ABSTRACT

Angelika Maria Gomolińska

THE VARIABILITY OF THE ORGANELLAR GENOMES AND SELECTED NUCLEAR GENES OF THE GENUS *ERICA*

The doctoral thesis was written under the supervision of prof. dr hab. Jakub Sawicki in the Department of Botany and Evolutionary Ecology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn.

The genus *Erica* (heath) is the second most abundant genus belonging to the *Ericaceae* family (heath family). Evolution of this genus, despite of the dynamic development of phylogenetic research in recent years, still remains weakly understood and poorly described. This genus is particularly interesting from a research perspective, among other things, due to the fact that it consists mainly, almost 80%, of endemic species. In the genomic documentation of *Erica* representatives exists a explicit gap, which makes it impossible to reconstruct the evolutionary history, identify the ancestral species and determine the exact relationships between individual species within this genus. Evolutionary research on the *Ericaceae* family, including the acquisition of entire chloroplast genomes, has gained recently, but complete plastomes of *Erica* genus are still unknown. Research on this genus, which aimed to revise the classification system, was mainly based on morphological predictors and sets of genetic markers. Phylogenetic studies have not been conducted taking into account to complete groups / families of genes of the chloroplast genome, or to complete NOR regions. So far, no reconstruction of the evolutionary history has been made to select the most probable ancestor, and the relationship between individual members of the genus based on i.a. complete plastid genomes or NOR sequences has not been established.

This doctoral dissertation is an important stage of genetic research on selected representatives of the *Erica* genus, which may contribute to expanding knowledge about the genomic resources of these plants based on organelle genomes, plastid operons and NOR nuclear sequences. The main goals of this work were, among others: detailed understanding of the patterns of genetic variability of plastomes within the *Erica* genus, identifying genes that

are more effective genetic markers than those previously used, proposing potential barcodes for the *Erica* genus, cognizance the structure of NOR clusters in the *Erica* genus, which are a good markers for detecting hybrids, or reconstructing phylogenetic relationships between selected species of *Erica*, in order to determine whether the origin of this genus lies in the Europe or in the Africa.

In order to implement them, the following research tasks were performed:

1. the chloroplast genome of *Erica versicolor* was sequenced and a comparative analysis was made between this cpDNA and others available in the NCBI database from the same family - *Ericaceae*;
2. the chloroplast genome of *Erica arborea* was sequenced and a comparative analysis was made between this cpDNA and others available in the NCBI database from the same family - *Ericaceae*; and a comparative analysis was performed between both complete chloroplast genomes: *E. arborea* and *E. versicolor*;
3. complete sequences of all chloroplast genes of representatives of the genus *Erica* were obtained by mapping the libraries to the obtained complete genome of *Erica versicolor* and their comparative and variability analysis was performed;
4. phylogenetic informativeness and phylogenomic analysis were assessed based on the genes of the chloroplast genome in the genus *Erica*;
5. detailed characterization of the nuclear NOR regions of the *Erica* genus was performed;
6. phylogenomic analysis was performed based on the sequences of nuclear NOR regions in the genus *Erica*;
7. comparative analysis of the inverted repeat (IR) regions of the complete sequences of chloroplast genomes from the *Erica* genus was performed with the complete chloroplast genomes of other representatives of the *Ericaceae* family;
8. patterns of genomic variation among the chloroplast genomes of the genus *Erica* were calculated and visualized; showing their mutual degree of sequence similarity.

The results obtained during the implementation of this work and analyses allow to conclude that:

1. the most likely ancestral species in the genus *Erica*, at the current level of research, seems to be *Erica arborea*;
2. the *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J and *psb*B genes turned out to be the most promising phylogenetic markers for the genus *Erica*, which may be an incentive to develop barcode sequences for the genus studied;
3. relatively small differences between the juxtaposed genomes of the genus *Erica* concern the layout and arrangement of genes, which may indicate a specific nature of the evolution rate of South African plastomes in relation to the dispersed bicontinental species, but it is difficult to assess this character as highly conservative, because the comparison of genomes shows a relatively low level of conservation on level of 63%;
4. analysis of the regions of inverted repeats allows to significantly distinguish sequences belonging to the genus *Erica* from other members of the family *Ericaceae*;
5. the comparison of individual phylogenetic trees prevents effective phylogenetic reconstruction within the genus and establishing relationships between the analyzed species;
6. the species studied show a large number of repeat sequences, mostly concentrated in non-coding regions of the plastomes;
7. non-coding regions are characterized by a much higher level of changes and divergence than the coding regions of the examined cpDNA;
8. analyzes of the NOR regions, the most phylogenetically informative area showed 3'ETS + NTS, the patterns of diversity of repeated sequences within IGS reflect, to some extent, the evolutionary processes taking place in the genus *Erica*.

The phylogenetic reconstruction of the *Ericaceae* family based on whole genomes and constructed on the basis of inverted repeat regions included for the first time not only species of the genus *Erica*, but also the species *Chamaedaphne calyculata*. Moreover, this work also included for the first time the precise structure of the IR for the genus *Chamaedaphne*. The conducted research provided extremely interesting information in the field of comparative analyses, providing new data extending the existing knowledge of genetic variability and genomic resources of the *Erica* genus, also enriching the documentation of the entire *Ericaceae* family.