# THESIS

Lad Priyankaben Rakeshkumar

2025

Hungarian University of Agriculture and Life Sciences

Institute of Genetics and Biotechnology, Buda Campus

MSC. In Agricultural Biotechnology

# Identification of basil species with chloroplast markers

Insider consultant: Zsuzsanna György PhD

Insider consultant's Institute/department

Department of Plant Biotechnology

Created by: Lad Priyankaben Rakeshkumar

Budapest, 2025

# Contents

ABBREVIATIONS ..........vi

1. INTRODUCTION ..........1

2. OBJECTIVE ..........2

3. LITERATURE REVIEW ..........3

3.1 Taxonomy and Distribution of Basil ..........3

3.2 Taxonomy ..........4

3.3 Morphology of *Ocimum* species ..........4

3.4 Economic and culinary importance of basil ..........6

3.5 Cultivation ..........7

3.6 Diseases ..........8

3.7 Breeding of basil ..........8

3.7.1. Chilling tolerance ..........9

3.7.2. Disease resistance ..........10

3.8 Basil essential oils ..........10

3.9 Basil Cultivars ..........11

3.10 Challenges in species identification using traditional methods ..........12

3.11 Molecular markers ..........13

3.12 Classification of Markers ..........13

3.12.1 Sequence Nonspecific Marker ..........14

3.12.1.1 Restriction Fragment Length Polymorphism (RFLP) ..........14

3.12.1.2 Random Amplified Polymorphic DNA (RAPD) ..........15

3.12.1.3 Amplified Fragment Length Polymorphism (AFLP) ..........15

3.12.1.4 Start Codon Targeted Marker Technique (SCoT) ..........15

3.12.2 Sequence Specific Markers ..........16

3.12.2.1 Simple Sequence Repeats (SSR) ..........16

3.12.2.2 Cleaved Amplified Polymorphic Sequences (CAPS) ..........16

3.12.2.3 Single Nucleotide Polymorphisms (SNPs) ..........17

3.12.3 Chloroplast DNA as a molecular marker ........ 17
3.12.3.1 Benefits of Plant Species Identification by Utilization of Chloroplast Markers ........ 18
3.12.3.2 Regions of Chloroplast DNA Most Widely used in Taxonomic studies of Plants ........ 19
3.12.3.3 Chloroplast Markers for Basil Species Identification ........ 20
3.12.3.4 Comparison of Their Efficiency in Differentiating Basil Species ........ 21
4. MATERIALS AND METHODS ........ 22
4.1 Material ........ 22
4.1.1 Plant Material ........ 22
4.1.2 DNA Extraction ........ 22
4.2 METHODS ........ 23
4.2.1 Polymerase Chain Reaction (PCR) ........ 23
4.2.2 Gel Electrophoresis and DNA cleaning ........ 25
4.2.3 Sequence Aliments and evolutionary study ........ 25
5. RESULT ........ 26
5.1 The amplification of chloroplast markers ........ 26
5.2 Cluster analysis ........ 28
5.2.1 Locus psbA5’R-matk8F (P2) ........ 28
5.2.2 Locus psbA – trnHGUG(P3) ........ 28
5.2.3 Locus psbB - psbH (P4) ........ 29
5.2.4 Locus rpl16F71 - rpl16R1516 (P5) ........ 30
5.2.5 Locus 5’rps12 - rpl20 (P6) ........ 31
5.2.6 Locus rps16F - rps16R (P8) ........ 32
5.2.7 Locus trnLUAAF - trnF (P9) ........ 33
5.2.8 Locus trnSUGA - trnGUCC (P10) ........ 34
5.2.9 Locus ycf6R (P12) ........ 35
5.3 Alignment and analysis of the concatenated chloroplast sequences ........ 36
6. DISCUSSION ........ 38
7 CONCLUSION ........ 40
8 SUMMARY ........ 41

ACKNOWLEDGMENTS ..... 42
BIBLIOGRAPHY ..... 43
DECLARATION ..... 47
STATEMENT ON CONSULTATION PRACTICES ..... 48

# ABBREVIATIONS

| | |
|---|---|
| AFLP | Amplified Fragment Length Polymorphism |
| BDM | Basil Downy Mildew |
| Bp | Base Pair |
| CAPS | Cleaved Amplified Polymorphic Sequences |
| DMSO | Dimethyl Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| dNTPs | Deoxynucleotide Triphosphates |
| ISSR | Inter Simple Sequence Repeats |
| MAS | Marker-Assisted Selection |
| NGS | Next-Generation Sequencing |
| PCoA | Principal Coordinate Analysis |
| QTLs | Quantitative Trait Loci |
| RAPD | Random Amplified Polymorphic DNA |
| RFLP | Restriction Fragment Length Polymorphism |
| Rp | Resolving Power |
| SCAR | Sequence Characterized Amplified Region |
| SCoT | Start Codon Targeted |
| SNP | Single Nucleotide Polymorphism |
| SSR | Simple Sequence Repeats |
| STS | Sequence Tagged Sites |

# 1. INTRODUCTION

Basil is a principal genus of the subfamily Nepetoideae in the Lamiaceae family, specifically the *Ocimum* genus. *Ocimum* is named after the Greek name "ozo," meaning a smell, and it is also known as the "king of herbs" as it has widespread application in the field of herbal medicine, perfumery, and the pharmaceutical industry.

Pushpangadan informed in 1995 that, in the *Ocimum* genus, there are around 160 species, the largest within the Lamiaceae family, and with them, 65 species are native to *Ocimum* and the rest are regarded as synonyms. Basil is a known taxon, which is considered to be wide in the taxonomic correspondence and thus makes the classification process difficult. This intricate situation is suggested to be the result of the genetic variability noticed in floral cross-pollination and environmental influences.

*Ocimum's* geographical existence is primarily focused on three areas of diversity such as tropical and subtropical parts of Asia, tropical Africa, and some from the tropical parts of America like Brazil, with altitudes usually around 1800 meters above the sea level.

Basil has many wild and cultivated, annual and perennial species distributed in the tropical and subtropical regions of Asia, Africa, Central and South America. Its application in nutritional, industrial and medicinal uses is due to the presence of biologically active compounds of antioxidant, insecticidal, nematicidal, fungistatic, and antimicrobial properties (Patel et al., 2015). Linalool, linalyl, geraniol, citral, camphor, eugenol, methyleugenol, methyl chavicol, methyl cinnamate, thymol, safrol, taxol, urosolic acid etc. are some of the remarkable metabolites which have been extracted from Ocimum species (Upadhyay et al., 2015).

It is utilized as perfumes and colognes in cosmetics and as culinary spices (Makri and Kintzios, 2008). Numbers of species in the genus reported in the literature variesy from 30 to 300 (Paton, 1992, Mahajan et al., 2015). The most popular species most used, utilized as spices and medicinal herbs, are *Ocimum basilicum, O. americanum (syn. O. canum), O. gratissimum, O. kilimandscharicum, O. tenuiflorum (syn. O. sanctum), and O. × citriodorum (syn. O. × africanum)*, which is the hybrid of *O. basilicum and O. americanum* (Paton et al., 1999, Carović-Stanko et al., 2010, Moghaddam et al., 2011). *O. tenuiflorum* (subtype Krishna) genome only has been sequenced and large gene set responsible for metabolite production has been found (Upadhyay et al., 2015).

## 2. OBJECTIVE

The main objective of this study was to assess the genetic variation within thirteen *Ocimum* species and cultivars using twelve chloroplast DNA loci. The study also sought out to understand if there could be any relationship between the species clusters and their morphological characteristics. The overall hypothesis was that there was a clear delimitation between the studied cultivars either based on the species or morphological traits.

# 3. LITERATURE REVIEW

## 3.1 Taxonomy and Distribution of Basil

Paton et al. (1999) estimates that roughly 65 kinds of basil shared globally are classified under one of the two tropical and subtropical regions along with the Family of Lamiaceae. Harley et al. (2004) discussed how polyploidy coupled with large genetic variation and different species hybridization made taxonomy identification for basil challenging. Work done by Simon et al. (1999) focused on the adoption of molecular markers instead of trait-based markers which struggled to differentiate closely related species. Basil is among the most cultivated plants and spices in the globe having major species constituents such as *Ocimum basilicum* or sweet basil popular in Mediterranean regions, United States, and Southeast Asia and in contrast, more scripted sites like Holy Basil or *Ocimum sanctum* celebrated medicinally by the people of India and Nepal counters India and Nepal (Kothari et al., 2004). *Ocimum gratissimum* is an African native known for its antimicrobial use while in rest of Africa Asia and the Americas *Ocimum americanum* grow in a wild or partial cultivated state (Vieira & Simon 2006). Carović-Stanko et al, (2010) noted how variation due to selection was becoming commonplace owing to the use of domestication and commercial farming aimed at increasing production of Basil.

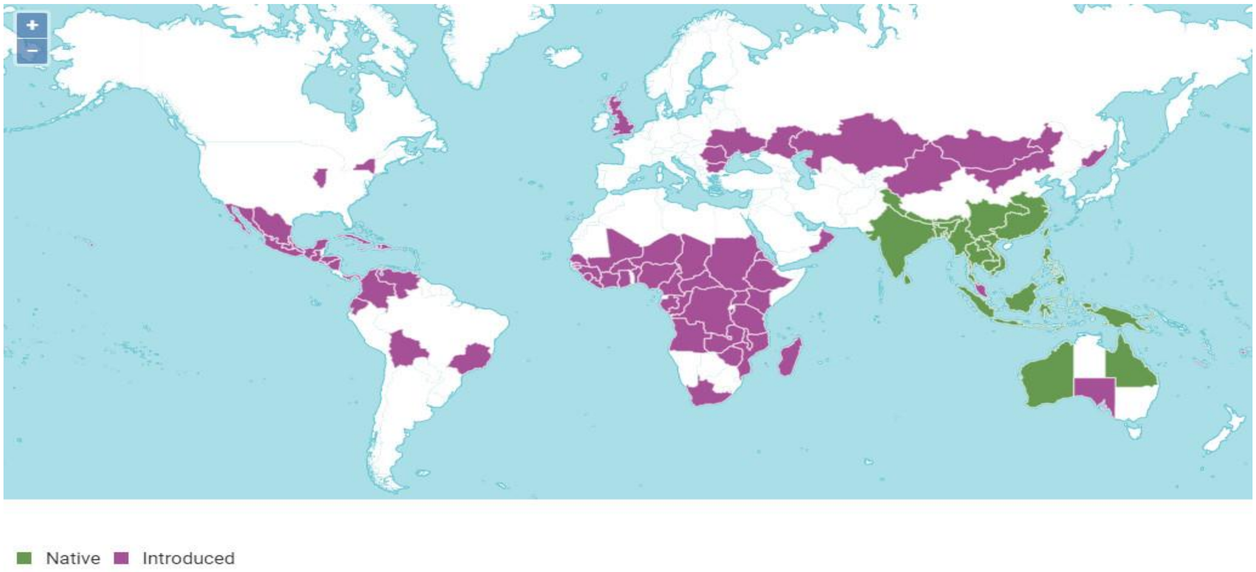

Figure 1. Worldwide Distribution of Basil

## 3.2 Taxonomy

*Ocimum* is classified as follows: it is in the Kingdom Plantae (Plants), in the Phylum Angiosperms (Flowering plants), and more precisely in the Class Eudicots (True dicotyledons). It is in the Order Lamiales, and is a part of the Family Lamiaceae (the Mint family). The genus *Ocimum* is in the Subfamily Nepetoideae and the Tribe Ocimeae, with subtribe Ociminae. The species *Ocimum* basilicum, or sweet basil, is a popular aromatic herb, extensively utilized in culinary and medicinal applications.

*Ocimum* is from the Ocimeae tribe, so it is not similar to such economic herbs as Rosmarinus and Salvia lives in the Labiatae family. Ocimeae tribe is different from the other tribes because of the declinate stamens they have. It means that in this type of flower, the stamen, which is the pollen-bearing organ, projects out laterally to the lower lip of the corolla, not in its center beneath the upper lip (Paton, 1999). Labiatae is a plant family that is very easy to recognize because of their opposite, bifid leaves with some reddish dots on them. The plants also have squareish stems.

Briquet (1897) has inserted *Ocimum* into the third subtribe of the Ocimeae tribe which is composed of only three subtribes. Afterwards, Ryding (1992) added the Ociminae subtribe to the previous two subtribes, thus making the number of the subtribes equal to three. He then placed *Ocimum* within the *Ocimum* group. In this paper, we will analyze cultivars of *Ocimum basilicum* that people usually call sweet basil. Some of the basil cultivars that are popular are Large Leaf, Osmin Basil, Mammoth Basil, Dark Opal, Red Rubin, Genovese, Lettuce Leaf, and Purple Ruffles. Among the countless cultivars and hybrids of basil, sweet basil is one of the most produced ones.

## 3.3 Morphology of *Ocimum* species

The growth form of *Ocimum basilicum* (basil), like other herbs in the family, is quite aromatic because of the essential oil content. This plant attains a height of 30 to 90 cm and has a stem that is square in nature. In addition to this, there are opposite, ovate, and lanceolate leaves on the stem. Some cultivars display purple or variegated leaves while some other cultivars have them in a finely serrate and glossy green color (Sivapalan et al., 2020).

*Ocimum basilicum L.*, or basil, is a scented annual herb belonging to the dicotyledonous family and having a diploid chromosome number of 48 (Prakash, 1990). The *Ocimum* genus is very variable in morphology (Aghaei et al., 2012). The basil flowers are of white, pink, or purple colors in racemes or spikes. The flower is bilabiate corolla with the lower lip being in three lobes and the upper lip in two. Basil flowers are hermaphroditic with both male and female reproductive organs, thus allowing pollination by bees, butterflies, and insects (Putievsky and Galambosi, 1999).

Basil leaves are found in opposite pairs, have a simple structure, and range from ovate to lance-shaped with serrated edges. Their overall shape is between 2–6 cm long and 1–3 cm wide. Leaves are shiny green, purple, or reddish, based on the variety. Cultivars also influence the general size and shape of the plant, with some plants being upright and more compact in shape while others are bushy (Hiltunen, 1999). Basil has small, dry fruit containing four extremely small, dark-colored seeds.

The androecium of the basil flower consists of four stamens, with two placed close to the corolla opening and two closer to its base. All the stamens are fertile except those in *O. drdnatum*, whose posterior stamens are sterile. While the anterior stamens are mostly free, united anterior pairs occur in some genera like Hemizygia and Syncolostemon. The corolla in plants like *O. campechianum* and *O. tenuiflorum* is extremely minute so that their sites of stamen attachment become difficult to compare with other species (Paton, 1999).

Regarding the gynoecium, the ovary in all the species is four-parted, subsequently developing into single-seeded nutlets called mericarps. Nutlets are of different shapes, elliptic in *O. basilicum* and spherical in *O. gratissimum*. Although smooth under normal circumstances, nutlets of certain species like *O. cufodontii* are pubescent. Secretion of mucilage by nutlets upon contact with water is different from species to species, where *O. basilicum* secretes abundant mucilage, *O. gratissimum* little, and *O. lamiifolium* nothing (Paton, 1999). Nutlets of Hemizygia and Syncolostemon possess a clear vein on the calyx-facing side (Ryding, 1992).

Most plants contain volatile essential oils which are responsible for the odor and these are often members of various plant families (Basile et al., 2016). This variation of *Ocimum* species morphology is due more to the cultivar or selection of the plant and environmental adaptation which is investigated for plant ecology and breeding (Chalchat & Hô, 1995).

Because of the great morphological diversity of *Ocimum*, it is time-consuming and inaccurate to distinguish species based on physical features alone. Hence, knowledge of the genetic diversity of various cultivars is crucial for scientific studies.

Figure 2. Morphology of *Ocimum basilicum* (basil)

## 3.4 Economic and culinary importance of basil

Basil (*Ocimum basilicum*) is considered a major cash crop in India, Italy and the United States. Apart from domestic uses, dried basil is exported as well. Basil is economically important because of its essential oils for perfumes and pharmaceuticals due to linalool, eugenol, and methyl chavicol compounds (Chalchat & Hô, 1995). Basil-flavored dishes like pesto, soups, sauces and salads are well known in Mediterranean and Southeast Asian regions. It is consumed dried and fresh. Italian cuisine shows the contribution of basil on a larger scale. Bianchini et al. (2018) stated its importance in improving the taste of traditional dishes. Rungrotmongkol et al. (2017) noticed a number of herbal treatments available with anti-inflammatory and antioxidant effects signifying basil's value. Therefore, both culinary and economic significance of basil is not only agricultural, but also industrial and gastronomic.

## 3.5 Cultivation

Basil has been extensively studied, with research expanding to areas of organic cultivation, nutrient optimization, and the numerous environmental conditions influencing basil growth within the various climatic zones (Putievsky and Galambosi, 1999). For commercial cultivation, basil is most frequently reproduced from seeds, either by direct sowing into the field or first grown in nurseries and subsequently planted. Basil seeds grown under laboratory conditions germinated in four days when humidity exceeded 80% and day and night temperatures were optimal, according to Putievsky (1983). Field germination periods were variable, taking 7–14 days in central Europe and 4–7 days in India.

Basil flourishes best in tropical conditions, when it is able to fully mature without the exposure to frost, which seriously retards its growth. Planting basil should only be done when the threat of frost has passed. In temperate climates, planting and transplanting should be carried in early summer for field planting. But in protected environments like greenhouses, planting cycles can be scheduled based on market demand (Putievsky, 1983).

Availability of nutrients is key in dictating basil quality and yield. The proportion of fertilization to be applied will depend on the current nutrient level of the soil. Most of the research done on fertilizer use has aimed at achieving an optimum ratio of three basic macronutrients, namely nitrogen (N), phosphorus (P), and potassium (K). A study conducted by Weichman (1948) in Germany concluded that an NPK ratio of 104:12:73 kg/ha would lead to optimum yields of basil. Of these nutrients, nitrogen was the most to affect yield the greatest. Subsequently, this effect was confirmed by Wahab and Hornok (1981), with increased amounts of nitrogen demonstrating to significantly elevate fresh as well as dry basil yields, an effect later reinforced by Hälvä and Puukka (1987).

Similarly, a Polish three-year trial by Czabajski (1978) demonstrated that nitrogen fertilizer increased dry basil yield by 44%. Nevertheless, the trial further indicated that additional phosphorus and potassium did not make any considerable contribution to yield. Surprisingly, even though increasing amounts of nitrogen increased yields, the effect plateaued beyond a certain point (Shahram & Omid, 2011).

While referred to as the "king of herbs," basil culture has challenges in relation to pests and disease that otherwise would have a significant impact on production.

## 3.6 Diseases

Several diseases can have a negative impact on basil growth, yield, and quality. Some of the most common basil diseases are:

1. Leaf Spot caused by Cercospora: Fungus *Cercospora ocimicola* causes it and appears as round to angular, dark-colored spots with light-colored centers on leaves.

2.Gray Mold: The cause of this disease is *Botrytis cinerea* and is seen as thick gray to brown fuzzy growth on stems, leaves, and dropped plant trash.

3.Leaf Spot & Root Rot: Bacterial leaf spot is caused by *Pseudomonas cichorii*, and root rot is caused by *Rhizoctonia solani*.

Fungal infections in basil can be managed by adopting drip irrigation and minimizing water splashes onto the infected leaves. Spraying a fungicide containing potassium bicarbonate weekly can also inhibit infections (Garibaldi et al., 1997).

## 3.7 Breeding of basil

Basil breeding has been an area of extensive study, focuss on improving characteristics such as disease resistance, yield, essential oil composition, and adaptation to fluctuating climates. Scientists have worked extensively towards developing superior cultivars with desired agronomic and phytochemical characteristics.

One of the key areas of basil breeding has been resistance to diseases, specifically against Basil Downy Mildew (*Peronospora belbahrii*). Wyenandt et al. (2015) researched genetic resistance of basil to this disease and worked on breeding resistant varieties. Similarly, Pyne et al. (2017) conducted research to breed basil lines that are more resistant to fungal and bacterial infections.

The other major breeding priority has been to enhance the content and quality of basil's essential oil. Simon et al. (1999) investigated the genetic diversity of basil to select high-yielding varieties with high essential oil content. Their research resulted in the development of aromatic varieties that are more suitable for food and medicinal use.

Further, Kiferle et al. (2011) researched basil morphological and physiological traits to generate varieties with improved growth performance in different environments. They enhanced traits such as plant architecture, leaf size, and biomass yield.

Advances in molecular breeding have also contributed to the enhancement of basil. Fracchiolla et al. (2020) utilized genetic markers to assist in selecting high-yielding basil cultivars with favorable traits such as improved stress tolerance and increased yields. These advances have significantly contributed to the commercial production of basil by ensuring the growth of robust and high-yielding varieties.

### 3.7.1. Chilling tolerance

Chilling tolerance is a trait that is sought after in basil, particularly in areas where temperature tends to dip sharply. Basil is a tropical plant that is highly sensitive to low temperature and frost, and these can critically affect its development and growth. Research on chilling tolerance in basil has been concerned with understanding the physiological and biochemical mechanisms that make the plant able to withstand cooler temperatures.

Kondrak et al. (2006) studied the impact of chilling stress on basil and found that cold tolerance in the plant was related to its ability to maintain membrane integrity and prevent oxidative damage. The research suggested that the enhancement of antioxidant synthesis would make basil more tolerant to mild chilling stress.

In addition, Mishra et al. (2012) explored the genetic regulation of chilling tolerance in basil and found potential cold stress response genes. They laid the foundation for the creation of basil cultivars with enhanced tolerance to lower temperatures, particularly in temperate regions.

Sánchez et al. (2016) also validated the understanding of chilling tolerance in basil, with emphasis on the significance of osmotic regulation in preventing freezing damage to the plant cells. By maintaining improved water and ion balance under stress at low temperatures, basil could survive and resume growth even in moderately cold temperatures.

### 3.7.2. Disease resistance

Basil resistance to disease is a pressing area of research to improve crop quality and yield. Basil suffers from several diseases such as Basil Downy Mildew (Peronospora belbahrii) and Gray Mold (Botrytis cinerea). Disease resistance has been of concern to most breeders.

Wyenandt et al. (2015) focused on developing Downy Mildew-resistant basil cultivars, one of the most devastating diseases in the United States. They focused on finding genetic resistance traits. Pyne et al. (2017) also examined resistance to numerous fungal and bacterial diseases through genetic screening, which improved the general health of basil plants.

## 3.8 Basil essential oils

Basil essential oils, which are derived from the flowers and leaves of *Ocimum basilicum*, are widely used in the perfume, pharmaceutical, and food industries due to their scented nature. The oils contain a variety of major aroma compounds such as citral, eugenol, linalool, methyl chavicol, and methyl cinnamate, which are traded globally in the essential oil industry (Quinn et al., 1990). Research has demonstrated the chemical makeup, bioactivity, and functional uses of basil essential oils widely and reported their diverse benefits.

The chemical composition of basil essential oils is a function of several factors, including seasonality, temperature, water and nutrient supply, and pest or disease pressure. The concentration of essential oil compounds is also influenced by genetic factors. The factors can yield different yields and qualities of oil, with synergistic, additive, or antagonistic activity between the compounds affecting its overall effectiveness (Araújo Couto et al., 2019).

1. Antioxidant Activity:

Antioxidant activity of basil essential oils is one of the prominent research areas. A research study was carried out in Brazil on 24 genotypes of basil and used various assays (DPPH, FRAP, ABTS+) to assess antioxidant activity. Basil essential oils containing eugenol and linalool were found to scavenge free radicals and provide protection against oxidative stress. Eugenol was found to be the most effective antioxidant among the compounds studied, and the presence of which may contribute to enhancing the antioxidant activity of basil oil. The interactions between the compounds may vary, however, due to the complexity of their action on antioxidant activity (Araújo Couto et al., 2019).

2. Anti-inflammatory Activity:

Basil essential oils have been found to be anti-inflammatory in nature primarily due to the flavonoids and phenolic compounds that are found within them. Khare (2008) and Akoto et al. (2020) have mentioned that basil consists of several bioactive compounds like flavonoids, phenolic acids, and essential fatty acids, which are the reason behind the anti-inflammatory activity of basil. In vitro studies, such as the Egg Albumen Denaturation Method, have established the plant to be inhibiting protein denaturation, a phenomenon linked with inflammation and diseases like rheumatoid arthritis. These compounds have been found to be inhibiting inflammation, thus making the application of basil oil in the treatment of inflammatory diseases (Mohammed et al., 2014) justifiable.

3. Antimicrobial Activity:

Research carried out in Thailand by Rattanachaikunsopon and Phumkhachorn (2010) revealed that basil's essential oil is rich in antibacterial and antifungal activity. Basil oil is highly inhibitory against pathogens such as Candida albicans, Escherichia coli, Staphylococcus aureus, and Pseudomonas aeruginosa. Notably, basil oil showed the highest rate of antibacterial activity against Salmonella enteritidis, which is the main foodborne causative agent of gastroenteritis. This antimicrobial activity renders basil essential oil a valuable natural compound employed in both pharmaceuticals and food preservation.

## 3.9 Basil Cultivars

Basil is a versatile herb with numerous cultivars, each of which possesses certain characteristics and applications. The most cultivated species of basil is Sweet Basil (*Ocimum basilicum*), which is well known for its sweet, fragrant flavor, used primarily in Italian cuisine. Thai Basil (*Ocimum basilicum var. thyrsiflora*) is used in Southeast Asia for its licorice flavor and is a staple in Thai cooking. Holy Basil (*Ocimum sanctum*), or Tulsi, is well known in India for its strong, clove-scented aroma and is also found to be extensively used in traditional medicine (Cohen, 2014).

Other popular types of basil include Purple Basil (*Ocimum basilicum* 'Purpurascens'), with dark purple leaves and a two-way purpose as an ornament and in cooking. Lemon Basil (*Ocimum x citriodorum*) has a strong lemon flavor and is commonly used in teas, desserts, and beverages.

Table 1. Information on *Ocimum basilicum* cultivars.

| **Category** | **Basil Cultivar** | **Description** |
|---|---|---|
| **Sweet Basil Varieties** | ‘Sweet’ | Classic basil with a sweet, aromatic flavor |
| | ‘Genovese’ | A popular Italian variety with a sweet, slightly spicy flavor, |
| | ‘Large Leaf’ | Known for its large, broad leaves, |
| | ‘Lettuce Leaf’ | Features large, tender, and crinkled leaves, |
| | ‘Mammoth’ | Large, bold leaves with a strong flavor, |
| **Purple Basil Varieties** | ‘Dark Opal’ | A purple variety with a strong flavor and ornamental appeal, |
| | ‘Purple Ruffles’ | Attractive, curly purple leaves with a mild, aromatic flavor |
| | ‘Red Rubin’ | Deep red leaves with a strong, slightly spicy flavor, |
| | ‘Osmin’ | A dark purple basil with a spicy and aromatic flavor |
| **Cultivars with Distinct Fragrances** | ‘Lemon Scented’ | Lemon-scented basil with a citrusy aroma, |
| | ‘Cinnamon’ | A basil with a spicy cinnamon aroma, |
| | ‘Anise’ | Has a sweet, licorice-like flavor, often used in Mediterranean and Asian cuisine. |
| | ‘Licorice’ | Characterized by a strong licorice scent, |
| | ‘Camphor’ | Known for its strong camphor-like scent, |
| | ‘Spicy’ | Offers a spicy, peppery aroma, typically used in savory dishes. |

## 3.10 Challenges in species identification using traditional methods

The classical techniques for identifying species through morphology have numerous limitations. One of the main problems is the plasticity of appearance, particularly that which is caused by environmental circumstances. Such phenomena render it impossible to separate strains that are, at first glance, rather similar (Lepage, 2012). Furthermore, a number of taxa experience pronounced changes in form with advances in ontogenetic stages, which renders precise identification problematic (Baker et al., 2003). Such techniques usually require some level of specialist judgment, which is often inadequate or too subjective, especially in highly biodiverse areas or places where taxonomic knowledge is limited (Schmidt-Lebuhn et al., 2016). On the other hand, overlap of characteristics such as flower color, shape of leaves, or even size is common among close relatives, therefore, the biodiversity is frequently reported even less than it exists. More troubling is the problem of separating distinct cryptic species (Pfenninger & Posada, 2002). All of these difficulties prompt the immediate use of additional methods such as molecular barcoding to enhance the effectiveness of species identification.

## 3.11 Molecular markers

A genetic marker is a recognizable piece of DNA which is generally located at a specific location of a genome and is therefore identifiable by a simple assay for the detection of nucleotide sequence polymorphisms between different specimens. The polymorphisms may be due primarily to mutations, including deletions, insertions, duplications, and translocations (Mondini et al., 2009). DNA markers which are applicable to the process of molecular-assisted selection (MAS) can significantly increase the efficiency and predictability of plant breeding. The recent advances in QTL research have identified many such marker-trait associations (Collard & Mackill, 2008). Molecular markers are typically classified as PCR-based markers and RFLPs (Restriction Fragment Length Polymorphisms), ISSR being included in the group of PCR-based markers.

## 3.12 Classification of Markers

Genetic markers can be categorized according to their type, inheritance mode, and molecular techniques involved in their detection. Morphological markers depend on observable characteristics such as flower color or seed shape but are subject to environmental influence. Biochemical markers entail protein variation, for example, enzyme variants, and are nowadays less utilized. Molecular markers, which are most common, are methods such as RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), SSR (Simple Sequence Repeats), SNP (Single Nucleotide Polymorphisms), and Indels (Insertions/Deletions). These markers are either dominant or codominant, and codominant markers are more informative since they can differentiate between alleles.

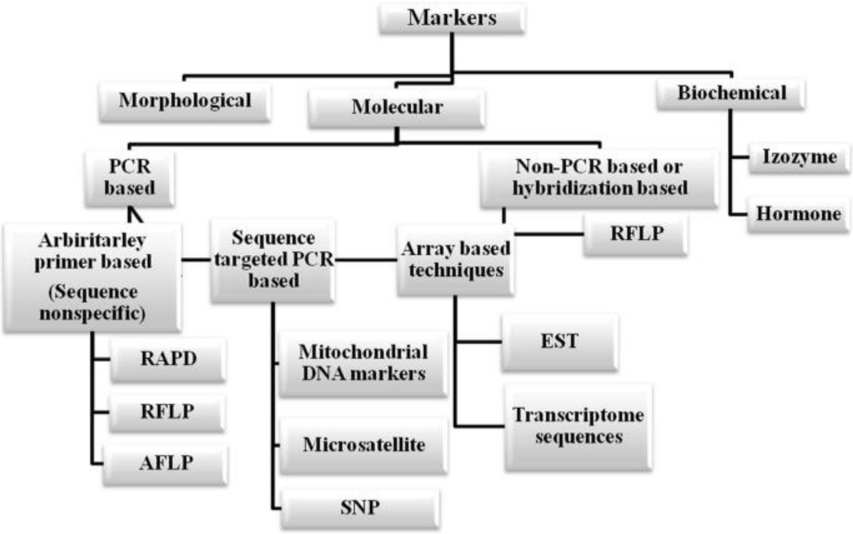

Figure 3. Classification Markers based on Morphology, Molecular and Biochemical. Kalia et al. (2011)

## 3.12.1 Sequence Nonspecific Marker

### 3.12.1.1 Restriction Fragment Length Polymorphism (RFLP)

RFLP, developed by Grodzicker et al. in 1974, was the initial DNA marker system and was the only one for about a decade. RFLP is a process of hybridizing digested DNA samples with restriction enzymes, like those from *E. coli*, and then performing Southern blotting. The digested DNA fragments are then placed on a membrane, where they are hybridized with labeled probes (e.g., radioactive or fluorescent) for detection. The limitations of RFLP are the need for a lot of DNA, the use of radioactive labeling, and the labor-intensive nature of the method itself (György, 2022). It is also able to identify other mutations like deletions, insertions, duplications, and translocations (Mondini et al., 2009).

With the help of marker-assisted selection (MAS), DNA markers have the ability to greatly increase the efficiency and precision of traditional plant breeding. The majority of recent studies of Quantitative Trait Loci (QTLs) have identified numerous marker-trait associations (Collard & Mackill, 2008).

#### 3.12.1.2 Random Amplified Polymorphic DNA (RAPD)

RAPD, initially established in 1990 by Williams et al., is an amplification method of genomic DNA using one random primer, typically 8-12 nucleotides in size. The success and yield of the PCR product depend on the primer and genome size and length and the GC content of the primer (ideally 40%). If the GC content is low, it may not bind well at 72°C in extension. The resulting products are then resolved in an agarose gel and stained with ethidium bromide to visualize the result. Polymorphisms may be identified by testing for the presence or absence of particular bands in electrophoresis.

The benefits of RAPD are that it does not require prior sequence data, can be applied to any species, is inexpensive, quick, simple, and does not involve labeled probes. The limitations are that it is low in reproducibility, non-specific annealing with short primers, challenging for PCR conditions, dominant in nature, and sensitive to DNA quality (György, 2022).

#### 3.12.1.3 Amplified Fragment Length Polymorphism (AFLP)

AFLP is a polymorphic and reproducible marker that does not require prior sequence information. It is a combination of RAPD and RFLP: total DNA is cleaved with two restriction endonucleases (a frequent cutter and an rare cutter), and restriction site-specific adaptors are then ligated to all fragments. PCR is performed, which amplifies part of these fragments employing adaptor specific primers. AFLP produces numerous fragments, and for the sake of clarity and reliability, visualization is usually obtained through Polyacrylamide Gel Electrophoresis (PAGE), although it's a more complex process and limitation of AFLP (Nadeem et al., 2017).

#### 3.12.1.4 Start Codon Targeted Marker Technique (SCoT)

The Start Codon Targeted (SCoT) marker technique, which was developed in 2009 by Collard and MacKill, is a novel technique specifically developed to identify genetic variation in plants. It is directed against a conserved region that is located near the ATG initiation codon in plant genes, a region which is well established in earlier studies (Joshi et al., 1997). SCoT is a dominant marker, which detects only the dominant allele at a locus, and is unable to differentiate between homozygotes and heterozygotes.

This technique utilizes single 18-mer primers for PCR, which leads to an annealing temperature of around 50°C, thereby causing higher specificity in primer binding than other random primers like RAPD. SCoT is also repeatable, easy, not laborious, and does not require prior sequence knowledge. SCoT is also time-efficient. The technique is convenient for most plant research laboratories because it can be performed with basic equipment and analyzed by routine agarose gel electrophoresis. SCoT is applicable in various fields, including phylogenetics, DNA fingerprinting, cultivar identification, and mapping of Quantitative Trait Loci (QTLs) in important crops such as sugarcane, grapes, potatoes, peanuts, and mangoes.

### 3.12.2 Sequence Specific Markers

#### 3.12.2.1 Simple Sequence Repeats (SSR)

SSRs or microsatellites are co-dominant markers which are of extreme use in population genetics research (Kumar 2021). First reported by Litt and Lutty in 1989 (Mishra et al., 2022), SSRs are found in the microsatellite regions of genomes. They are repetitive, short sequences (of 1–6 base pairs) in the eukaryotic DNA. The fact that there is variability in the number of such repeats makes SSRs extremely polymorphic, making them perfectly suited to distinguish individuals from each other. Furthermore, the surrounding regions of such repeats are conserved and, hence, primers can easily be designed locus-specific. These primers utilized in PCR are specific for such sequences. Visualization of the PCR product is typically carried out using PAGE or capillary electrophoresis, since differences between them are typically only a few nucleotides apart (György, 2022). The biggest benefits of SSR markers are their ability to process small quantities of DNA (10–100 ng), genomic frequency, high polymorphism, automation potential, and broad application.

#### 3.12.2.2 Cleaved Amplified Polymorphic Sequences (CAPS)

CAPS, or PCR-RFLP, employ endonucleases such as RFLP but with a distinction: in CAPS, the endonuclease only seeks out certain fragments of DNA, whereas in RFLP, it seeks out the entire genome. CAPS possess an important advantage, which is that they are extremely reproducible and co-dominant (György, 2022). The technique was first utilized in 1993 by Ausubel and Konieczny for the purpose of mapping genes within the Arabidopsis genome. CAPS primers are constructed from genomic database DNA sequences, cloned cDNA, or RAPD fragments. Being a chameleon marker, CAPS is versatile to pair with other markers like SCAR, AFLP, or RAPD to maximize its polymorphism detection potential (Nadeem et al., 2017).

#### 3.12.2.3 Single Nucleotide Polymorphisms (SNPs)

SNPs are the most common form of genetic variation, making them an ideal choice for the study of polymorphisms in populations. For a gene difference to be an SNP, it must be found at a frequency of 1% or higher among the population. To discover SNPs, one must know the entire genome sequence. SNPs are generally found *in vitro* but most frequently *in silico* through next-generation sequencing (NGS) data where transcripts are grouped and compared. In a comparative genome study of holy basil and sweet basil, 6,565 SNPs were detected, out of which 66.16% were transitions and 33.84% transversions (Rastogi et al., 2014). The *in-silico* approach is more cost-effective since it utilizes available information from existing open-source databases like dbEST or HarvEST. SNPs are co-dominant, bi-allelic, and genotypically compatible with a variety of technologies (György, 2022). Numerous platforms have been developed for genotyping of SNPs, the simplest one being SNP-RFLP. CAPS method is also utilized to identify SNPs (Nadeem et al., 2017).

### 3.12.3 Chloroplast DNA as a molecular marker

Chloroplast DNA (cpDNA), is found within the chloroplast of plant cells and is composed of a double-stranded, circular molecule. Nuclear DNA is biparentally inherited, but cpDNA is maternally inherited in most plants and, therefore, proves to be ideal for analysis of maternal lineages. Chloroplast and mitochondrial genomes are much smaller than nuclear genomes. Chloroplast DNA typically encodes for 100 to 120 genes of major functions like photosynthesis, protein synthesis, and chloroplast viability. Although architecture of chloroplast DNA is a mirror of drastic preservation across all plants, its certain parts have variability and therefore aid in specie identification and also in the study of phylogenetics.

### 3.12.3.1 Benefits of Plant Species Identification by Utilization of Chloroplast Markers

Maternal Lineage: Chloroplast DNA is largely maternally inherited that lowers hybridization and recombination issues with nuclear DNA, hence it becomes simple to identify maternal lineage (Rosenberg, 2013).

Conservation and Low Mutation Rate: Certain areas of cpDNA mutate at slower rates than the nuclear DNA, and therefore these areas are more conservative for species differentiation at higher taxonomic ranks (e.g., genus, family) than at the species level. However, conservation of cpDNA aids in resolving higher phylogenetic relationships (Zhang et al., 2016).

Even as sealing parts of a nuclear marker slows down mutation, preservation of some regions within cpDNA c is less affected by changes compared to the nuclear DNA, making it easy to differentiate species to higher taxa like family or genus. The nature of conservativeness of cpDNA is valuable in solving for larger scales of phylogenetic relationships (Zhang et al., 2016).

Being a more homogenous area compared to the rest of cpDNA, it is also more widespread in all plants to render it an excellent species marker for the identification of species, especially cryptic or diverged lineages (Ruhlman & Jansen, 2014).

The sensitivity and simplicity of PCR based techniques for the identification of target sequences in cpDNA make it even simpler because each plant cell contains multiple copies of targeted DNA sequence, so it is more probable to be identified by the PCR techniques.

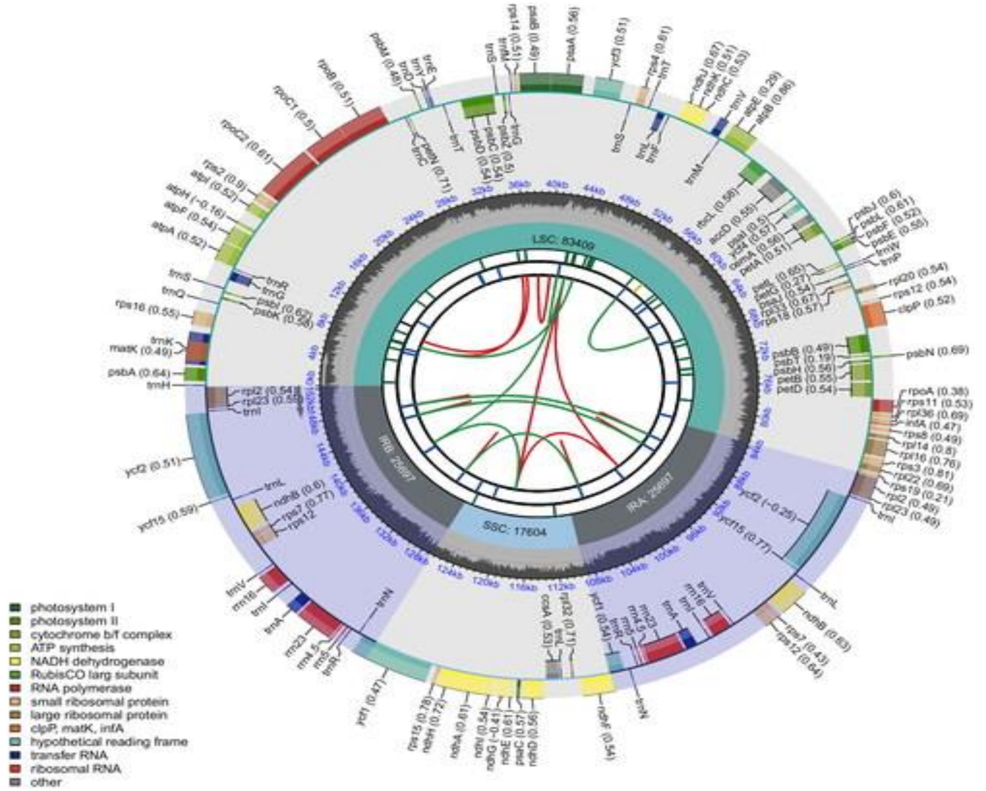

Figure 4. Circular map of the chloroplast genome of O. basilicum var. basilicum. From the center going outward, the first circle shows the distribution of the repeats connected with red (the forward direction) and green (the reverse direction) arcs. The second circle displays the tandem repeats marked with short bars. The third circle shows the LSC, SSC, IRa, and IRb regions. The fourth circle shows the percent of GC content. The next circle shows the genes having different colors based on the functional groups. The functional classification is shown at the bottom left. Genes inside the circle are transcribed in a clockwise direction, and those outside are in a counter-clockwise direction.

### 3.12.3.2 Regions of Chloroplast DNA Most Widely used in Taxonomic studies of Plants

1.*rbcL* Gene: The gene codes for the enzyme subunit Rubisco and is a great standard DNA barcode in plant identification because it is required by the overwhelming majority of species and it is simple to identify over a broad variety of different plant families (Kress et al., 2005).

2.*matK* Gene: *matK* gene use is common as a marker among plant species identification and phylogenetic research. It has been known that some regions of the gene are more divergent than the *rbcL* gene and are informative at lower taxonomic levels such as species and genus (Li et al., 2017).

3.*trnL-F* Intergenic Spacer: This two-gene spacer, *trnL* (Leu) and *trnF* (Phe), is often used for plant species differentiation as it differs among species but is fairly stable within species (Taberlet et al., 1991).

4.*psbA-trnH* Spacer Region: The intergenic spacer between the *psbA* and *trnH* genes is extremely variable in terms of species and has been used extensively for plant barcoding (Sang et al., 1997).

5.*ndhF* Gene: The *ndhF* gene encoding one of the subunits of the NADH dehydrogenase complex has been reported to be utilized in plant relationship studies, particularly in more recently diverged clades where other markers are too conserved at the sequence level to dissect phylogenetic relationships (Gielly & Taberlet, 1994).

### 3.12.3.3 Chloroplast Markers for Basil Species Identification

1.*matK* Gene: The *matK* gene is very variable and is therefore of value for species level identification and is commonly used in plant systematics. The gene produces a protein called the maturase that plays a role in splicing chloroplast RNA introns. *matK* has successfully been applied to a number of different dissected varieties and species of the genus *Ocimum* that, in comparison with other genes, have the highest variation and least conservation, respectively. *matK* is employed in most cases together with the other markers to enhance resolution. (Li et al., 2017).

2. *rbcL* Gene: The *rbcL* gene, the coding sequence for a ribsulose 1,5 bisphosphate carboxylase/oxygenase (Rubisco), is quite possibly the most universally used utilized chloroplast markers for plants used for initial identifications. Further, it is extremely divergent in other species, thus is perfectly suited to use in recognizing genera. Nonetheless, while helpful in the delimitation of higher taxonomic ranks, it cannot separate species close to a particular genus, for example, those belonging to the *Ocimum* because of low mutation rate in some of its sectors (Kress et al., 2005).

3. *trnH-psbA* Spacer Region: The intergenic spacer region of *trnH-psbA* has also received considerable prominence as a DNA barcode marker owing to its internal location between the

*trnH* (Leu) and *psbA* genes that is exposed to high frequency sequencing, hence being more variable than other loci. This makes the identification of species more precise, particularly in the case of narrow plant genera such as the *Ocimum* species complex. It has been successfully employed in the discrimination of different species of Ocimum Basilicum and clarification of cultivar uncertainties (Taberlet et al. 1991, Sang et al. 1997).

### 3.12.3.4 Comparison of Their Efficiency in Differentiating Basil Species

1.*matK* and *rbcL*: Even though both genes *matK* and *rbcL* have been effective in the recognition of members of a given plant genus, *matK* is more involved than *rbcL* in discrimination of species within the genus Basil, particularly among the closely related species. The highly repetitive regions of the *matK* gene are of greater variability giving matK greater resolution at the species level. In contrast, *rbcL* is too conservative among the *Ocimum* taxa and may possibly not be useful in discriminating certain extremely close-related taxa (Li et al. 2017).

2.*trnH-psbA* vs. *matK*: Because of its higher level of sequence divergence, the spacer region of *trnH-psbA* is also highly effective in distinguishing between basil species, especially when the use of the matK gene alone may be incapable of doing so. Combining *trnH-psbA* with *matK* can increase the overall Integrating *matK* with *trnH-psbA* can increase the accuracy of species identification in *Ocimum* (Sang et al., 1997).

# 4. MATERIALS AND METHODS

## 4.1 Material

### 4.1.1 Plant Material

Leaves of thirteen cultivars of *Ocimum* species were used as a source of DNA. Table 2. lists the applied genotypes giving their code, species and when possible, the cultivar name. The plant material has been collected in the experimental field of the MATE, Horticultural Institute, in Soroksár, Budapest in the summer of 2024.

Table 2. The *Ocimum* genotypes used in the study.

| Sample | Species Name | Cultivar Name |
|---|---|---|
| V1 | *Ocimum basilicum* | Purple Ruffles |
| V6 | *Ocimum basilicum* | Mrihani |
| V17 | *Ocimum sanctum* | Rama Tulsi |
| V18 | *Ocimum gratissimum* | Tulsi Vana |
| V19 | *Ocimum gratissimum/africanum* | African |
| V20 | *Ocimum africanum* | Penang Lemon (Pinang) |
| V21 | *Ocimum gratissimum* | African nunum |
| S7 | *Ocimum basilicum* | O.Sanctum Ethiopian |
| S12 | *Ocimum basilicum* | AdiF1 |
| S13 | *Ocimum americanum* | O.Americanum |
| S14 | *Ocimum basilicum* | Thai Basil |
| S18 | *Ocimum sanctum* | O.sanctumDanish |
| S24 | *Ocimum americanum* | O.Citronum lemon basil |

### 4.1.2 DNA Extraction

The fresh basil leaves which are collected were ground using sterile pestle and mortal in liquid nitrogen ready for further processing. The SP Plant DNA kit (Omega BIO-TEK Company, USA) was used for DNA Extraction. The isolated DNA was stored in a freezer at -20°C at the Buda Campus, Hungarian University of Agriculture and Life Sciences, Institute of Genetics and Biotechnology.

## 4.2 METHODS

### 4.2.1 Polymerase Chain Reaction (PCR)

The PCR mixture was prepared to a total volume of 45 µl for PCR thermocycling where a Swift MaxPro thermocycler (Esco Healthcare Pte, Singapore) was used. The following ingredients were used in the PCR mixture: 4.5 µl 10x PCR buffer, 0.9 µl dNTPs, 0.09 µl DreamTaq polymerase, 1 µl of Forward and Reverse primer,0.45 µl BSA, 0.9 µl DMSO, 35.16 µl sterile distilled water and finally 1 µl of extracted DNA sample. 20–80 ng of template DNA was used in each reaction. A standard PCR cycle was employed, consisting of an initial denaturation step at 94°C for 5 min, then 35 cycles of denaturation at 94°C, 30 seconds to 1 minute, after that the second step is annealing in between 50-65°C (according to the applied primers, see Table 3.), 30 seconds extension at 72°C for, 1 minute per kb of target DNA. At the end an extra 5 min of extension temperature was applied before, cooling the mixture to 4°C.

Table 3. Chloroplast marker Forward and Reverse Primer information.

| Primer Name | Sequence (5' to 3') | Fragment Length | Annealing temperature (°C) |
|---|---|---|---|
| P1 | *matk5* (F) – TGTCAATAACCTGCATTTCC<br>*matk6* (R) - TGGGTGGTGACTACAAATGG | 713 bp (tobacco) | 50 |
| P2 | *psbA5'F* (F) - AACCATCCAATGTAAAGCGGTTT<br>*matk8R* (R) - TCGACTTTCTGTAGAAGCTTT | 500 bp (tobacco) | 50 |
| P3 | *psbA (F*) - GTTATTGAAGCTGGAATGCTC<br>*trnHGUG (R)* - CGGCGCATGGTGATTCAACATCC | 454 bp (tobacco) | 53 |
| P4 | *psbB (F)* - TCCAANAANKGGAGATCCAAC<br>*psbH (R)* - TCAAYRGTTYGGTAGCCAT | 387 bp (tobacco) | 57 |
| P5 | *rpl16F71 (F)* - GCTATGCTAGTGTGATGCTGTTG<br>*rpl16R1516 (R)* - CCCTTCACTTTCTCCTCTAGTTG | 1021 bp (tobacco) | 50 |
| P6 | 5'*rps*12 (F) - ATTAAGAAANRGAAGCAGCCAAT<br>*rpl*20 (R) - CGYYAYGGAAGGAATGAACTC | 812 bp (tobacco) | 53 |
| P7 | *rpoB* (F) - CKAAANAYCYCTRAATTGG<br>*trn*CGCA R (R) - CACCCRGATTYGAATGGGG | 1282 bp (tobacco) | 53 |
| P8 | *rps*16F (F) - AAAAGTGGAATGTAARAACACATC<br>*rps*16R (R) - AACTACWATTGCAAGGATTGCATA | 861 bp (tobacco) | 53 |
| P9 | *trnL*UAAF (F) - CGAAATCGGTAGACGCTACG<br>*trnF* (R) - ATTGAACGCTACACTGAGCAG | 1015-547 bp (tobacco,rice, liverwort) | 50 |
| P10 | *trn*SUGA (F) - GAGAGAGGAGGATTGACGAC<br>*trn*GUCC (R) - ACCAAATTGAACATCGAACG | 867 bp (tobacco) | 57 |
| P11 | *trn*TGGU (F) - CTACCCTGAGTTAAAGGGG<br>*trn*CGCA F (R) - CCAGTTCRAATGYCGGTG | 750 bp (tobacco) | 57 |
| P12 | *ycf*6R (F) - GCCCAAGGACTCATACATTCA | 671 bp (tobacco) | 53 |

### 4.2.2 Gel Electrophoresis and DNA cleaning

For separation, all PCR amplification products were electrophoresed on 1% agarose gels made with TBE (Tris-borate EDTA) buffer, that were stained using 1% ethidium bromide (the staining was done in a laminar flow due to the hazardous nature of ethidium bromide). Gel was run in TBE buffer solution for 15 to 20 minutes at 100 volts.

Using a 1 kb bp DNA ladder as a reference, we evaluated the quality and size of the DNA fragments and visualized the DNA bands under UV light in Spectrophotometer.

For sequencing the successfully amplified PCR products were cleaned using the Thermo Scientific GeneJET Purification Kit. Sequencing was performed at a contracted partner in an automated sequencer ABI PRISM 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA). For each fragment, the nucleotide sequences were determined in both directions.

### 4.2.3 Sequence Aliments and evolutionary study

Forward and reverse sequences were edited and assembled and sequences were aligned using MEGAX manually (REF). Phylogenetic tree was constructed using the maximum likelihood approach with Bootstrap method using 500 replications also in MEGAX.

# 5. RESULT

## 5.1 The amplification of chloroplast markers

In course of my work, 13 basil genotypes were subjected to amplification using a set of 12 different chloroplast primers. Amplified fragments were checked on 1% TBE agarose gel (Figure5). Among them, 11 samples are amplified effectively within 9 of the 12. Fragments were sequences with Sanger sequencing technology.

Table 4 summarizes the amplification efficiency and polymorphism among the sequences gained in course my work. Primer pairs 1, 7 and 11 didn't amplified any products.

| Primers | Total amplified samples | Sequences length(bp) | SNPs | Indels |
|---|---|---|---|---|
| P1 | 0 | 0 | 0 | 0 |
| P2 | 10 | 839 | 21 | 27 |
| P3 | 8 | 375 | 10 | 11 |
| P4 | 9 | 615 | 26 | 6 |
| P5 | 8 | 877 | 11 | 3 |
| P6 | 10 | 782 | 22 | 8 |
| P7 | 0 | 0 | 0 | 0 |
| P8 | 12 | 843 | 35 | 41 |
| P9 | 6 | 838 | 29 | 24 |
| P10 | 7 | 746 | 24 | 22 |
| P11 | 0 | 0 | 0 | 0 |
| P12 | 11 | 213 | 2 | 1 |

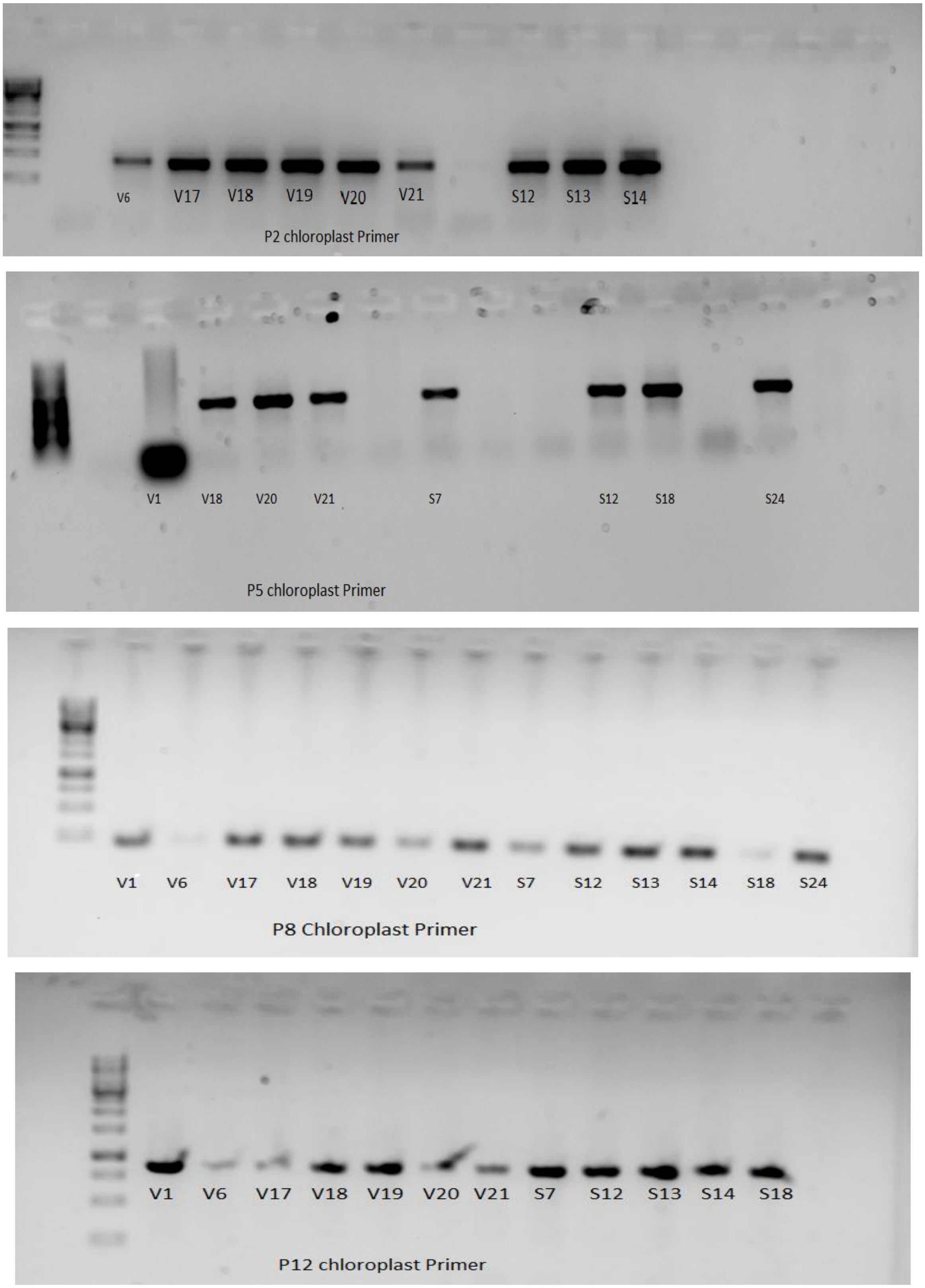

Figure 5. Electrophoresis gel Picture showing PCR amplification of chloroplast DNA loci P2, p5, P8 and P12 across different *Ocimum* species. To the left is 1k bp DNA ladder.

## 5.2 Cluster analysis

### 5.2.1 Locus psbA5’R-matk8F (P2)

Primer pair P2 successfully amplified ten sequences, and detected 21 SNPs and 27 indels were obtained. The sequence length was 839 bp, providing a good dataset for multiple sequence alignment that allowed variation between amplified fragments to be detected. These polymorphisms showed great diversity, and this Locus is therefore suitable for further genetic relationship analysis. From the aligned sequences, a dendrogram was constructed to reveal clustering of the variants and depicting well-differentiation in the amplified regions.

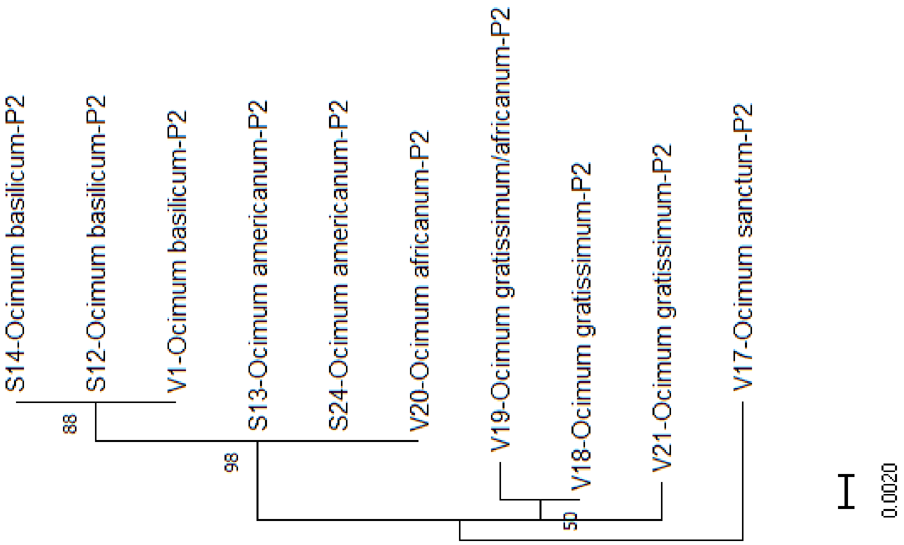

Figure 6. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1252.15) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. This analysis involved 10 nucleotide sequences. There was a total of 839 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

### 5.2.2 Locus psbA – trnHGUG(P3)

Primer pair psbA – trnHGUG(P3) amplified eight sequences, however, only seven could be successfully aligned, revealing 10 SNPs and 11 indels. The sequences obtained were 375 bp in length.

The length of the sequence was relatively shorter but sufficient for sequence alignment to detect polymorphic sites. This Primer pair was useful for genetic diversity analysis and was also found in the dendrogram representing moderate variation among the aligned sequences.

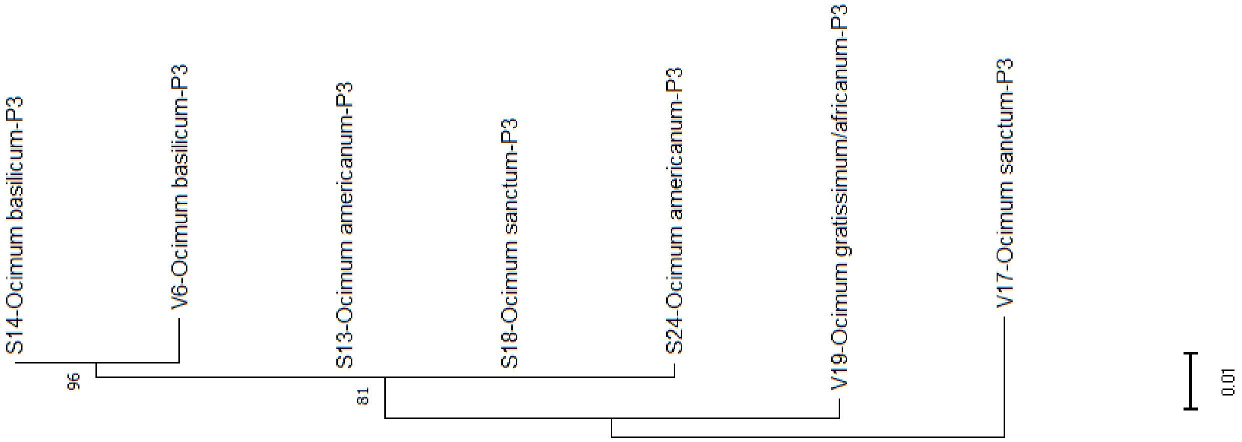

Figure 7. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-614.07) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. This analysis involved 7 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There was a total of 375 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

### 5.2.3 Locus psbB- psbH (P4)

Through nine positive amplifications, Primer pair P4 revealed a relatively high count of 26 SNPs and 6 indels, in a sequence range of 615 bp. The results reveal that Primer pair P4 is targeting a variable genomic region. Upon multiple sequence alignment, various polymorphic sites were identified, enhancing the resolution of genetic analysis.

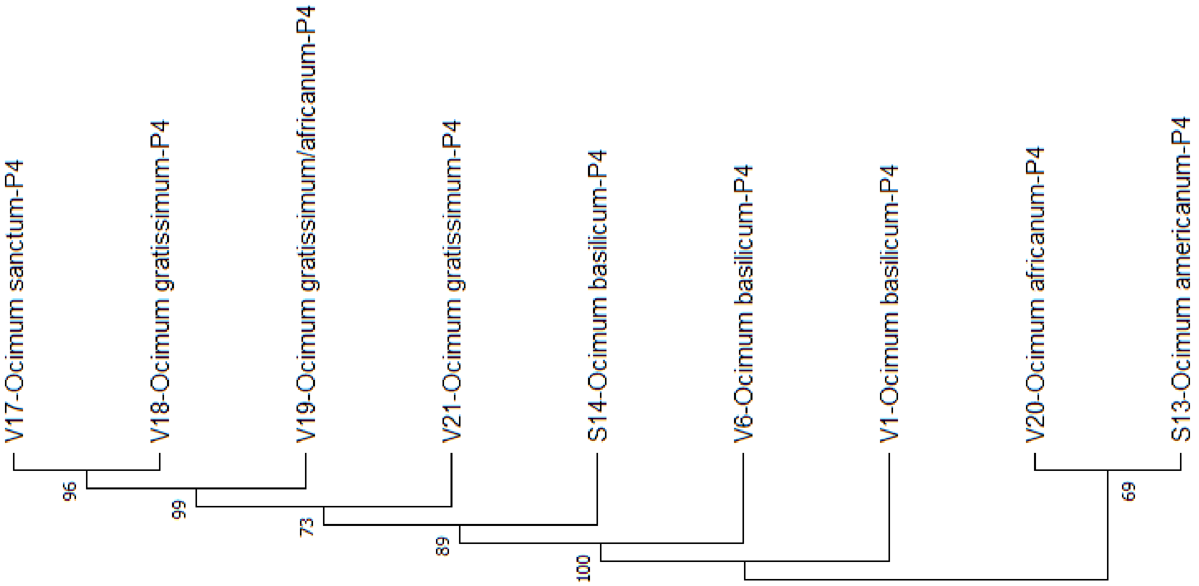

Figure 8. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-995.37) is shown. This analysis involved 9 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There was a total of 615 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

### 5.2.4 Locus rpl16F71- rpl16R1516 (P5)

Primer pair P5 produced 8 amplifications, revealing 11 SNPs and 3 indels from 877 bp long sequences. The high sequence length improved the accuracy of alignment and provided a clear visualization of nucleotide variation among samples. Although it had a relatively moderate level of polymorphism, it still contributed to the dendrogram by facilitating clustering of moderately variable sequences.

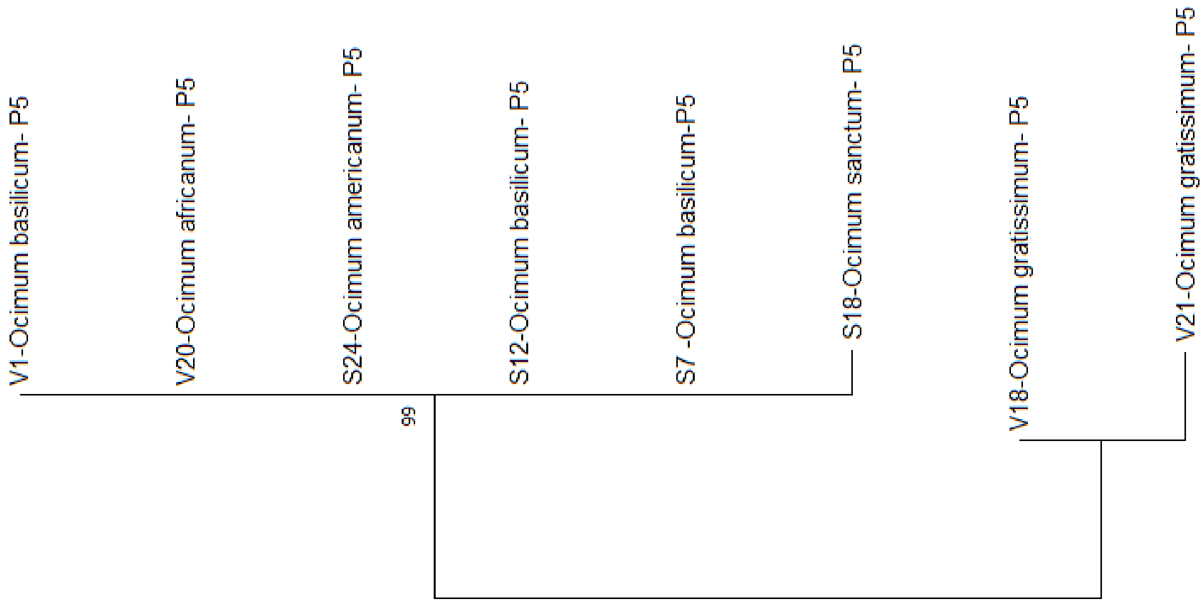

Figure 9. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1248.53) is shown. This analysis involved 8 nucleotide sequences. There was a total of 877 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

### 5.2.5 Locus 5’rps12- rpl20 (P6)

This primer pair amplified 10 sequences, with 22 SNPs and 8 indels, and an average length of 782 bp. Multiple sequence alignment revealed significant genetic diversity in the amplified segment, rendering Primer pair P6 useful for diversity analyses. It revealed high divergence in the dendrogram and clustering.

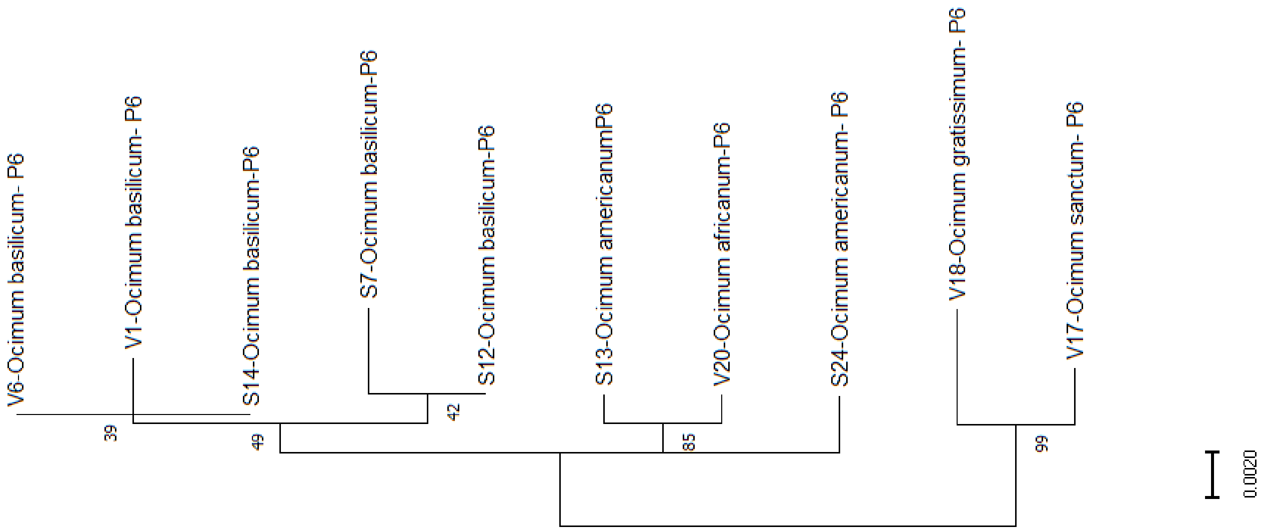

Figure 10. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1164.82) is shown. This analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There was a total of 782 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

### 5.2.6 Locus rps16F- rps16R (P8)

Primer pair P8 performed well, yielding 12 amplified sequences, and the maximum number of single nucleotide polymorphisms 35 and 41 indels, from sequences of varied lengths (843 bp). High polymorphism revealed by sequence alignment testifies to P8 being one of the most informative primers to investigate genetic diversity. Its ability to generate dissimilar clusters in the dendrogram shows high sequence variability.

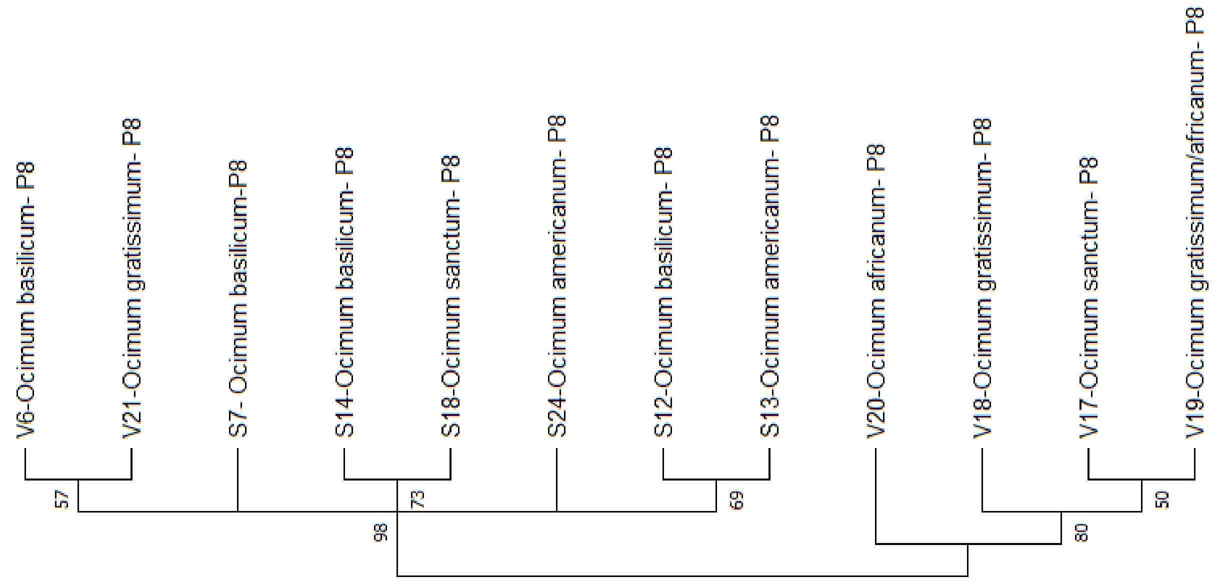

Figure 11. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1516.11) is shown. This analysis involved 12 nucleotide sequences. There was a total of 889 positions in the final dataset.

### 5.2.7 Locus trnLUAAF- trnF (P9)

With an average length of 838 bp, Primer pair P9 amplified 8 sequence fragments, depicting a long and informative region of DNA. In MEGA, sequence alignment revealed 29 SNPs and 24 indels, which indicates considerable genetic diversity. Using the aligned sequences, a dendrogram was constructed which displayed distinctive samples clustering and differentiating, thus proving that P9 is a robust marker for evaluating genetic variance in basil.

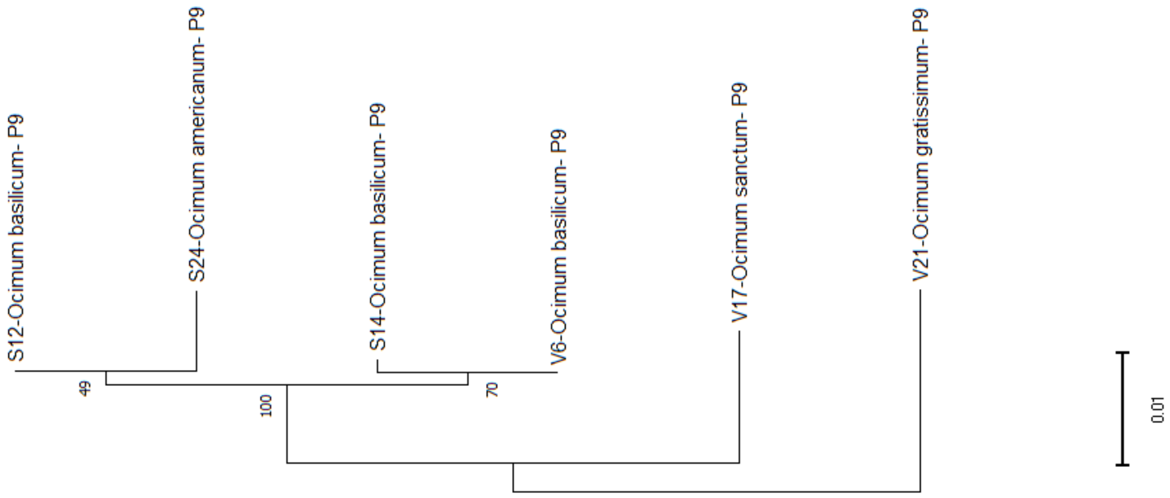

Figure 12. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1495.27) is shown. This analysis involved 6 nucleotide sequences. There was a total of 838 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

### 5.2.8 Locus trnSUGA- trnGUCC (P10)

Locus trnSUGA - trnGUCC (P10) did manage to amplify 7 sequences, the lowest among the successful primers, but still managed to discover 24 SNPs and 22 indels, high variation for a relatively small sample. This suggests that while it will amplify fewer fragments, those regions are rich in variation. Its size was 746 bp, a normal size. P10 can perhaps be optimized for more efficient amplification.

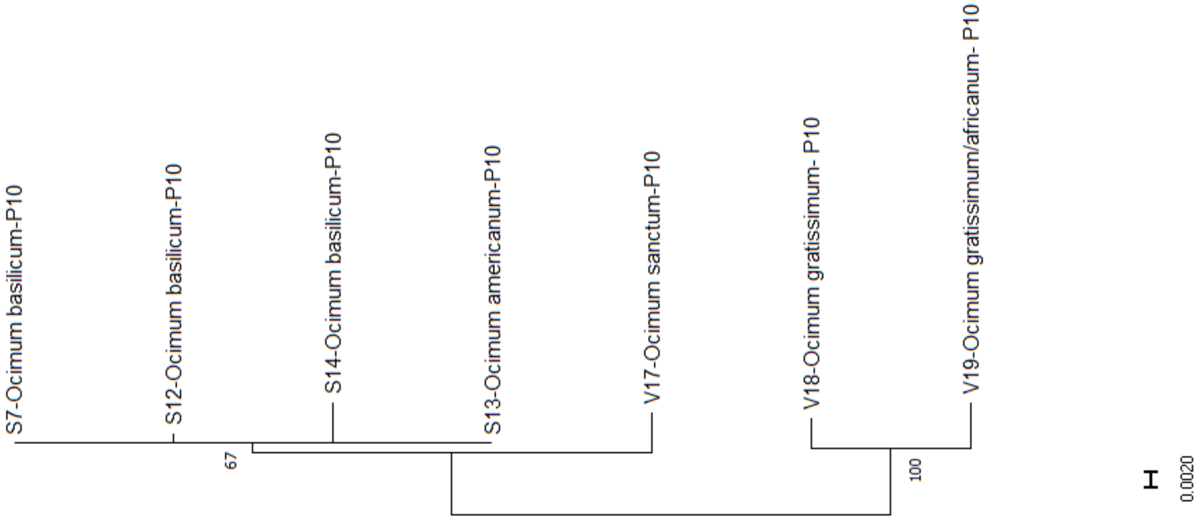

Figure 13. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-1240.14) is shown. This analysis involved 7 nucleotide sequences. There was a total of 746 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2]

### 5.2.9 Locus ycf6R (P12)

From Primer pair ycf6R (P12), 11 sequences were amplified which revealed 2 SNPs and 1 indel within a sequence length of 213 bp. While the number of amplifications achieved was high, the lack of polymorphism indicates that Primer P12 may target a more conserved region. It was observed that the alignment for the sequences showed little variation, this along with grouping with low diversity primers in the dendrogram indicated their limited applicability for genetic differentiation.

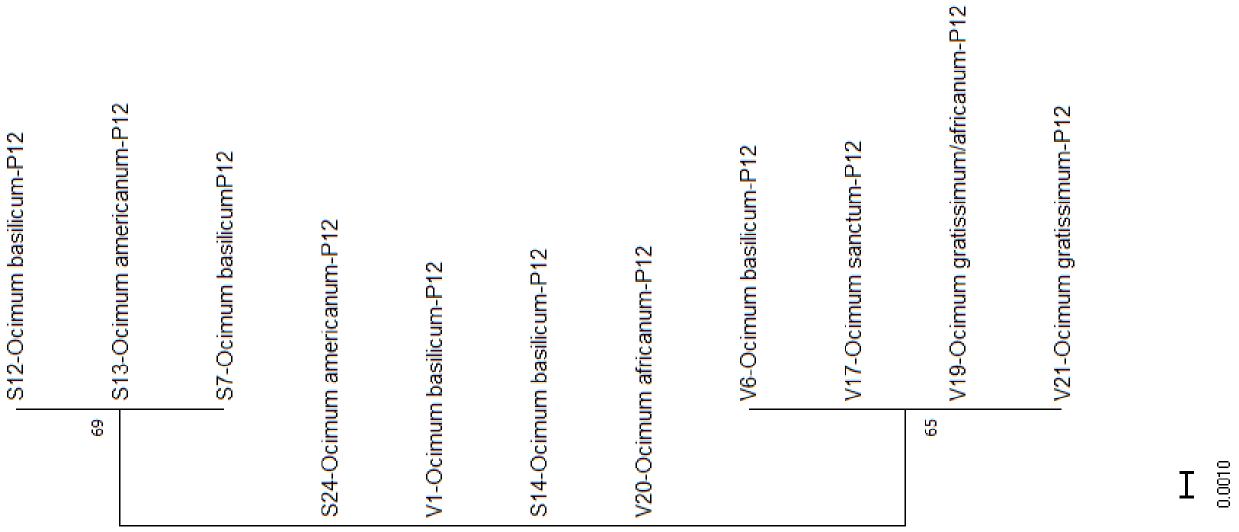

Figure 14. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-300.28) is shown. This analysis involved 11 nucleotide sequences. There was a total of 213 positions in the final dataset.

## 5.3 Alignment and analysis of the concatenated chloroplast sequences

The dendrogram based on the concatenated sequence alignment of six samples illustrates the evolutionary pattern between different species of the *Ocimum* genus, including *Ocimum basilicum, Ocimum americanum, Ocimum sanctum,* and *Ocimum gratissimum.* It revels 112 SNPs in 4079 bp sequence length. Every branch of the tree, as identified by "*S12-Ocimum basilicum*" or "*V18-Ocimum gratissimum*," is one sample or accession to a specific species. The dendrogram demonstrates that the samples belonging to the same species, for example, *O. basilicum* (S12 and S14) and *O. americanum* (S13 and S24) are tightly grouped, indicating a high level of genetic similarity within the species. These groups are well-supported clades with high bootstrap values of 100, indicating strong statistical support for these groupings. In contrast, *O. sanctum* (V17) and *O. gratissimum* (V18) are on single branches, indicating a higher level of genetic divergence from the rest of the species. It's indicating the close genetic relationships among species and also the evolutionary divergences among various *Ocimum* species and thereby gives information about their genetic diversity and evolutionary history.

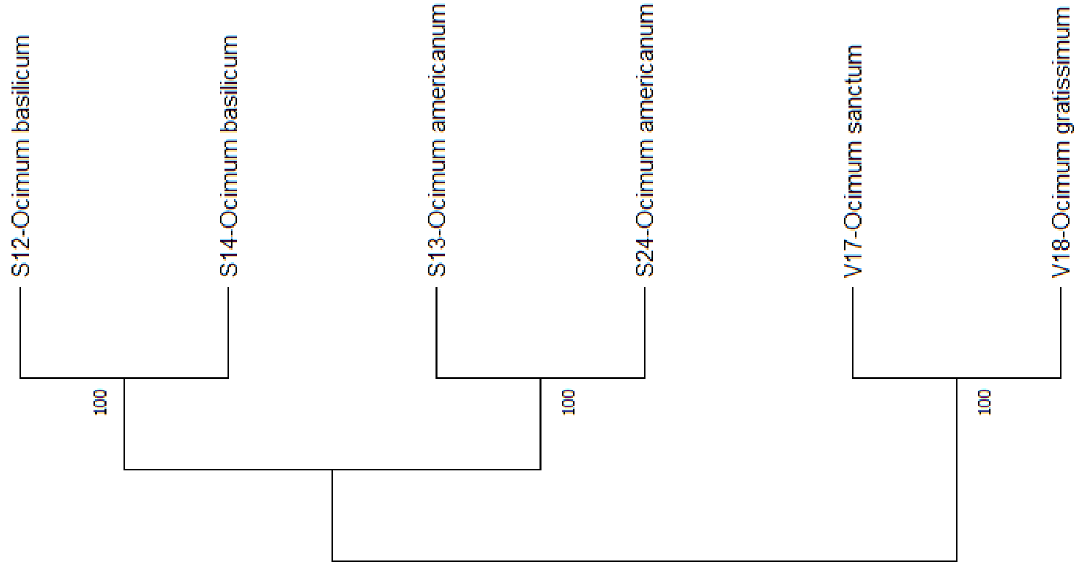

Figure 15. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-6597.65) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved 6 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated (complete deletion option). There was a total of 4076 positions in the final dataset.

## 6. DISCUSSION

The Molecular characterization of *Ocimum* species using various chloroplast regions (P1–P12) shows valuable information regarding the genetic diversity and relationship among the various species, namely *Ocimum basilicum, O. americanum, O. sanctum,* and *O. gratissimum*. Evidence from SNPs, indels, and sequence lengths reveals intra- as well as interspecific variation in support of their varied taxonomic statuses and probable evolutionary relationships.

Primers varied widely in their amplification success and the number of polymorphisms detected. Whereas locus matk5 - matk6(P1), locus rpoB - trnCGCA R(P7), and locus trnTGGU - trnCGCA F (P11) did not amplify any sequences, other locus like rps16F - rps16(P8) and psbA5' - matk8R (P2) amplified well with high SNP and indel numbers. Such differences indicate differential primer efficiency, which could be due to either sequence mismatches at primer binding sites or a gradient of sequence conservation across loci. Such locus like rps16F - rps16(P8) with 35 SNPs and 41 indels had high resolution power and are thus convenient in uncovering genetic variability in *Ocimum* species. The concatenated dendrogram demonstrates the evolutionary relationship among different species of the genus *Ocimum*, namely *O. basilicum, O. americanum, O. sanctum, and O. gratissimum*. Each label on the dendrogram, such as *Ocimum basilicum(S12)* and *Ocimum gratissimum*(V18), represents a specific sample or accession of the respective *Ocimum* species The phylogeny of the tree reveals that the identical species, e.g., *O. basilicum* (S12 and S14) and *O. americanum* (S13 and S24), exhibit a close clustering pattern, reflecting a high level of genetic similarity among the same species. The clusters form clear-cut clades that are characterized by high bootstrap values of 100, corresponding to high statistical support for the observed groupings. In contrast, *O. sanctum* (V17) and. *O. gratissimum* (V18) appears in isolated branches, indicating a significant genetic divergence from the other species. In general, the phylogenetic tree indicates the close genetic relationships within species and the evolutionary distances among various *Ocimum* species, thereby offering valuable information on their genetic diversity and evolutionary history. The results agree with a phylogeographic investigation of *O. tenuiflorum* in India that reported a highly diminished intraspecific genetic diversity using chloroplast DNA markers. Bast et al.'s (2014) research demonstrated that Indian *O. tenuiflorum* isolates formed a monophyletic clade and that North-Central Indian isolates were basal on the phylogenetic tree. The suggestion of this observation is that the region could be the center of origin for the species and that its lowered plastid diversity is a consequence of widespread human cultivation activities, minimized gene flow, and vegetative propagation via cuttings or selective seed populations. Chloroplast DNA markers have also been referred to as valuable tools in phylogenetic research. Paton et al.'s (2004) work demonstrated the utility of *trnL-trnF, rbcL*, and *matK* markers in investigating the phylogenetic relationships of the Ocimeae tribe and implicated the existence of polyphyly in *Ocimum*. Jena et al. (2009) and Vieira et al. (2001), also emphasizes the utility of chloroplast markers in resolving intricate phylogenetic relationships and genetic diversity in closely related plant groups, providing a complementary strategy to nuclear markers.

Utilization of chloroplast genetic markers enables more comprehensive insight into genetic diversity, evolutionary history, and biogeographical patterns throughout the genus *Ocimum*. The nuclear markers offer informative data on genetic diversity within and among species, whereas the chloroplast markers are more appropriate for examining evolutionary histories at the lineage level along with maternal inheritance characteristics. This multi-locus strategy is essential for comprehending taxonomic interactions and evolutionary mechanisms within the genus and holds great significance in terms of *Ocimum* species conservation and cultivation, especially those known for their medicinal and cultural importance, e.g., *O. tenuiflorum*.

# 7 CONCLUSION

In this study, the analysis of thirteenth *Ocimum* cultivars using twelve different chloroplast loci (P1-P12), provided valuable insights into the genetic diversity within and between species. Out of the twelve used for screening, only nine amplified DNA effectively, among which locus rps16F - rps16R (P8) and Locus psbA5'R-matk8F (P2) showed maximum amplification with 12 and 10 amplicons, respectively. These primers showed high polymorphism, with the highest SNP (35) and indel (41) counts. The sequence lengths were very varied, ranging from 213 bp to 877 bp, which reflected the genetic variation among the cultivars. The results indicated intra-species similarity that was very high, especially among *O. basilicum*, while inter-species, especially between *O. basilicum* and *O. gratissimum*, were moderate to low, confirming the distinct taxonomic identity of each species. These indicate the efficacy of primers chosen to detect genetic diversity, despite differing amplification efficiency and requiring careful scoring of gels.

# 8 SUMMARY

Basil (*Ocimum sp*.) is a well-established plant whose economic importance, agronomic value and ecological impetus is deserving research. To understand *Ocimum* species in an impactful approach, it`s clear that researchers have to define it`s taxonomic boundaries. Basil hails from the Lamiaceae family, it is documented to have about 50 to 150 species (Avetisyan et al. 2017). The taxonomical intricacies in basil are attributed to several factors like crossbreeding in basil species that allows gene flow (Avetisyan et al. 2017). Its uses span from pharmaceutical therapy, antimicrobial activity, culinary and also in the ornamental industry (Cohen, 2014). While there are numerous examples of basil species, the following are the most common *Ocimum sanctum, Ocimum gratissimum, Ocimum viride, Ocimum basilicum,* and *Ocimum americanum*. Different molecular markers present a plethora of novel platforms to significantly study the variation in species and cultivars. In this study, chloroplast Locus were the choice. My study had the following objectives; (i) to assess the genetic variation within thirteen *Ocimum* cultivars using twelve chloroplast loci. (ii) To understand if there could be any morphological resemblance between clustered species. The overall hypothesis was that there was a clear delimitation between the studied cultivars either based on the species or morphological traits. The plant materials used in this experiment were obtained from the MATE Institute of Horticulture gene bank. To get acquainted with the genetic diversity and the relationship among the species, a molecular analysis of the thirteen *Ocimum* cultivars was conducted applying chloroplast markers. The results showed the varying efficiency of the primer, where some of the primers are strongly effective at amplification, while others cannot amplify. Still, *Ocimum basilicum* varieties that proved to be similar genotypically amongst each other were found to be genetically different to a limited extent. However, the difference between *O. basilicum* and other species, such as *O. gratissimum*, represented the high level of their genetic similarity with lower rates of 45-60%, which clearly indicate that significant interspecific divergence has taken place. Also, *Ocimum sanctum* fell within the range of both high and low genetic similarity among the samples, thus indicating intraspecific variation or possible hybridization. Thus, noting the various traits of chloroplast markers in the change in the *Ocimum* genus. Therefore, the little gaps that came up within the genetic profiles of the species would mean a larger sample size and more replicates are what is needed in order to have clear genetic distinctions.

# ACKNOWLEDGMENTS

I would like to take this opportunity to extend my heartfelt thanks to the individuals who have supported me during my academic journey towards the completion of my Master's degree and the preparation of this thesis.

Above all, I would like to offer my sincerest gratitude to my supervisor, Dr. Zsuzsanna György, whose direction and constant encouragement throughout the last two years have been invaluable. Her sage counsel, patience of Job, and vast knowledge—always dispensed with a kind heart—have had a profound effect on my ability to achieve this important milestone. I am exceedingly thankful for the time, constructive criticism, and ongoing commitment she has devoted to my scholarly progress.

I would also like to thank Prof. Dr. Attila Hegedüs, Head of the Department of Plant Biotechnology, for his inspiring lectures that have motivated me since the beginning of the program. My sincere thanks also go to Dr. Júlia Halász for her dedication, both in teaching and in supporting our departmental activities with care and commitment.

There is a special mention for my classmates, whose collaboration, respect for each other, and encouragement throughout the practical sessions, group discussions, and thesis writing have enriched the experience more. I am forever thankful for the camaraderie we enjoyed while studying together.

# BIBLIOGRAPHY

1.Pushpangadan, P. (1995). ‘The Taxonomy, Distribution, and Uses of Ocimum Species.’ Indian Journal of Traditional Knowledge, 1(2), pp. 13–19.

2.Paton, A. J., et al. (1999). ‘A Review of the Genus Ocimum (Lamiaceae) in the Tropics: Diversity, Distribution, and Uses.’ Taxon, 48(2), pp. 407–421.

3.Harley, R. M., et al. (2004). ‘Polyploidy and Hybridization in the Taxonomy of Basil (Ocimum spp.): A Review.’ Plant Systematics and Evolution, 245(1), pp. 55–72.

4.Simon, J. E., et al. (1999). ‘Molecular Markers for Identifying and Differentiating Ocimum Species.’ Genetic Resources and Crop Evolution, 46(3), pp. 325–332.

5.Kothari, S. L., et al. (2004). ‘Medicinal Properties and Cultivation of Holy Basil (Ocimum sanctum) in India and Nepal.’ Indian Journal of Medicinal Plants, 27(1), pp. 21–29.

6.Vieira, A. R., & Simon, J. E. (2006). ‘Antimicrobial Properties of Ocimum Gratissimum and Other Basil Species.’ Phytotherapy Research, 20(6), pp. 517–522.

7.Carović-Stanko, K., et al. (2010). ‘The Domestication and Commercial Farming of Basil: Implications for Crop Production.’ Horticultural Science, 45(3), pp. 168–175.

8.Paton, A. J. (1999). Ocimum: An overview of classification and relationships. In R. Hiltunen & Y. Holm (Eds.), Basil: The genus Ocimum (pp. 1–38). Harwood Academic Publishers.Horizon e-Publishing Group

9.Briquet, J. (1897). Labiatae. In A. Engler & K. Prantl (Eds.), Die Natürlichen Pflanzenfamilien (Vol. IV, Part 3a, pp. 183–375). Wilhelm Engelmann.Horizon e-Publishing Group

10.Ryding, O. (1992). Pericarp structure and phylogeny of Lamiaceae subfamily Ocimoideae. Nordic Journal of Botany, 12(6), 689–710.

11.Sivapalan, S., Premarathna, A. D., & Weerakkody, W. A. P. (2020). Morphological characterization and essential oil composition of selected Sri Lankan basil (Ocimum basilicum L.) accessions. Journal of the National Science Foundation of Sri Lanka, 48(1), 63–74.

12.Bianchini, R., Tomi, F., & Casanova, J. (2018). Composition and chemical variability of the essential oil from leaves of Ocimum basilicum L. cultivated in Corsica (France). Chemistry & Biodiversity, 15(5), e1800068.

13.Basile, A., Sorbo, S., Giordano, S., Ricciardi, L., Ferrara, S., Montesano, D., Castaldo Cobianchi, R., Vuotto, M. L., & Ferrara, L. (2016). Antibacterial and antioxidant activities of Ocimum basilicum L. (Lamiaceae) essential oils. Journal of Ethnopharmacology, 104(1–2), 113–115.

14.Chalchat, J. C., & Hô, T. L. (1995). Essential oil composition of Ocimum basilicum L. and Ocimum minimum L. in relation to their morphological and anatomical characteristics. Scientia Pharmaceutica, 63(4), 375–384.

15.Rungrotmongkol, T., Pakotiprapha, D., Kokpol, S., & Hannongbua, S. (2017). Molecular dynamics simulations of eugenol derivatives in cyclooxygenase active sites: Insight into anti-inflammatory activity. Journal of Molecular Graphics and Modelling, 75, 319–327.

16.Mondini, L., Noorani, A., & Pagnotta, M. A. (2009). Assessing plant genetic diversity by molecular tools. Diversity, 1(1), 19-35.

17.Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 557-572.

18.Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ali, F., Chung, G., & Baloch, F. S. (2017). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing.

19.György, Z. (2022). Molecular markers in plant genetics and breeding. Advances in Plant Science and Biotechnology, 3(1), 45-67.

20.Grodzicker, T., Williams, J., Sharp, P. A., & Sambrook, J. (1974). Physical mapping of temperature-sensitive mutations of adenovirus. Cold Spring Harbor Symposia on Quantitative Biology, 39, 439-446.

21.Collard, B. C. Y., & MacKill, D. J. (2009). Start codon targeted (SCoT) markers: A novel class of markers for studying genetic diversity in plants. Molecular Breeding, 23(3), 73-79.

22.Joshi, C. P., et al. (1997). The sequence and organization of the chloroplast genome of Brassica napus and its comparison with other plants. Theoretical and Applied Genetics, 94(6-7), 757-764.

23.Kumar, S. (2021). Advances in molecular marker techniques for plant genetic studies. Plant Biotechnology Reports, 15(2), 195-209.

24.Litt, M., & Lutty, J. (1989). DNA fingerprinting and its application to plant breeding. Journal of Agricultural and Food Chemistry, 37(3), 715-722.

25.Mishra, M., et al. (2022). Use of microsatellites in the study of genetic diversity in plants. Plant Molecular Biology Reporter, 40(1), 36-47.

26.Nadeem, M. A., et al. (2017). Molecular markers in plant breeding: Tools for identifying and mapping genetic diversity. Plant Breeding Reviews, 41, 1-29.

27.György, J. (2022). Limitations and challenges of using molecular markers in plant breeding. Plant Molecular Biology and Biotechnology, 32(6), 729-741.

28.Ausubel, F. M., & Konieczny, A. (1993). The utility of cleaved amplified polymorphic sequences (CAPS) markers for genetic mapping in Arabidopsis thaliana. Plant Journal, 3(6), 587-598.

29.Rastogi, S., et al. (2014). SNP discovery in Ocimum species using de novo sequencing. Plant Science, 219, 45-52.

30.Rosenberg, M. (2013). Molecular markers and maternal inheritance. Plant Biology Journal, 50(1), 19-26.

31.Zhang, X., et al. (2016). Conserved regions of cpDNA for plant species identification and their role in taxonomy. Plant Genetic Resources, 14(4), 455-461.

32.Ruhlman, T. A., & Jansen, R. K. (2014). The plastid genome and its evolutionary dynamics. Current Opinion in Plant Biology, 18, 56-63.

33.Kress, W. J., et al. (2005). Plant DNA barcodes: The standardization of plant identification. Proceedings of the National Academy of Sciences, 102(23), 9155-9160.

34.Li, L., et al. (2017). Phylogenetic application of the matK gene in plants. Molecular Phylogenetics and Evolution, 113, 72-84.

35.Taberlet, P., et al. (1991). Universal primers for amplification of chloroplast DNA. Plant Molecular Biology Reporter, 9(3), 210-216.

36.Sang, T., et al. (1997). The psbA-trnH intergenic spacer: A widely used DNA barcode in plants. Molecular Ecology, 6(5), 727-732.

37.Gielly, L., & Taberlet, P. (1994). The ndhF gene as a marker for plant phylogenetic studies. Molecular Phylogenetics and Evolution, 3(2), 85-92.

# DECLARATION

on authenticity and public assess of final essay/thesis/mater's thesis/portfolio[1]

| | |
|---|---|
| Student's name: | Lad Priyankaben Rakeshkumar |
| Student's Neptun ID: | ELVTZJ |
| Title of the document: | Identification of basil species with chloroplast markers |
| Year of publication: | 2025 |
| Department: | Department of Plant Biotechnology |

I declare that the submitted final essay/thesis/master's thesis/portfolio[2] is my own, original individual creation. Any parts taken from an another author's work are clearly marked, and listed in the table of contents.

If the statements above are not true, I acknowledge that the Final examination board excludes me from participation in the final exam, and I am only allowed to take final exam if I submit another final essay/thesis/master's thesis/portfolio.

Viewing and printing my submitted work in a PDF format is permitted. However, the modification of my submitted work shall not be permitted.

I acknowledge that the rules on Intellectual Property Management of Hungarian University of Agriculture and Life Sciences shall apply to my work as an intellectulal property.

I acknowledge that the electric version of my work is uploaded to the repository sytem of the Hungarian University of Agriculture and Life Sciences.

Place and date: Budapest, 2025. April 15.

P.R.LAD

Student's signature

---

# STATEMENT ON CONSULTATION PRACTICES

Lad Priyankaben Rakeshkumar (student Neptun code: ELVTZJ) as a consultant, I declare that I have reviewed the final thesis/thesis/dissertation/portfolio1 and that I have informed the student of the requirements, legal and ethical rules for the correct handling of literary sources.

[✓ I recommend] / do not recommend2 the final thesis / dissertation / portfolio to be defended in the final examination.

The thesis contains a state or official secret: yes [✓ no]*3

Date: 2025 year April month 15. day

Insider consultant