Livestock Genomic Resources in a Changing World Conference

SUBSAMPLING AS AN ECONOMIC CONSEQUENCE OF USING WHOLE GENOME SEQUENCE DATA IN LANDSCAPE GENOMICS: HOW TO MAXIMIZE ENVIRONMENTAL INFORMATION FROM A REDUCED NUMBER OF LOCATIONS?

K Leempoel, S Stucki and S Joost
Email: kevin.leempoel@epfl.ch

The recent availability of whole genome sequence (WGS) data implies to reconsider sampling strategies in landscape genomics for economic reasons. Indeed, while we had many individuals and few genetic markers ten years ago, we now face the contrary with high costs of WGS limiting the number of sequenced samples. In others words, molecular resolution is becoming excellent but it is achieved at the expense of spatial representativeness and statistic robustness. Therefore, when starting from a standard sampling, it is necessary to apply sub-sampling strategies in order to keep most of the environmental information. To study local adaptation of goats and sheep's breeds in Morocco, we used a sampling design based on a regular grid overlaid on the territory. In each cell of this grid, 3 individuals were sampled in 3 different farms. Then, the final subset destined to sequencing had to meet two criteria in order to ensure a regular cover of both environmental and physical spaces. The first was met by using stratified sampling techniques over a range of climatic variables, previously filtered using a PCA. The second was by minimising a clustering index in order to ensure spatial spread. The sub-sampling procedure using a hierarchical clustering resulted in two datasets of 162 goats selected over 1283, and 162 sheep over 1412 based on variables such as temperature, pluviometry and solar radiation. By maximising the environmental information collected, we were able to select individuals that are the most relevant to study adaptation.