

## Exploring the genome of Norway spruce (*Picea abies*) in Swedish/Finnish populations

Venkata Raghava Pavankumar Thunga

The main goal of this project was to broaden the earlier attempts to estimate genetic parameters, such as nucleotide diversity and linkage disequilibrium, in populations of the ecologically and economically important conifer tree species Norway spruce (*Picea abies*). Nucleotide diversity, the level of genetic diversity in a population, holds information about the size and changes in a population. Pattern of linkage disequilibrium (LD), which is the occurrence of non-random association of genetic variants along the genome, can also contain information on population size, but is in general more related to the recombination rate. Both nucleotide diversity and LD needs to be known to efficiently design an association study that is a study aiming at connecting genetic variants with phenotype of interest e.g. a certain disease.

In this study DNA from 23 different loci (i.e. places) of the genome of Norway spruce were sequenced and analyzed to obtain more detailed information on population genetic parameters. The mean nucleotide diversity was 0.002 lower than expected from a species with a large current distribution range, but in line with previously published genetic data from Norway spruce. This low diversity at genetic level is likely the result of fluctuations in population size over time and expected from a growing population. The average Tajima's  $D$  a value that calculates the deviation in a population ended up in a negative value -0.86 suggesting the data comes from an expanding population. To be precise, demographic analyses also conducted on three sub populations and the results gave support for the conclusion that the Swedish/Finnish spruces are a growing population with a large effective population size.

Despite the fact the sampled populations are distributed over large parts of Sweden and Finland no clear population genetic structure could be identified. A population is described as structured if there is any inbreeding in population which means mating non-randomly. The lack of structure in the northern range of the species could also be confirmed by re-analysis of a publicly available population genetic data set suggesting that the recent establishment of spruce in Scandinavia together largely like a single large population. In contrast to this there was a clear differentiation between populations from southern Europe and the more northern population.

Patterns of LD in the sampled populations verified earlier data where LD decayed to less than 0.2 even with a few 100bp also consistent with earlier data. The level of linkage is explained by how long it extends over the genome and mostly in spruce the decay is faster within hundreds of base pairs. If one extrapolates these patterns to the whole genome it makes easy for association studies but due to the given size of the spruce genome this likely prohibits attempts with current genotyping technologies.