

Joanna Sikora

Title of doctoral dissertation: Application of genome skimming to characterize the genetic diversity of closely related taxa in the *Pinus mugo* complex

Abstract

The *Pinus mugo* complex is a large group of closely related European pines that are a component of the ecosystems of the most important mountain ranges, such as the Alps, the Carpathians and the Pyrenees. The high morphological variability of individual taxa, sympatric occurrence, hybridization and introgression processes, the presence of hybrids and many synonymous names, as well as the convoluted history of recolonization, make the *Pinus mugo* complex an interesting, but at the same time very complicated and demanding research object. Taxonomic and genetic relationships between taxa in this complex have been the subject of research and discussion for many years. Despite the use of many different research approaches, they still require clarification and extension with new data and results, especially those obtained using high-throughput methods and bioinformatic tools.

In this doctoral dissertation, genome skimming and next-generation sequencing were used to: 1) characterize the genetic diversity of the *Pinus mugo* complex pines and closely related taxa, i.e. *Pinus sylvestris* and *Pinus* × *rhaetica*, 2) select regions potentially useful for their identification and further analysis, and 3) make phylogenetic inferences based on sequences from the chloroplast, mitochondrial, and nuclear genomes.

Genome skimming is a novel approach that allows obtaining DNA fractions that occur in a cell in a high copy number, i.e. nuclear ribosomal DNA (nrDNA) and partial or complete sequences of plastomes and mitogenomes.

On this basis, a detailed comparative analysis of the chloroplast genomes of the three main representatives of the *Pinus mugo* complex, i.e. *Pinus mugo* Turra, *P. uncinata* Ramond and *Pinus uliginosa* Neumann, as well as the closely related *Pinus sylvestris*, was performed for the first time. The obtained results showed a high similarity of the chloroplast genomes of the three analyzed taxa, in terms of length, structure and number of genes, and confirmed their close phylogenetic relationships. Additionally, a first-ever assessment of the validity of using the nuclear internal transcribed spacer 2 (ITS2) region for identifying taxa in the Pinaceae family, including taxa from the *Pinus mugo* complex, showed that the usefulness of the ITS2 region for this purpose is quite limited. Seven data sets composed of nucleotide sequences

from three organellar genomes were also generated using genome skimming and used to assess the genetic diversity of taxa and populations belonging to the *Pinus mugo* complex and to determine their suitability for identifying the analyzed taxa using three different delimitation methods. The highest percentage of success in distinguishing taxa from this complex was achieved using the distance-based method and a set composed of highly variable regions (so-called hotspots).

The obtained results provide new and extensive genomic resources for further studies of the *Pinus mugo* complex, and confirm the effectiveness and validity of using the genome skimming method to analyze the genetic diversity of closely related taxa.

Keywords: *Pinus mugo* complex, genome skimming, next-generation sequencing, taxa identification