

Title of dissertation: Building a Model: Developing Genomic Resources for *Ferula communis* (Apiaceae), a Traditional Medicinal Plant

Author: Mergi Daba Dinka

Abstract

Angiosperms, representing the most diverse group of land plants, have complex genomes that are believed to underpin their evolutionary success. However, due to the absence of data for numerous angiosperm groups, our understanding of genome architecture remains limited. In this work, I address this knowledge gap by assembling and analyzing the genome of *Ferula communis*, a species from the economically important plant family Apiaceae.

The nuclear genome of *F. communis* was assembled utilizing sequences obtained from next-generation sequencing technologies, encompassing both Illumina short-read and Oxford Nanopore Technology (ONT) libraries. A wide spectrum of bioinformatic tools, including short-read, long-read, and hybrid genome assemblers, was used in this process. The most superior assembly, based on multiple metrics, was produced by the long-read assembler Flye. This resulted in a genome size of 2.8 Gb, composed of 59,178 contigs, an N50 of 0.17 Mb, and a genome completeness covering 93.2% of BUSCO genes. Notably, transposable elements constituted 86.67% of the genome, with *Gypsy* and *Copia* retrotransposons being particularly prevalent, highlighting their pivotal role in genome evolution. The functional genome annotation identified 68,318 protein-coding genes, and 1,262 tRNA genes. When compared to the genomes of carrot, celery, and coriander from the Apiaceae family, *F. communis* exhibits a higher count of annotated genes and a more extensive gene family expansions. Within these expanded gene families, I identified genes tied to defense mechanisms, stress responses, and vernalization. Surprisingly, *F. communis* shares more unique genes with the distantly related coriander than with carrot and celery. A notable instance is the genes tied to terpenoid biosynthesis. The probable reason for this enhanced gene sharing with coriander over other studied Apiaceae members is genome reduction in carrot, leading to extensive gene deletion. Conversely, in celery, its primary aquatic habitat may have necessitated a specific gene set distinct from those required in terrestrial environments.

The plastid genome of *F. communis* was assembled using two assemblers: GetOrganelle and Novoplasty. Spanning 166,696 bp, it encodes 132 genes, comprising 87 protein-coding, 37 tRNA, and 8 rRNA genes. This genome reflects the standard quadripartite structure observed in other angiosperm plastid genomes, which includes paired IR, LSC, and SSC regions. The *F. communis* plastid DNA

houses 217 simple sequence repeats (SSRs), with a predominant concentration in the LSC region. Several intergenic spacer (IGS) regions demonstrated pronounced nucleotide diversity and have potential as informative phylogenetic markers. Interestingly, only the *ccsA* gene, which is responsible for heme attachment to cytochrome c, displayed sites of strong positive selection within the *F. communis* plastid genome.

The mitochondrial genome was assembled using GetOrganelle and analyzed for gene content and structural variations. Interestingly, rather than forming a single master circle, the *F. communis* mtDNA genome assembly resulted in 16 scaffolds with a total genome size of 250,278 bp. This suggests a non-circular genome structure in this species. Within these scaffolds, 37 protein-coding genes, 3 rRNA genes, and 20 tRNA genes are present. Among the annotated genes, eight of the protein-coding genes and three tRNAs contain introns of varying lengths. In agreement with other assembled mtDNA genomes in the Apiaceae family, 183 SSRs were identified, 82% of which are mono- and dimeric repeats. Moreover, 385 RNA editing sites were identified within 25 protein-coding genes. Of these edits, 61.14% transition amino acids from hydrophilic to hydrophobic states, while 31.09% involve alterations between hydrophobic amino acids.

The results presented here establish a foundation and provide a valuable genomic resources for future investigations into the genomics, evolution, and medicinal properties of this important plant species.