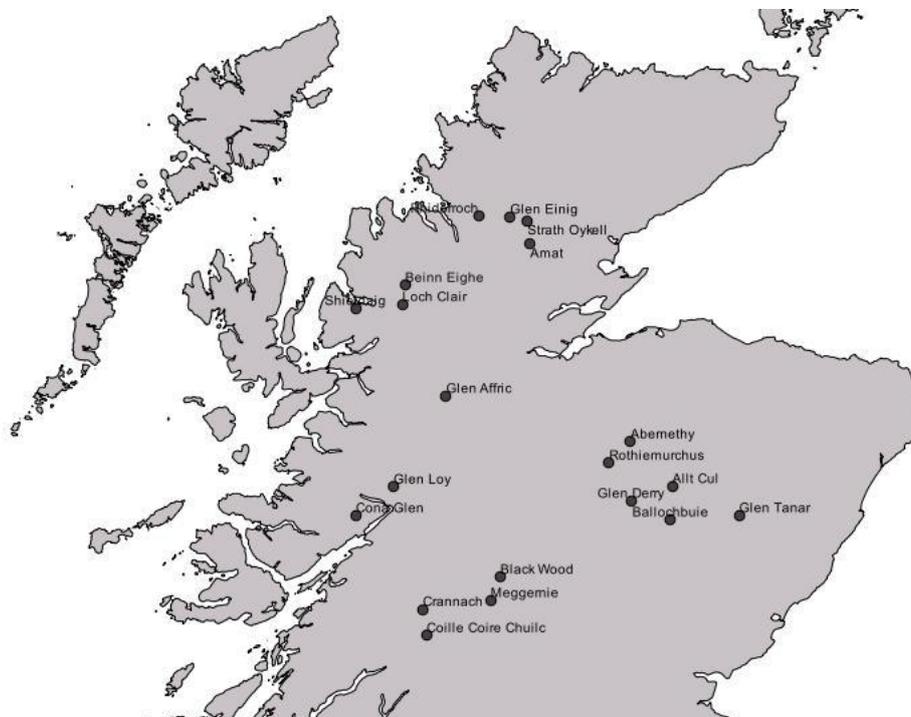

Figure 1: Populations from Scotland that have been studied.

2.1.1. Studying contemporary genetic structure of Scots pine

To study the contemporary genetic structure the training and use of different packages software was carried out with STRUCTURE v.2.3.3 (Pritchard et al. 2000), BAPS and TESS. To detect the optimum value of K, Structure Harvester website was used (Earl and vonHoldt, 2012). Also, several software for calculating summary statistics were used, as GenAlex, Arlequin (Excoffier et al., 2010), and FSTAT (Goudet, 2002).

2.1.2. Studying the evolutionary dynamics of Scots pine in Scotland.

To study the genetic processes and evolutionary dynamics of Scots pine in Scotland, we used the ABC framework of DIYABC v2.0 (Cornuet et al., 2014). Four scenarios were tested (Figure 2) to detect and quantify population size change. To estimate the timing in the geographical processes, three scenarios were also tested (Figure 3).
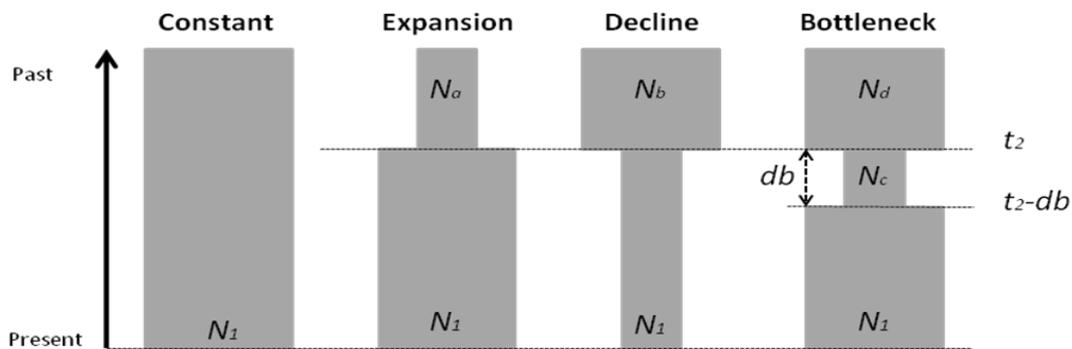

Figure 2: Description of the four scenarios of **demographic change** tested in DIYABC for Scots pine.
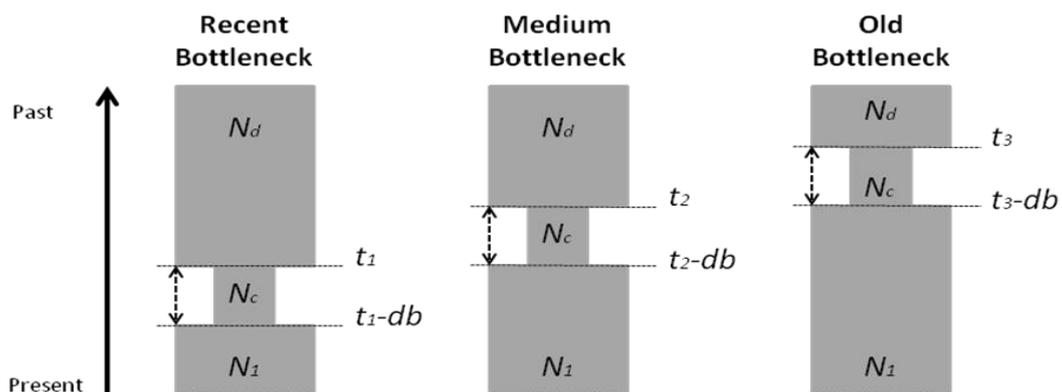

Figure 3: Description of the three scenarios of **timing** for the bottleneck scenario tested in DIYABC for Scots pine.

Other software were also used for previous exploratory analysis of demographic history and to detect deviation of particular genes from neutrality, as the software BOTTLENECK and BAYESCAN respectively.

3. Description of the main results obtained

3.1. Genetic structure and diversity

Little genetic structure associated with geographical patterns was observed in the Scottish populations (Figure 4). Only for the SNPs markers, the bayesian inference of genetic structure with STRUCTURE suggest a more significant division of three clusters (K=3). Nevertheless, individuals representing clusters were mixed between all the 15 populations, with no correspondence to geographical areas. Genetic structure was also analysed with BAPs and TESS, and similar results were found.
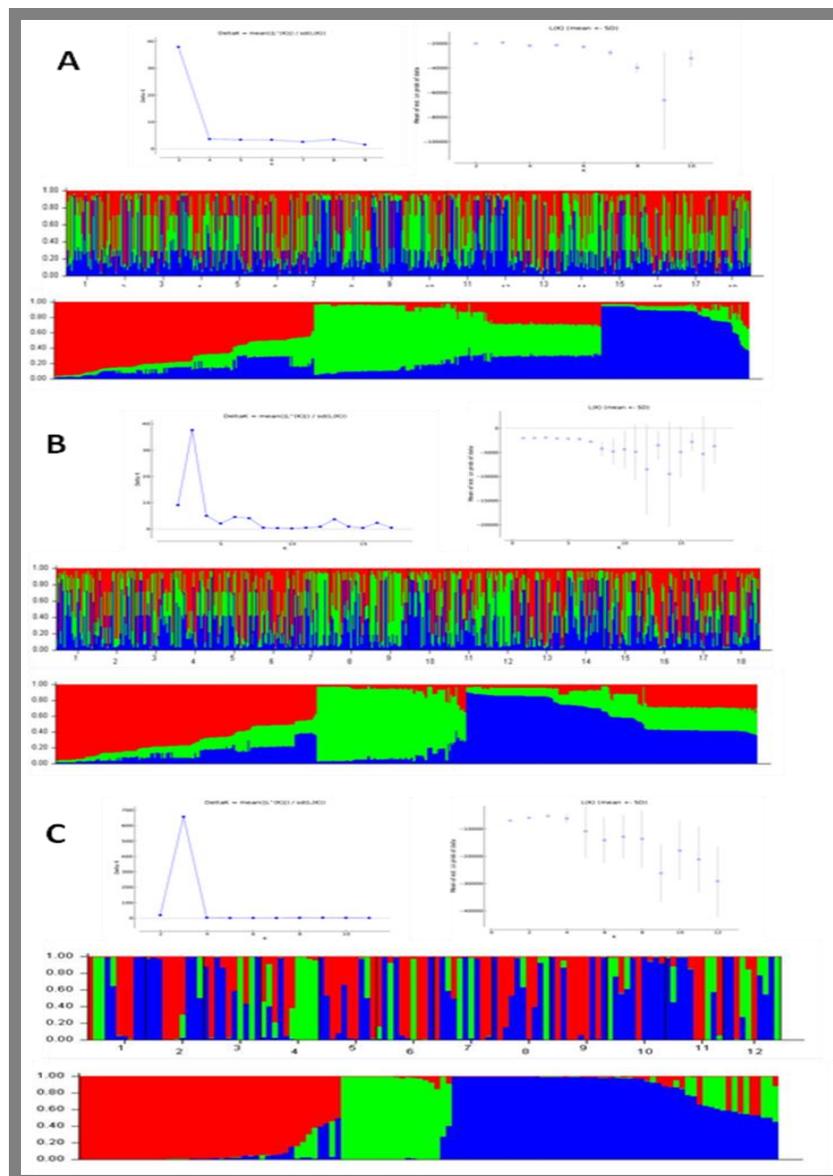

Figure 4: Population structure as estimated using STRUCTURE 2.3.3. A, B and C correspond to the structure results from nSSR, cpSSR and SNPs respectively.

3.2. Evolutionary dynamic of Scots pine in Scotland

More than 200 run of around 400.000 simulations each have been tested for selecting the best scenario possible for the Scots pine populations.

Nuclear markers showed the **bottleneck** scenario as the most likely, dated between 1160 and 570 generations ago (Scots pine have an estimated generation time of 20-25 years). These dates correspond to 23,000-29,000 and 11,000-14,000 years ago respectively. On the other hand, chloroplast SSR showed the **expansion** as the best scenario. The expansion was dated about 400 generations ago, which correspond to 8.000 – 10.000 years ago. SNPs markers will be processed soon.

Some examples of the results in DIYABC are described below:

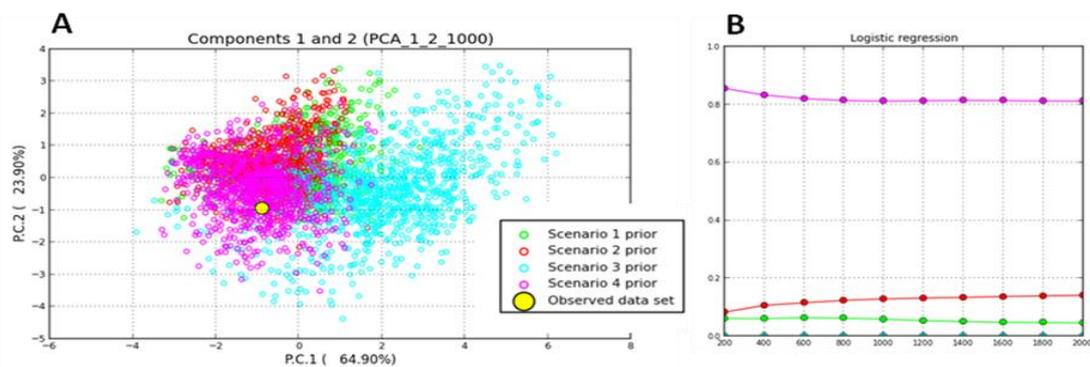- Testing **demographic change**

Figure 4: A, principal component analysis on the summary statistics when processing pre-evaluation of scenarios in DIYABC for the four tested **demographic change scenarios** for **nSSR;** and B, posterior probability of scenarios. Scenario 1 is equivalent to a constant population size; scenario 2 correspond to an expansion in a population; scenario 3 is a decline in the population size, and scenario 4 correspond to a bottleneck.
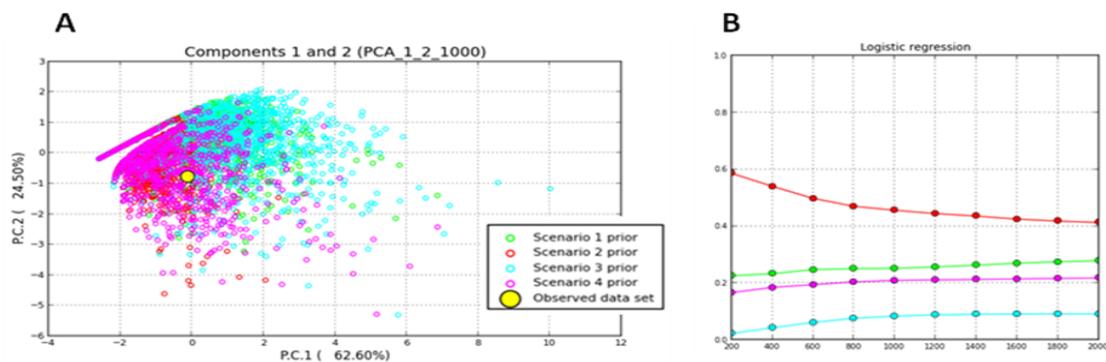

Figure 5: A, principal component analysis on the summary statistics when processing pre-evaluation of scenarios in DIYABC for the four tested **demographic change scenarios** for **cpSSR;** and B, posterior probability of scenarios. Scenario 1 is equivalent to a constant population size; scenario 2 correspond to an expansion in a population; scenario 3 is a decline in the population size, and scenario 4 correspond to a bottleneck.
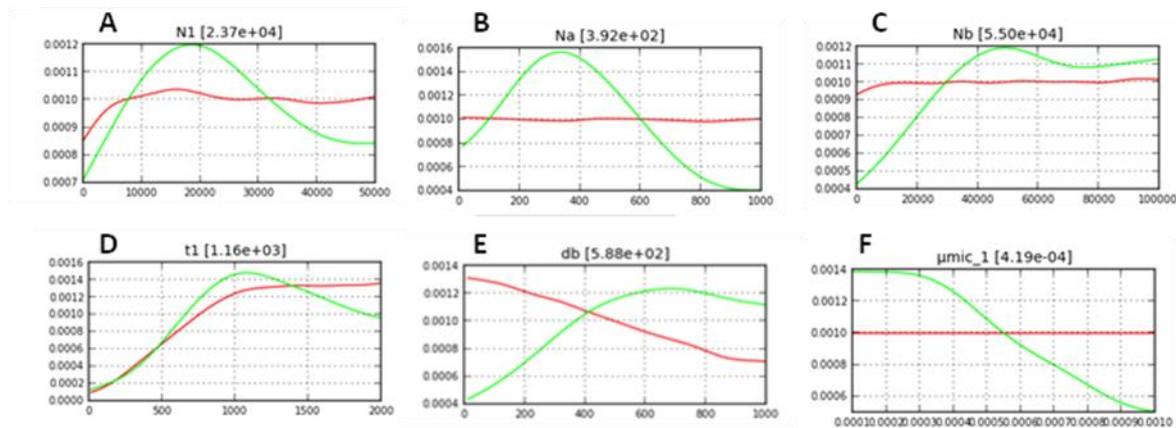
- Estimating parameters

Figure 5: Parameters values when processing posterior distribution of parameters in DIYABC for the recent bottleneck scenario for nSSR. A, B, C correspond to the contemporary effective population size, population size during the bottleneck, and ancient population size respectively; D and E are the time when the bottleneck started and the duration respectively, and finally F correspond to the mutation rate.

Scottish populations showed signs of a "recent" demographic bottleneck which corresponded with the LGM. This bottleneck was followed by an expansion in the population size, which was also supported by the chloroplast markers.

The information provided in this STSM detail insights into the evolutionary dynamics of Scots pine in Scotland, regarding past population sizes and timing of evolutionary processes. These preliminary results are in concordance with other studies (Wachowiak et al. 2010) and fit with the LGM event. However, drawing conclusions about the past demography and colonization history of the Scottish populations will need more analysis. SNPs results will add some information to the demographic history of Scots pine in the near future, and the effect of compound environmental variables will be also investigated.

Overall, the STSM was a high success, as the main goals were reached and analyses were performed. The work in close collaboration and frequent discussions and exchanges with Beppe Vendramin and Francesca Bagnoli, introduced me successfully to the novel statistical approaches and helped me to understand results and following steps of the analysis.

As a result of the present work, a strengthening of the existing networks between the Centre of Ecology and Hydrology (CEH) and the Institute of Biosciences and Bioresources of Florence have taken place. Future collaboration with the host institution will be carried out to finalize the present work. A scientific publication regarding demographic evolutionary dynamics of Scots pine will result of the present work when all the analysis will be ready.

REFERENCES

Bennett, K.D., 1984. The post-glacial history of Pinus sylvestris in the British Isles. *Quaternaty Science Reviews*, 3(1973), pp.133–155.

Birks, H.J.B., 1989. Holocene Isochrone Maps and Patterns of Tree-Spreading in the British Isles. *Journal of Biogeography*, 16(6), pp.503–540.

Beaumont, M. a, Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–35.

Comes, H.P. & Kadereit, J.W., 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, 3(11), pp.432–438.

Cornuet J.M., Veyssier J., Pudlo P., Dehne-Garcia A., Gautier M., Leblois R., J.M. Marin and A. Estoup. DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data. *submitted to Bioinformatics.*

Csilléry, K. et al., 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in ecology & evolution*, 25(7), pp.410–8.

Earl, D. A. and B. M. von Holdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361.

Excoffier, L. and H.E. L. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Froyd, C.A., 2006. Long-term ecology of native pinewood communities in East Glen Affric, Scotland. *Forestry*, 79(3), pp.279–291. Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907–13.

Goudet, J.,2002. FSTAT version 2.9.3.2 (Feb. 2002). Institute of Ecology. Biology Building, UNIL. CH-1015, Lausanne. Switzerland.

Hewitt, G.M., 1996. Some genetic consequences of ice ages , and their role , in divergence and speciation. *Biological Journal of the Linnean Society*, (58), pp.247–276.

Lander, T.A. et al., 2011. Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Molecular ecology*, 20(24), pp.5182–96.

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

Wachowiak, W., Salmela, M. J., Ennos, R. a, Iason, G., & Cavers, S. (2011). High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (Pinus sylvestris L.) in Scotland. *Heredity*, *106*(5), 775–87.