



Fakulta zemědělská
a technologická
Faculty of Agriculture
and Technology

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

UNIVERSITY OF SOUTH BOHEMIA IN ČESKÉ BUDĚJOVICE

FACULTY OF AGRICULTURE AND TECHNOLOGY

Department of Genetics and Biotechnology

Ph.D. thesis

**Genetic diversity and molecular barcoding of caraway
(*Carum carvi* L.) genetic resources**

Author: Samuel Goitom Misginna
Supervisor: prof. Ing. Vladislav Čurn, Ph.D.
Consultant: Mgr. et Ing. Ondřej Hejna, Ph.D.

České Budějovice

2026

Abstrakt

Kmín kořený (*Carum carvi* L.) je významná aromatická, léčivá a kořeninová plodina a správný management genetických zdrojů a jejich využití ve šlechtění závisí na spolehlivé charakterizaci genetické diverzity. Tato práce hodnotila genetickou diverzitu, populační strukturu a potenciál molekulárního barcodingu u geograficky rozmanité kolekce genetických zdrojů kmínu pomocí SSR a ISSR markerů. Celkem bylo analyzováno 130 položek genetických zdrojů pocházejících z 24 zemí a šesti geografických regionů. Po úvodním testování markerů a kvalitativním filtrování bylo do SSR analýzy zařazeno 85 položek genetických zdrojů a do ISSR analýzy 105 položek, přičemž 76 položek bylo společných pro oba datasety.

Pro deset vybraných SSR lokusů bylo získáno 83 alel, z nichž 45 (54,2 %) bylo polymorfních, zatímco osm ISSR primerů generovalo 71 skórovatelných markerů, z nichž 59 (83,1 %) bylo polymorfních. ISSR markery vykazovaly vyšší polymorfismus a vyšší informativnost na marker, zatímco SSR markery umožnily úplnou diskriminaci jednotlivých položek, přičemž všech 85 položek genetických zdrojů vykazovalo unikátní multilokusové profily (minimální Jaccardova vzdálenost = 0,044). Oba markerové systémy prokázaly značnou genetickou diverzitu, přičemž genetické zdroje ze střední Evropy vykazovaly nejvyšší diverzitu i nejvyšší počet privátních alel/markerů. Analýza populační struktury identifikovala v obou datasetech šest hlavních genetických klastrů ($K = 6$) a signifikantní Mantelův test ($r = 0,257$; $P = 0,001$) potvrdil shodu mezi oběma markerovými systémy. Celkově se ukázalo, že SSR a ISSR markery jsou komplementární nástroje, přičemž SSR byly účinnější pro identifikaci jednotlivých genetických zdrojů a ISSR informativnější pro hodnocení širších vzorců genetické diverzity a geografické struktury. Získané výsledky představují užitečný molekulární základ pro ochranu genetických zdrojů kmínu, ověřování identity genetických zdrojů a jejich budoucí využití ve šlechtění.

Klíčová slova: *Carum carvi* L.; genetická diverzita; populační struktura; SSR; ISSR; molekulární barcoding; genetické zdroje

Abstract

Caraway (*Carum carvi* L.) is an important aromatic, medicinal, and spice crop whose conservation and utilization depend on the reliable characterization of genetic diversity. This study assessed the genetic diversity, population structure, and molecular barcoding potential of a geographically diverse caraway germplasm collection using SSR and ISSR markers. A total of 130 accessions from 24 countries and six geographic regions were analyzed. After marker screening and quality filtering, 85 accessions were retained for SSR analysis and 105 for ISSR analysis, with 76 accessions shared between both datasets.

The 10 selected SSR loci generated 83 allele bands, of which 45 (54.2%) were polymorphic, whereas 8 ISSR primers produced 71 scorable bands, of which 59 (83.1%) were polymorphic. ISSR markers showed higher polymorphism and informativeness per band, whereas SSR markers provided complete accession-level discrimination, with all 85 accessions displaying unique multilocus profiles (minimum Jaccard distance = 0.044). Both marker systems revealed substantial genetic diversity, with Central Europe showing the highest diversity and the greatest number of private bands. Population structure analysis identified six major genetic clusters in both datasets ($K = 6$), and a significant Mantel test ($r = 0.257$, $P = 0.001$) confirmed concordance between the two marker systems. Overall, SSR and ISSR markers proved complementary, with SSR being more effective for accession identification and ISSR being more informative for broader diversity patterns and geographic structure. These findings provide a useful molecular framework for caraway germplasm conservation, accession verification, and future breeding.

Keywords: *Carum carvi* L.; genetic diversity; population structure; SSR; ISSR; molecular barcoding; germplasm

Table of Contents

1. INTRODUCTION	1
2. HYPOTHESES	2
3. AIMS OF THE THESIS	3
4. LITERATURE REVIEW	4
5. MATERIALS AND METHODS	7
6. RESULTS	10
7. DISCUSSION	14
8. CONCLUSION.....	17

1. INTRODUCTION

Caraway (*Carum carvi* L.) is an important aromatic, medicinal, and spice crop that occupies a valuable position among cultivated plant genetic resources. Its importance is linked not only to its economic and pharmaceutical uses, but also to its biological and agronomic variability. Caraway is cultivated across different temperate regions and shows variation in adaptation, reproductive performance, and seed-related quality traits. Such variability makes it an appropriate subject for genetic study, especially when the objective is to understand how diversity is organized within germplasm collections and how that diversity can be used in breeding and conservation.

The introduction of the thesis establishes genetic diversity as the fundamental basis for crop improvement and long-term germplasm utilization. In breeding programs, genetically diverse materials provide the variation necessary for identifying useful traits and for combining desirable characteristics into improved cultivars. These may include better agronomic performance, environmental adaptation, higher quality seed traits, improved essential oil profiles, or other breeding-relevant features. At the same time, genetic diversity is central to the rational management of gene bank collections because it helps distinguish unique accessions from redundant ones and supports the maintenance of a collection that is both representative and efficient. The introductory chapter also emphasizes the growing importance of molecular markers in genetic resource research. Although morphological and agronomic traits remain useful, they are often affected by environmental conditions, plant age, and phenotypic plasticity. Molecular markers overcome these limitations by examining genetic variation directly at the DNA level. As a result, they allow more stable and precise evaluation of relationships among accessions and make it possible to detect patterns that are not visible through phenotype alone. This is particularly relevant in a crop such as caraway, where broad phenotypic and ecological diversity exists, but molecular data remain comparatively limited.

Within this context, the thesis introduces ISSR and SSR markers as the two main molecular systems selected for study. Both are widely used in plant diversity research, yet

they differ in their biological properties and practical strengths. ISSR markers are useful for broad polymorphism detection across the genome, while SSR markers are especially well suited to accession-level identification because of their locus specificity and reproducibility. The introduction therefore frames the thesis not simply as a diversity study, but as a broader investigation into the molecular characterization of caraway germplasm, including genetic diversity, population structure, accession discrimination, and molecular barcoding. A final important point in the introduction is that caraway has received less molecular attention than many major crop species, despite its agricultural, medicinal, and commercial value. Because of this, a deeper understanding of its genetic diversity and structure is needed to support both conservation and breeding. The thesis is therefore presented as a response to that gap. Its purpose is to generate molecular information that can help characterize a large and geographically diverse collection of caraway genetic resources and to provide a scientific basis for their future use in research, germplasm management, and crop improvement.

2. HYPOTHESES

The thesis is built upon three related hypotheses that define the central expectations of the research. The first hypothesis states that the analyzed set of caraway genetic resources exhibits a high level of molecular polymorphism detectable by both ISSR and SSR markers. This hypothesis is rooted in the broad origin of the germplasm collection and in the expectation that material collected from multiple countries, regions, and biological categories should contain substantial DNA-level variation.

The second hypothesis proposes that ISSR and SSR molecular markers provide sufficient resolution for the reliable discrimination of individual caraway accessions and for the assessment of genetic relationships among them. This hypothesis moves beyond the simple detection of polymorphism and addresses the practical value of the marker systems. In other words, it asks whether the selected markers are informative enough not only to show diversity, but also to distinguish accessions and define their genetic

relatedness with sufficient confidence for germplasm characterization and molecular fingerprinting.

The third hypothesis states that the genetic variation within the caraway collection is non-randomly distributed and that a distinct population and genetic structure related to accession origin or biological type can be identified. This is an especially important hypothesis because it links molecular diversity to biological interpretation. It asks whether the collection is simply diverse overall, or whether that diversity is patterned according to geography, domestication background, or germplasm category.

Taken together, the three hypotheses create a coherent framework for the thesis. They combine a descriptive question, how much polymorphism is present, with a methodological question, whether the markers are strong enough for discrimination, and a biological question, whether the detected variation has an interpretable structure. This makes the thesis both analytically focused and practically relevant.

3. AIMS OF THE THESIS

The main aim of the thesis was to assess the genetic diversity and population structure of caraway genetic resources using ISSR and SSR molecular markers and to generate molecular information useful for germplasm characterization, accession discrimination, conservation, and future breeding applications. This aim is broad but clearly targeted. It combines the evaluation of diversity with practical outputs relevant to gene banks and breeding systems.

To achieve this general aim, the thesis defined five specific objectives. The first objective was to evaluate the level of genetic diversity among selected caraway genetic resources using ISSR and SSR markers. This objective addressed the extent of polymorphism present in the collection and laid the basis for all later analyses.

The second objective was to determine the genetic relationships among the studied caraway accessions. This required pairwise comparisons, cluster analysis, and ordination

methods in order to describe similarity and divergence across the collection. The third objective was to analyze the population structure of the germplasm collection. Here the emphasis shifted from pairwise relationships to broader grouping patterns and to the question of whether the collection could be partitioned into genetically meaningful clusters.

The fourth objective was to identify genetically distinct and potentially redundant accessions within the studied material. This objective is highly relevant for genetic resource management because gene banks benefit from being able to identify duplicate material, rare diversity, and particularly valuable accessions. The fifth objective was to assess the usefulness of ISSR and SSR markers for the molecular characterization and discrimination of caraway genetic resources. This objective introduced the comparative dimension of the thesis and required both marker systems to be evaluated not only biologically but also methodologically.

The aims therefore show that the thesis was designed as more than a simple marker survey. It combined population genetics, marker evaluation, and practical germplasm management into a single research framework. This gave the thesis both scientific depth and applied value.

4. LITERATURE REVIEW

The literature review provides the scientific foundation for the thesis and is structured in a way that gradually moves from the crop itself to the broader conceptual and methodological tools used in the study. It begins with a botanical and agronomic overview of caraway and then expands into germplasm conservation, genetic diversity theory, molecular markers, molecular barcoding, and population structure.

The first part of the review discusses the taxonomy and biological characteristics of *Carum carvi*. Caraway is described as a member of the Apiaceae family, one of the largest and most economically important families of flowering plants. It is a biennial or annual

herbaceous species depending on genotype and flowering requirement, with broad distribution across Europe, North Africa, and Asia. This section of the review establishes the crop as biologically diverse and historically important, including its ancient cultivation history and its distribution across temperate agricultural systems. It also highlights its reproductive and morphological features, such as compound umbels, aromatic fruits, and contrasting annual and biennial life forms. These details are important because they help explain why caraway may show structured diversity linked to region, adaptation, and breeding history. The next part of the review focuses on the economic and agricultural importance of caraway. Caraway seeds and essential oil are widely used in the food, medicinal, and nutraceutical sectors, largely because of their bioactive compounds, especially carvone and limonene. The review notes that caraway is important in the spice trade and in several European agricultural systems, and that countries such as Finland, the Netherlands, Germany, and the Czech Republic play important roles in its production. This section also discusses the crop's agronomic relevance in diversified systems, including its fit in rotations and mixed production systems. At the same time, it points out several production challenges, such as the biennial growth habit, sensitivity to environment, and variable yield stability. By doing so, the literature review justifies the need for improved germplasm characterization and better breeding support.

A major component of the literature review is devoted to germplasm and plant genetic resources. Plant germplasm is described as the total heritable biological material of a species, including seeds, vegetative material, pollen, and tissue cultures. The review connects germplasm directly to genetic diversity, emphasizing that greater variation within a collection increases the chance of identifying useful alleles for crop improvement, adaptation, and resilience. It also explains the roles of *ex situ* and *in situ* conservation and highlights the growing risk of genetic erosion resulting from habitat destruction, climate change, and the replacement of traditional landraces by a narrower range of modern cultivars. In this context, caraway germplasm is presented as a strategically important resource whose conservation and characterization are essential for future breeding and for the maintenance of crop biodiversity. The literature review then moves into the concept of genetic diversity itself. Genetic diversity is presented as the total genetic variation within and among populations and as the basis of evolutionary adaptation and breeding

progress. The thesis review explains that diversity can be measured in many ways, including allelic richness, heterozygosity, and the proportion of polymorphic loci. It also discusses the biological factors affecting diversity, such as mutation, recombination, gene flow, drift, mating system, and environmental heterogeneity. Special attention is given to outcrossing species, which generally retain higher within-population diversity than selfing species. This is directly relevant to caraway, whose reproductive biology supports substantial gene flow and therefore a large proportion of within-group variation. The review uses this conceptual background to show why diversity studies in caraway should be expected to reveal both broad variation and relatively weak sharp boundaries between groups.

A further major section addresses molecular markers in plant diversity studies. Molecular markers are described as DNA-based tools that allow direct analysis of genetic variation and that are more stable than morphological markers because they are less affected by environment and development. The review surveys several marker systems, including RFLP, RAPD, AFLP, ISSR, SSR, and SNP-based approaches. It shows how PCR-based markers became especially influential because of their simplicity, speed, and reduced requirement for prior genomic information. In the case of caraway, earlier molecular studies relied largely on dominant markers such as RAPD, ISSR, and AFLP, while more recent work has begun using SNPs. However, SSR markers remain underused in caraway despite their value in related Apiaceae crops. This creates a clear rationale for the thesis, which applies SSRs more extensively to caraway germplasm and evaluates them alongside ISSR markers.

The review gives particular emphasis to ISSR and SSR markers. ISSR markers are described as useful in species with limited genomic resources because they amplify genomic regions between microsatellites without needing prior sequence information. They are relatively inexpensive, technically straightforward, and capable of detecting high polymorphism. However, they are dominant markers and therefore do not distinguish heterozygous from homozygous states. SSR markers, on the other hand, are described as microsatellites with codominant inheritance, high reproducibility, multiallelic variation, and strong discriminatory power. The review points out that SSRs have become a standard

tool in diversity studies and germplasm fingerprinting in many crops and that their development and transferability in Apiaceae species make them a promising option for caraway. This theoretical contrast between ISSR and SSR becomes one of the central comparative themes of the thesis. The literature review also discusses molecular barcoding and molecular fingerprinting in plants. It distinguishes between classical DNA barcoding, which is mainly used for species identification, and the more practical within-species fingerprinting used in crop germplasm research. In this practical sense, molecular barcoding refers to the use of multilocus marker profiles to generate unique molecular identities for cultivars or accessions. The review highlights that SSR markers are particularly suitable for this because they generate stable, accession-specific profiles. This section is important because it provides the conceptual justification for the thesis's emphasis on accession discrimination and its claim that the marker data can be used as a barcoding system for caraway germplasm.

Finally, the literature review addresses population structure. Population structure is defined as the non-random distribution of genetic variation within a species, usually due to geographic separation, partial reproductive isolation, selection, or historical processes. The review explains that understanding population structure is relevant in plant breeding because it helps identify distinct genetic groups, choose genetically complementary parents, reduce false marker–trait associations, and develop more representative core collections. For caraway, which has both wild and cultivated forms and a wide geographic range, population structure analysis is especially relevant for understanding whether genetic variation is organized by region, domestication background, or both. Overall, the literature review is comprehensive and functionally well integrated with the rest of the thesis, because each topic it introduces becomes directly relevant in the experimental design or later interpretation of results.

5. MATERIALS AND METHODS

The Materials and Methods chapter provides the experimental foundation of the thesis and is one of its strongest parts because it combines practical laboratory detail with a clear

analytical design. The study used a broad germplasm collection of 130 caraway accessions obtained from two Czech sources, the National Gene Bank of the Czech Republic and Agritec Plant Research Ltd. These accessions represented 24 countries and were grouped into six geographic regions: Central Europe, Western Europe, Northern Europe, Eastern Europe and the Former Soviet Union, Southern Europe, and Other, which included non-European material. The biological material categories included wild accessions, improved cultivars, breeding materials, and accessions of unknown status. This sampling design was suitable for testing both diversity and structure because it incorporated broad geographic and biological variation.

Before molecular work began, the thesis assessed seed viability through germination tests. Seeds were incubated on moistened filter paper in Petri dishes under controlled conditions. Germination was monitored for seven days, and the results were used to decide how each accession would be processed. Accessions with moderate or high germination were grown further to produce fresh young leaf tissue, while poorly germinating accessions were retained by extracting DNA directly from seeds. This strategy was practical and scientifically important because it ensured broad inclusion of the original germplasm collection rather than limiting analysis only to accessions with good germination. In studies of plant genetic resources, such decisions matter because excluding poor germinators can bias the final dataset toward better-preserved or more vigorous materials.

DNA extraction was performed using two complementary methods: a modified CTAB protocol and a commercial column-based extraction kit. The CTAB method served as the main approach, while the kit was used where CTAB extraction yielded poor-quality or insufficient DNA. DNA concentration and purity were measured using a NanoDrop spectrophotometer, and DNA integrity was verified by agarose gel electrophoresis. This dual extraction approach allowed the study to accommodate diverse source materials, including seeds and leaves, and to maximize the quality of downstream PCR amplification.

The thesis then optimized two different marker systems, SSR and ISSR. For SSR analysis, 47 primer pairs originally developed in related Apiaceae species were screened for

amplification success, reproducibility, and polymorphism. Gradient PCR was used to optimize annealing conditions. Thirteen primer pairs initially showed acceptable amplification, and after fragment analysis on the ABI 3500 Genetic Analyzer, 10 loci were retained for full-scale genotyping. These loci were evaluated using fluorescent labeling, capillary electrophoresis, internal size standards, and manual review of peak quality. The optimization also included empirical adjustment of dilution ratios because different loci produced different fluorescence intensities. This careful technical work was necessary for accurate SSR allele scoring and demonstrates that the thesis did not treat SSR genotyping as a routine procedure, but as a locus-specific analytical process.

For ISSR analysis, 16 primers from the UBC set were initially screened. Primers were evaluated for amplification efficiency, band clarity, polymorphism, and reproducibility between replicate PCR reactions. Eight primers were finally selected. All ISSR amplifications were performed in duplicate, and only bands that were reproduced in both independent reactions were scored. This is an important methodological strength because ISSR markers, as dominant markers scored from gels, are especially dependent on reproducibility. By requiring repeatable banding patterns, the thesis increased the reliability of the ISSR dataset and reduced the risk of overestimating diversity due to scoring artifacts.

A key methodological decision was the use of binary presence/absence matrices for both marker systems. For ISSR this is standard, since bands are inherently scored as present or absent. For SSR, binary scoring was adopted because several loci exhibited multi-band amplification behavior, a pattern known from transferred SSR markers in some Apiaceae crops. Instead of treating each primer pair as a simple codominant locus, the study scored all SSR allele classes as binary characters across the collection. This allowed both marker systems to be analyzed in a common comparative framework, especially for Jaccard distance analysis, UPGMA, PCoA, AMOVA, and Mantel testing. Although this approach reduces some classical codominant advantages of SSRs, it was justified by the actual amplification behavior observed in the dataset.

The quality filtering procedure was systematic and clearly defined. In the SSR dataset, accessions with zero scored alleles were removed, followed by accessions with very sparse profiles. In the ISSR dataset, accessions with too few scored bands were excluded for the same reason. The purpose was to ensure that Jaccard distances were calculated only from informative profiles with sufficient data. After filtering, the final SSR dataset consisted of 85 accessions, the final ISSR dataset of 105 accessions, and the overlap between them contained 76 accessions. This overlap set was especially important for the Mantel comparison between the two marker systems.

The analytical framework combined multiple complementary statistical methods. Diversity indices included the number of observed bands or allele classes, effective number of alleles, expected heterozygosity, Shannon's information index, and polymorphism information content. Pairwise genetic relationships were analyzed using the Jaccard coefficient and summarized by the Weighted Average Mean similarity. UPGMA clustering and Principal Coordinates Analysis were used to visualize genetic relationships. AMOVA partitioned variation among and within groups under two grouping criteria, geographic region and biological material type. Nei's G_{ST} measured regional differentiation. Population structure was inferred using a Gaussian Mixture Model after PCA reduction, and the optimal number of clusters was determined by BIC and Delta-K. Finally, a Mantel test quantified concordance between the SSR and ISSR distance matrices. This broad analytical design is a major strength of the thesis because it allows the results to be interpreted from several perspectives and reduces reliance on a single method.

6. RESULTS

The Results chapter is substantial and structured in a logical progression from plant material preparation to marker performance, diversity analysis, structure analysis, and accession discrimination. This organization makes it possible to follow how the final biological conclusions emerged from the raw material through stepwise filtering and increasingly complex analyses.

The first result concerns seed germination and plant growth. Germination varied strongly among accessions, confirming that the collection was heterogeneous not only genetically but also in seed performance. Based on the seven-day test, 29 accessions were classified as poor germinators, 42 as moderate, and 59 as excellent. Accessions with poor germination were processed by direct seed DNA extraction, while those with moderate or excellent germination were propagated further and supplied young leaf tissue for DNA isolation. These results are important because they show that the study retained accessions with very different viability profiles rather than excluding weak material from the beginning. This helped preserve the breadth of the original collection in the subsequent molecular analyses. The SSR optimization and amplification results showed that DNA of usable quality was obtained from the original accession set and that 10 SSR loci could be optimized for full-scale genotyping. The initial screening began with 47 SSR primer pairs, of which 13 amplified satisfactorily and 10 were eventually retained after evaluating fragment analysis quality. These 10 loci required different dilution ratios during electrophoretic optimization, indicating locus-specific variation in amplification strength and fluorescence signal. After quality filtering, the SSR dataset contained 85 accessions. All 10 selected loci were polymorphic across the collection. This demonstrates both technical success and biological informativeness of the selected SSR panel. The ISSR optimization and amplification results showed a parallel but distinct pattern. Sixteen ISSR primers were screened, and eight were retained on the basis of band clarity, reproducibility, and polymorphism. These eight primers generated a total of 71 scorable bands across the final ISSR dataset of 105 accessions. The banding profiles were sufficiently clear to permit reliable scoring, and the use of duplicate reactions ensured that only repeatable bands entered the matrix. These results confirm that the final ISSR panel was technically robust and suitable for diversity analysis.

The data quality filtering results are particularly important because they define the final datasets used for all downstream analyses. In the SSR dataset, 32 accessions were removed because they showed zero scored alleles and 13 more were removed because they carried fewer than five scored alleles. This left 85 accessions for analysis. In the ISSR dataset, 25 accessions with five or fewer scored bands were removed, leaving 105 accessions. No identical accession pairs were found within the retained SSR set, which

already suggested strong discriminatory potential of the SSR markers. The fact that 76 accessions passed the filters in both systems provided a sound basis for cross-marker comparison.

The characterization of ISSR markers showed that the eight primers generated 71 scorable bands, 59 of which were polymorphic at the selected frequency threshold. This corresponds to 83.1% polymorphism, a high value indicating that ISSR markers captured extensive variability in the caraway collection. Per-primer band numbers ranged from seven to twelve, and the most informative primers were ISSR-812 and ISSR-841, both of which reached 100% polymorphism. Across all ISSR bands, the mean expected heterozygosity was 0.321, mean Shannon's information index was 0.481, and mean PIC was 0.256. These values indicate that ISSR was highly effective in detecting polymorphism and in representing broad genome-wide variation in the collection. The SSR characterization results revealed a somewhat different pattern. The 10 SSR loci generated 83 binary allele bands across the 85 retained accessions. All loci were polymorphic at the locus level, but only 45 of the 83 allele bands met the frequency threshold for polymorphism, corresponding to 54.2%. Mean expected heterozygosity across the SSR bands was 0.191, mean Shannon's I was 0.304, and mean PIC was 0.158. Cumin 306 emerged as the most informative SSR marker, while GSSR 111 was the least informative. Compared with ISSR, SSR showed lower average per-band diversity values, but this did not reduce its usefulness; rather, it suggested that its strength lay more in accession-level discrimination than in broad per-character polymorphism.

The diversity analysis by geographic region produced one of the most meaningful results of the thesis. In the ISSR dataset, Central Europe had the highest number of observed bands, the highest diversity values, and the greatest number of private bands. Eastern/FSU and Western Europe also showed substantial diversity, while Southern Europe and the non-European category had lower total diversity due in part to small sample size. In the SSR dataset, Central Europe again had the highest diversity values and strong private-band contribution, while Western Europe contributed a notable number of private SSR bands relative to its size. This repeated pattern across both marker systems strongly supports the interpretation that Central Europe represents the primary diversity-rich region

in the analyzed collection. When accessions were grouped by biological material type, both marker systems showed that breeding material had the lowest diversity. In the ISSR dataset, improved cultivars and wild accessions displayed similarly high diversity values, whereas breeding material was markedly lower. In the SSR dataset, wild accessions were the most diverse among the well-defined groups, while breeding material again had the narrowest diversity profile. The unknown group often showed relatively high diversity and several private bands, which likely reflects heterogeneous origin. These findings suggest that breeding material has undergone some narrowing of the genetic base, while wild accessions and some improved cultivars still retain substantial variation.

The pairwise genetic similarity analysis showed broadly comparable overall divergence in the two datasets. The Weighted Average Mean similarity was 36.48% for SSR and 34.47% for ISSR. These values indicate substantial diversity in both marker systems and show that a randomly chosen pair of accessions is genetically rather dissimilar. One of the strongest applied findings is that all SSR-genotyped accessions were unique, with no identical multilocus fingerprints. This means the SSR panel achieved complete accession-level discrimination within the retained set. The UPGMA cluster analysis revealed broad, continuous patterns of genetic variation rather than a few highly discrete groups. The cophenetic correlation coefficient was 0.857 for SSR and 0.897 for ISSR, indicating good dendrogram fit in both systems and slightly stronger hierarchical structure in the ISSR dataset. Accessions from the same geographic regions were distributed across multiple clusters, though some partial association among neighboring European regions was visible. Non-European accessions tended to occupy more peripheral positions. Biological material types were intermixed throughout both trees, indicating no sharp molecular split between wild and cultivated material.

The Principal Coordinates Analysis supported the same interpretation. In the SSR dataset, the first three coordinates explained 40.2% of the total variation, while in the ISSR dataset they explained 46.0%. In both ordinations, non-European accessions were more clearly separated from the main European group, and Central European accessions occupied a broad region of the ordination space, consistent with their higher diversity. Again, accessions of different biological material types were intermixed. This result reinforced

the view that although some structure exists, the overall diversity pattern is broad and continuous rather than sharply partitioned. AMOVA revealed one of the most important contrasts between the two marker systems. In the SSR dataset, geographic region did not explain a significant fraction of the total variation, while biological material type did. In the ISSR dataset, both geographic region and biological material type were significant, with the geographic signal being detectable only in ISSR. In all AMOVA analyses, however, more than 90% of the variation occurred within groups rather than among them. This indicates that within-group diversity dominates in caraway and that regional or material-type structure exists but is relatively weak in comparison with the broad diversity present within groups.

Population structure analysis identified six major genetic clusters in both marker systems. The convergence of the optimal K value across SSR and ISSR is especially important because it suggests that both systems were capturing the same broad organization of diversity. At the same time, substantial admixture was observed, which is biologically consistent with the outcrossing nature of caraway and the absence of sharply discrete lineages. The Mantel test on the 76 shared accessions produced a significant positive correlation between the SSR and ISSR Jaccard distance matrices ($r = 0.257$; $P = 0.001$; 999 permutations), confirming concordance between the two marker systems. The final result of major applied importance was molecular barcoding and accession discrimination. The 10-locus SSR panel generated unique multilocus profiles for all 85 SSR-genotyped accessions, establishing complete accession-level discrimination. The ISSR panel alone did not fully separate all accessions in the larger ISSR dataset, but all indistinguishable ISSR pairs were resolved by SSR profiles. This means the dual-marker system, and especially the SSR component, can function as a molecular identity database for caraway germplasm.

7. DISCUSSION

The Discussion chapter interprets the results carefully and places them within the broader context of plant genetic resources, marker behavior, and caraway biology. One of the

strengths of the discussion is that it does not simply repeat the results; instead, it explains why the observed patterns are meaningful and how the two marker systems complement one another.

The discussion begins by returning to seed germination and plant establishment. Rather than treating germination only as a technical precondition for DNA extraction, the thesis interprets germination variability as a meaningful characteristic of the collection itself. The fact that the collection contained poor, moderate, and excellent germinators shows that the accessions differed not only genetically but also in seed lot performance and likely in physiological status. By retaining poorly germinating accessions through seed DNA extraction, the study avoided a major bias that could have favored only vigorous materials. This is important in germplasm research, where the goal is not to select the best-performing accessions at the start, but to characterize diversity as fully as possible. A major part of the discussion focuses on ISSR marker performance. The high percentage of polymorphic ISSR bands and the comparatively strong He, I, and PIC values are interpreted as evidence that ISSR markers sampled broad and informative genome-wide variability in caraway. The thesis discusses this in relation to earlier marker studies and argues that ISSR is particularly useful for species such as caraway where extensive marker resources have not historically been available. The significant geographic differentiation detected by ISSR is interpreted as a real biological signal rather than a statistical artifact, largely because ISSR markers surveyed more highly polymorphic genomic regions and provided stronger signal for subtle differences among regions. This leads to the conclusion that ISSR is especially informative for broad diversity detection and for identifying geographic structure in caraway germplasm.

The discussion of SSR markers highlights a different strength. Although SSR markers showed lower average diversity values per binary-scored allele band than ISSR, they provided much stronger accession-level discrimination. The thesis interprets this as a consequence of the marker system's ability to capture fine multilocus uniqueness, even when many bands occur at low or uneven frequencies. In practical terms, SSR markers were more effective for distinguishing individual accessions and for generating molecular fingerprints. This is why the thesis places particular emphasis on their value for molecular

barcoding. Rather than viewing the lower average H_e or PIC values as a weakness, the discussion argues that these values reflect the nature of the binary-scored SSR dataset and do not reduce its utility for accession identification. The contrast between the marker systems becomes especially important in the interpretation of AMOVA. ISSR detected significant geographic differentiation, while SSR did not. The thesis explains this difference in two main ways. First, ISSR had a higher proportion of polymorphic bands, which increased statistical power to detect subtle differentiation. Second, ISSR sampled broader genomic regions, while SSR was limited to ten defined loci. According to the discussion, this difference in genome coverage made ISSR more likely to capture regional allele frequency shifts. However, both marker systems agreed that the majority of variation was located within groups rather than among them. This agreement is treated as biologically meaningful and consistent with the outcrossing reproductive system of caraway, where high gene flow and broad within-group diversity are expected.

The discussion also addresses conservation and regional priorities. Because Central Europe consistently showed the highest diversity values and the highest contribution of private bands across both marker systems, it is interpreted as the principal center of diversity within the analyzed collection. Western Europe is also considered important because of its private-band contribution, and the non-European accessions are highlighted as especially distinct despite their small sample size. This leads to a practical conservation message: Central European wild material should receive high priority in ex situ and in situ conservation, and future sampling should be expanded particularly for the non-European component of the species range. The discussion thus converts analytical findings into concrete recommendations for genetic resource management. Another important interpretive point concerns biological material type and domestication. Significant differentiation among material categories was detected, yet ordinations and dendrograms showed complete intermixing of wild, improved, breeding, and unknown accessions. The thesis argues that these results are not contradictory. AMOVA reflects systematic frequency differences among groups, whereas ordination and clustering reflect overall relationships among individual accessions. The complete absence of a sharp cultivated-versus-wild boundary is interpreted as evidence that domestication in caraway has not produced a strong molecular divide. This is consistent with the relatively recent and

incomplete domestication history of the species and with the possibility that improved cultivars still contain substantial wild-derived variation.

The discussion also emphasizes the methodological importance of cross-marker concordance. The Mantel correlation between SSR and ISSR was moderate rather than very high, but the thesis argues that this is expected because the two systems sample different parts of the genome and differ fundamentally in marker type. More important than the absolute correlation value is the convergence of the major biological conclusions: both systems identified Central Europe as diversity-rich, both showed strong intermixing of material types, both supported six main genetic clusters, and both confirmed high within-group variation. The discussion therefore concludes that the two systems are concordant in their broad message, even though they differ in their sensitivity and resolution for different analytical tasks. The final major theme of the discussion is molecular barcoding. The thesis argues that the complete uniqueness of all SSR profiles establishes the 10-locus SSR panel as an effective molecular barcoding system for this caraway collection. In crop germplasm work, such a system has clear practical uses: verifying accession identity, detecting duplicates, supporting variety registration, and helping authenticate exchanged or commercial material. The discussion presents this as one of the most important applied contributions of the thesis. In this respect, the work goes beyond population genetics and contributes a tool directly relevant to gene bank management and crop improvement programs.

8. CONCLUSION

The thesis concludes that the analyzed caraway germplasm collection contains substantial molecular diversity and that this diversity can be effectively characterized using both ISSR and SSR marker systems. The broad geographic and biological scope of the collection made it possible to evaluate not only the level of diversity but also its distribution across regions and material types.

One of the central conclusions is that SSR and ISSR markers are complementary rather than redundant. ISSR markers were particularly effective for detecting high per-band polymorphism and revealing broader geographic structure. SSR markers, in contrast, showed outstanding power for accession-level discrimination and molecular fingerprinting. These differences demonstrate that the value of a marker system cannot be judged by a single diversity index alone; rather, each system contributes different strengths to a complete germplasm characterization strategy. Another major conclusion is that Central Europe represents the most important diversity-rich component of the analyzed collection. This conclusion is strongly supported by both marker systems and is one of the most biologically and practically meaningful findings of the study. Western Europe also contributes notable unique variation, while the small non-European group appears especially distinct and therefore important for future germplasm expansion. The thesis also concludes that most molecular variation in caraway is found within groups rather than among them. Although significant differences among regions and biological material types were detected in some analyses, especially with ISSR and in biological-material AMOVA, the overall pattern was one of broad continuous variation and strong intermixing. This is consistent with the biology of caraway as a predominantly outcrossing species and with the absence of a severe domestication bottleneck separating cultivated and wild germplasm.

Finally, the study establishes a strong molecular basis for future work on caraway genetic resources. It provides a validated comparative framework for SSR and ISSR analysis, identifies major diversity-rich groups, confirms the usefulness of dual-marker analysis, and demonstrates that the SSR panel can serve as a reliable molecular barcoding system for accession identification. Taken together, these achievements give the thesis both scientific significance and clear practical value for germplasm conservation, accession verification, and breeding-oriented use of *Carum carvi* resources.