

## Future Research Plans

My long-term goal is to conduct meaningful research in population and evolutionary genomics with special emphasis on understanding adaptation in forest tree genomes. I want to continue my *P. sylvestris* population genetics research, where I am well established and have good background knowledge and research networks. However, I also want to extend my current research focus into a wider set of tree species. This will allow deeper understanding on what are the factors determining the adaptive evolutionary dynamics. I am excited about the possible opportunity to work in Organismal and Evolutionary Biology research programme where both of my passions: plant evolutionary biology and genomics can be combined.

### Genetic basis of adaptive clinal variation in trees

Mutations creating phenotypic and fitness differences between individuals are the raw material of natural selection. There is mounting evidence that mutations in regulatory regions affecting the level of gene expression are more critical to adaptation than previously assumed. The recent 'functional space hypothesis' suggests that larger genomes have a larger mutational target in regulatory regions and that genome size has a critical effect on the prevailing evolutionary processes.

I plan to use a recently developed molecular method to identify regulatory open-chromatin regions of the genome (ATAC-seq). Together with gene expression measurements (RNA-seq), DNA re-sequencing, common garden experiments and quantitative genetics analyses it will be used to identify adaptive variation on coding and regulatory areas. The distribution of loci underlying adaptive variation and effects of natural selection along the genome will be compared among species with contrasting genome sizes.

The choice and number of species is flexible, but as an example, I propose to contrast *P. sylvestris* (Scots pine) and *B. pendula* (Silver birch), two tree species with phenotypically well-characterized local adaptation patterns. These two species differ dramatically in genome size: *P. sylvestris* has 22 000 Mb genome, whereas the *B. pendula* genome is only 440 Mb. *B. pendula* has an advantage of short generation time, facilitating classical genetic studies involving crossing experiments. As an angiosperm, it is closer to other model organisms that provide background information about possible molecular basis of adaptive changes. Combining research on birch and pines, two species with woody life style and phenotypic adaptations, yet very distant evolutionary histories, facilitates understanding plant adaptation in more general level.

The results have potential to have a fundamental impact on how we view environmental adaptation as a process. For example, if the genetic basis of adaptive quantitative variation is spread widely across the genome, the analytical tools designed to identify e.g. hard selective sweeps and positive selection in protein evolution are weak. New methods targeted more towards joint analysis of multiple loci and taking into account the regulatory networks may be

more suitable. Further, if the genome sizes of the two species have a major consequence in the evolutionary processes, evolutionary geneticists need to start taking the difference in functional space into account in genome-wide analyses of evolution. Finally, as this would be the first large scale study of open chromatin and its effects on gene expression and adaptation in gymnosperms, there is potential for identifying completely new strategies of adaptation via expression regulation.

I have submitted an R'Life consortium grant proposal as a main PI on this topic to the Academy of Finland in September 2019.

## **Gymnosperm haploid life-stage as an evolutionary model system**

From an evolutionary perspective, conifers provide a unique glimpse to plant evolutionary history. Evolutionarily, it is a vascular seed plant but has some more archaic properties compared to angiosperms. For example, the development of seed, reproductive cycle and extent of the haploid life stage of conifers are different from angiosperms. Detailed analysis of haploid tissue development, genetics and evolution allow me to answer questions on ploidy level evolution, genetic imprinting, allele specific expression, evolution of double fertilization in angiosperms, etc. An important practical advantage over angiosperms is the relatively large size of the haploid tissue, which allows easy determination of maternal haplotype and subsequent separation of maternal and paternal haplotypes in the following generation.

In diploid or polyploid organisms, different alleles do not necessarily have an equal contribution to the phenotype due to dominance variation. Further, diploid organisms also have life stages where some genes are expressed in a haploid manner. In these situations, natural selection can directly affect recessive alleles. In an ongoing 5-year Academy of Finland funded study, we have started to investigate this phenomenon, using *P. sylvestris* seed tissues as a model system. We have assembled several new high quality *de novo* transcriptomes specifically for the purpose of haploid and diploid tissue analyses. We have identified genes that are preferentially expressed in seed haploid tissue and measured allele specific expression in reciprocal crosses to find genes displaying genetic imprinting. In the next stage, we will measure genetic diversity and its distribution in a population. Based on allele frequency spectra, we will evaluate whether natural selection is stronger in haploid-expressed genes, as expected based on theory.

Finally, in collaboration with Dr. Brandvain (University of Minnesota), we have developed a model for evolution of polyzygotic polyembryony, another gymnosperm-specific trait. I plan to continue this branch of research to understand polyembryony's role in avoiding adverse effects of inbreeding depression and as an early stage of the evolution of self-incompatibility. Many conifers have a large genetic load and high inbreeding depression. Understanding the mechanisms responsible for it and distribution of deleterious alleles along the genome and individuals are not only evolutionarily important, but allow seed orchard managers and conifer breeders avoid possible unwanted unfavorable effects due to founder relatedness.