DNA BARCODING OF ROBERT COLLEGE CAMPUSES' TREES: A CITIZEN SCIENCE INITIATIVE

by

Gizem Yılmaz

B.S. in Biology, Hacettepe University, 2013

Submitted to the Institute of Environmental Sciences

in partial fulfillment of the requirements for the degree of

Master of Science

in

Environmental Sciences

Boğaziçi University

2024

DNA BARCODING OF ROBERT COLLEGE CAMPUSES' TREES: A CITIZEN SCIENCE INITIATIVE

APPROVED BY:

Prof. Dr. Raşit Bilgin Thesis Advisor	. . . . .
Prof. Dr. Bahar İnce	. . . . .
Prof. Dr. Emre Keskin	. . . . .

DATE OF APPROVAL: 16/05/2024

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor Prof. Dr. Raşit Bilgin for giving me opportunity to study, conduct experiment with my students in his laboratory, for his continuing support, endless patience, and guidance. I would like to extend my gratitude to Aygün Karaca and Maha Salameh for their invaluable help in the laboratory.

I am so grateful to my friends and colleagues Dr. Özgül Yahyaoğlu, Gökhan Ergan, Pınar Kohlmeier, Dr. Cansu Demirel Turan, Zehra Toprak, Burçin Çıngay, Dr. Ece Kara Terzioğlu, Dünya Önen Ringwald and Burak Özçakır for their encouragement, vision and support each step of this study.

I am thankful to my brother Gürkan Yılmaz for his never-ending support, inspiration, and assistance through every challenge in this study journey.

Most importantly, I would like to express my appreciation to my students from Robert College DNA Barcoding of RC Plants Initiative. This thesis would not have been possible without them. I am so thankful to Defne Duman, Elif İyihuylu, Elif Doğa Ergüven, Beren Avcı, Elif Nida Kına, Zeynep İpek Kayalı, Mehveş Hatemi, Ezgi Zehra Ovacıklı, Ada Tanbulut, Berra Arı, Puren Akbulut, Rona Çetin, Zeynep Nehir Çankaya, Defne Vural, Nil Kibar, Defne Su Baysal, Öykü Erciyes, Defne Seyfi, Begüm Yener, Deren Bulut, Ceyda Yeşillik, Han Atlamaz, Salim Sinan Tuncer, Defne Şen, Ayşe Umay Bahçivan, Selin Usluer, Eylül Doğan, Zeynep Alara Şeremet, Nehir Yatağanbaba, Dura Taş, Nazlı Çam, Alya Kurt, Balım Şener, Berra Kıyak, Kaan Akan, Rona Çetin, Ece Ata for their contribution.

ABSTRACT

DNA BARCODING OF ROBERT COLLEGE CAMPUSES' TREES: A CITIZEN SCIENCE INITIATIVE

Urban ecology is directly related to how people interact with their environments. Monitoring the urban ecosystem requires rapid and precise identification of biodiversity, especially flora that is used as habitat for various species. DNA barcoding is a remarkable molecular technique to accurately identify species with their genetic information using the global barcode reference library. We designed this study as a citizen science project and used the DNA Barcoding method as an educational tool to identify tree species of Robert College campuses with their students. Fifty-two students from different grade levels were involved in various steps such as data collection, morphological identification, and DNA extraction. The Robert College campus is an amazing urban green space over 24 hectares on the Bosporus hills and is home to about sixty species of trees. The study aims to identify around fifty tree species with DNA Barcoding method and investigate their phylogenetic relationships. Although we had 100% PCR amplification success for fifty samples, five clean sequences were obtained for the matK genetic marker and twenty-two clean sequences for the rbcL sequence. The seven families and thirteen genera were identified as the *Juglandaceae, Fagaceae, Apocynaceae, Cupressaceae, Oleaceae, Lauraceae, Malvaceae* families and *Cercis, Fraxinus, Pistacia, Celtis, Ficus, Morus, Paulownia, Cinnamomum, Aucuba, Carpinus, Rubus, Quercus, Viburnum* genera. Distribution data for Istanbul in the literature was assessed and filtered using information from taxonomy and genetic identification. seventeen native and five imported taxa were found in twenty-two samples for Turkey.

ÖZET

BİR VATANDAŞ BİLİMİ GİRİŞİMİ: ROBERT KOLEJİ KAMPÜSLERİNDEKİ AĞAÇLARIN DNA BARKODLAMASI

Kentsel ekoloji, insanların çevreleriyle nasıl etkileşimde bulunduklarıyla doğrudan ilgilidir. Kentsel ekosistemi izlemek, çeşitli türler için habitat olarak kullanılan özellikle floranın hızlı ve kesin bir şekilde tanımlanmasını gerektirir. DNA barkodlama, küresel barkod referans kütüphanesini kullanarak türleri genetik bilgileriyle doğru bir şekilde tanımlamaya olanak tanıyan dikkate değer bir moleküler tekniktir. Bu çalışmayı bir vatandaş bilimi projesi olarak tasarladık ve Robert Koleji kampüslerindeki ağaç türlerini öğrencileriyle birlikte tanımlamak için DNA Barkodlama yöntemini bir eğitim aracı olarak kullandık. Farklı sınıf seviyelerinden kırk sekiz öğrenci, veri toplama, morfolojik tanımlama ve DNA ekstraksiyonu gibi araştırmanın çeşitli aşamalarında yer aldı. Robert Koleji kampüsü, yerel floranın kalıntılarının yanı sıra dünya çapında çeşitli bitki türlerini barındıran muhteşem bir kentsel yeşil alandır. Kampüs, Boğaziçi tepeleri üzerinde 24 hektarlık bir alanı kapsar ve yaklaşık altmış farklı ağaç türüne ev sahipliği yapar. Çalışma, DNA Barkodlama yöntemiyle yaklaşık elli ağaç türünü tanımlamayı ve onların filogenetik ilişkilerini araştırmayı amaçlamaktadır. Elli numune için %100 PCR amplifikasyon başarısı elde etmemize rağmen, matK gen bölgesi için beş temiz dizi ve rbcL sekansı için yirmi iki temiz dizi elde edildi. Yedi aile ve on üç cins Juglandaceae, Fagaceae, Apocynaceae, Cupressaceae, Oleaceae, Lauraceae, Malvaceae aileleri ve Cercis, Fraxinus, Pistacia, Celtis, Ficus, Morus, Paulownia, Cinnamomum, Aucuba, Carpinus, Rubus, Quercus, Viburnum cinsleri olarak tanımlandı. Literatürde yer alan İstanbul için dağılım verileri değerlendirilmiş ve taksonomi ve genetik tanımlama bilgileri kullanılarak filtrelenmiştir. İstanbul'a ait bitki dağılım verileri ile taksonomi ve genetik tanımlamadan elde edilen bilgiler kullanılarak değerlendirildi ve filtrelendi. Türkiye için yirmi iki örnekde on yedi yerli ve beş yabancı kökenli takson bulunmuştur.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

ÖZET v

LIST OF FIGURES viii

LIST OF TABLES ..... **Hata! Yer işareti tanımlanmamış.**

LIST OF SYMBOLS/ABBREVIATIONS xii

1. INTRODUCTION 1

2. LITERATURE REVIEW 8

3. MATERIALS AND METHODS 12

3.1. Sampling 12

3.2. Morphological Identification 12

3.3 Molecular Studies 12

3.4 PCR Amplification 12

3.5. Gel Electrophoresis 13

3.6. Sequencing and Species Identification 13

4. RESULTS AND DISCUSSION 14

4.1. Cercis 17

4.2. Fraxinus 19

4.3. Pistacia 19

4.4. Celtis 23

4.5. Ficus 23

4.6. Morus 24

4.7. Paulownia 25

4.8. Cinnamomum 25

4.9. Juglandaceae 34

4.10. Fagaceae 37
4.11. Aucuba.....38
4.12. Carpinus.....38
4.13. Rubus.....39
4.14. Apocynaceae.....48
4.15. Cupressaceae 50
4.16. Oleaceae.....50
4.17. Lauraceae.....53
4.18. Malvaceae.....53
4.19. Quercus.....54
4.20. Viburnum.....54
5. CONCLUSION AND RECOMMENDATIONS.....60
6. REFERENCES.....61

LIST OF FIGURES

Figure 1.1. Result of DNA Plant Barcoding for publication years between 2004 to August 2023. ...3

Figure 1.2. Result of DNA Barcoding and Citizen Science for publication years between 2010 to August 2023. 5

Figure 4.1. The tree-based identification of sample 42. Sample 42 is highlighted in yellow. 18

Figure 4.2. The tree-based identification of sample 3. Sample 3 is highlighted in yellow. 21

Figure 4.3. The tree-based identification of sample 4. Sample 4 is highlighted in yellow 22

Figure 4.4. The tree-based identification of sample 5 for the rbcL marker. Sample 5 is highlighted in yellow.....27

Figure 4.5. The tree-based identification of sample 5 for the matK marker. Sample 5 is highlighted in yellow.....28

Figure 4.6. The tree-based identification of sample 7 for the matK marker. Sample 7 is highlighted in yellow.....29

Figure 4.7. The tree-based identification of sample 7 for the rbcL marker. Sample 7 is highlighted in yellow.....30

Figure 4.8. The tree-based identification of sample 8 for the rbcL gene region. Sample 8 is highlighted in yellow. 31

Figure 4.9. The tree-based identification of sample 9 for the rbcL gene sequences. Sample 9 is highlighted in yellow. 32

Figure 4.10. The tree-based identification of sample 10 for the rbcL gene region. Sample 10 is highlighted in yellow. 33

Figure 4.11. The tree-based identification of sample 11 for the rbcL sequence. Sample 11 is highlighted in yellow. 35

Figure 4.13. The tree-based identification of sample 12 for the rbcL marker. Sample 12 is highlighted in yellow. 40

Figure 4.14. The tree-based identification of sample 13 for the rbcL marker. Sample 13 is highlighted in yellow. 41

Figure 4.15. The tree-based identification of sample 13 for the matK marker. Sample 13 is highlighted in yellow. 42

Figure 4.16. The tree-based identification of sample 15 for the rbcL gene region. Sample 15 is highlighted in yellow. 43

Figure 4.17. The tree-based identification of sample 15 for the matK gene region. Sample 15 is highlighted in yellow. 45

Figure 4.18. The tree-based identification of sample 16 for the rbcL marker. Sample 16 is highlighted in yellow. 45

Figure 4.19. The tree-based identification of sample 18 for the rbcL gene sequences. Sample 18 is highlighted in yellow. 46

Figure 4.20. The tree-based identification of sample 18 for the matK gene sequences. Sample 18 is highlighted in yellow. 47

Figure 4.21. The tree-based identification of sample 19. Sample 19 is highlighted in yellow. 49

Figure 4.22. The tree-based identification of sample 29 for the rbcL gene region. Sample 29 is highlighted in yellow. 51

Figure 4.23. The tree-based identification of sample 31 for the rbcL marker. Sample 31 is highlighted in yellow. 52

Figure 4.24. The tree-based identification of sample 32 for the rbcL sequences. Sample 32 is highlighted in yellow. 56

Figure 4.25. The tree-based identification of sample 46 for the rbcL sequences. Sample 46 is highlighted in yellow. 57

Figure 4.26. The tree-based identification of sample 33 for the rbcL gene region. Sample 33 is highlighted in yellow. 58

Figure 4.27. The tree-based identification of sample 40 for the rbcL gene region. Sample 40 is highlighted in yellow. 59

Figure 4.28. The tree-based identification of sample 43 for the rbcL marker. Sample 43 is highlighted in yellow. 59

LIST OF TABLES

Table 3.1. Primers used to amplify plant DNA barcodes.....13

Table 4.1. BLAST results of the rbcL gene region 15

Table 4.2. Results of the matK gene region 16

LIST OF SYMBOLS/ABBREVIATIONS

A	Adenine
BOLD	Barcode of Life Data System
BLAST	The Basic Local Alignment Search Tool for nucleotide
C	Cytosine
°C	Celcius degree
CBOL	Consortium for the Barcode of Life
COI	Cytochrome oxidase I
DNA	Deoxyribonucleic Acid
G	Guanine
GESC	Genovese Environmental Study Centre
ha.	Hectare
IUCN	International Union for Conservation of Nature
ITS	Internal transcribed spacer region of nuclear ribosomal DNA
ITS2	Internal transcribed spacer region 2 of nuclear ribosomal DNA
mtDNA	Mitochondrial Deoxyribonucleic Acid
MatK	Maturase K
µL	Microliter
NGBB	Nezahat Gökyiğit Botanical Garden
NGSS	The Next Generation Science Standards
NYCDOE	New York City Department of Education
PCR	Polymerase Chain Reaction
rbcL	Ribulose 1,5-biphosphate carboxylase
SMTP	School Malaise Trap Program
T	Thymine
UK	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
trnH-psbA	Intergenic spacer region

1. INTRODUCTION

Currently, 55% of the world's population resides in cities, and by 2050, that number is projected to rise to 68% (The World's Cities in 2018, n.d.). Considering that most people on Earth live in urban regions, there is a growing interest in studying urban ecosystems (Lepczyk et al., 2017). According to urban ecology, urbanization significantly affects ecological and evolutionary processes and people who live in urban settings. Besides human density, urban environments can support a variety of species native and non-native to the area with species of conservation significance (Uchida et al., 2020).

Identification of plant species is crucial to understanding habitats in urban ecosystems (Faeth et al., 2011)DNA Barcoding technology allows for rapid and accurate identification to compare DNA sequences of plants with confirmed DNA barcode databases. The DNA Barcoding method is used to distinguish between species with sequence variation, in small, defined gene regions. A high degree of species discrimination is provided by the optimal DNA barcode region, which is successfully amplified and sequenced across huge assemblages of taxa (Braukmann et al., 2017).

Only a strong library of accurate DNA sequence information about individuals may be used for DNA-based identification. Following that, the Barcode of Life Data System (BOLD) Project was created to build a global barcode reference library from well-known species (Ratnasingham & Hebert, 2007). Different barcode regions are effective in the discrimination of various groups of taxa.

The DNA barcode region of the mitochondrial cytochrome oxidase I (COI) gene's 5' region in resolving a variety of animal species differentiation. COI was impossible since most plant lineages had very low nucleotide substitution rates in mitochondrial genes (Braukmann et al., 2017). In 2009, at the 3rd World DNA Barcode Conference, it was reported that the matK and rbcL markers are the primary sequences of plant DNA barcodes, with ITS and trnH-psbA serving as complementary sequences 2021 (Yu et al., 2021). Subsequently, two plastid gene sections were taken into consideration as DNA barcodes for vascular plants based on their extensive use in phylogenetics and molecular systematics, and the large subunit of rubisco, ribulose 1,5-biphosphate carboxylase (rbcL) together with an intron maturase (matK) was accepted as standards (Braukmann et al., 2017).

The Web of Science database was used to overview the articles under the category of Web of Science core collection with the topic containing the keyword ‘’Plant DNA Barcoding’’. The

keyword has appeared in 3065 articles until August of 2023. Result of DNA Plant Barcoding for publication years between 2004 to August 2023.

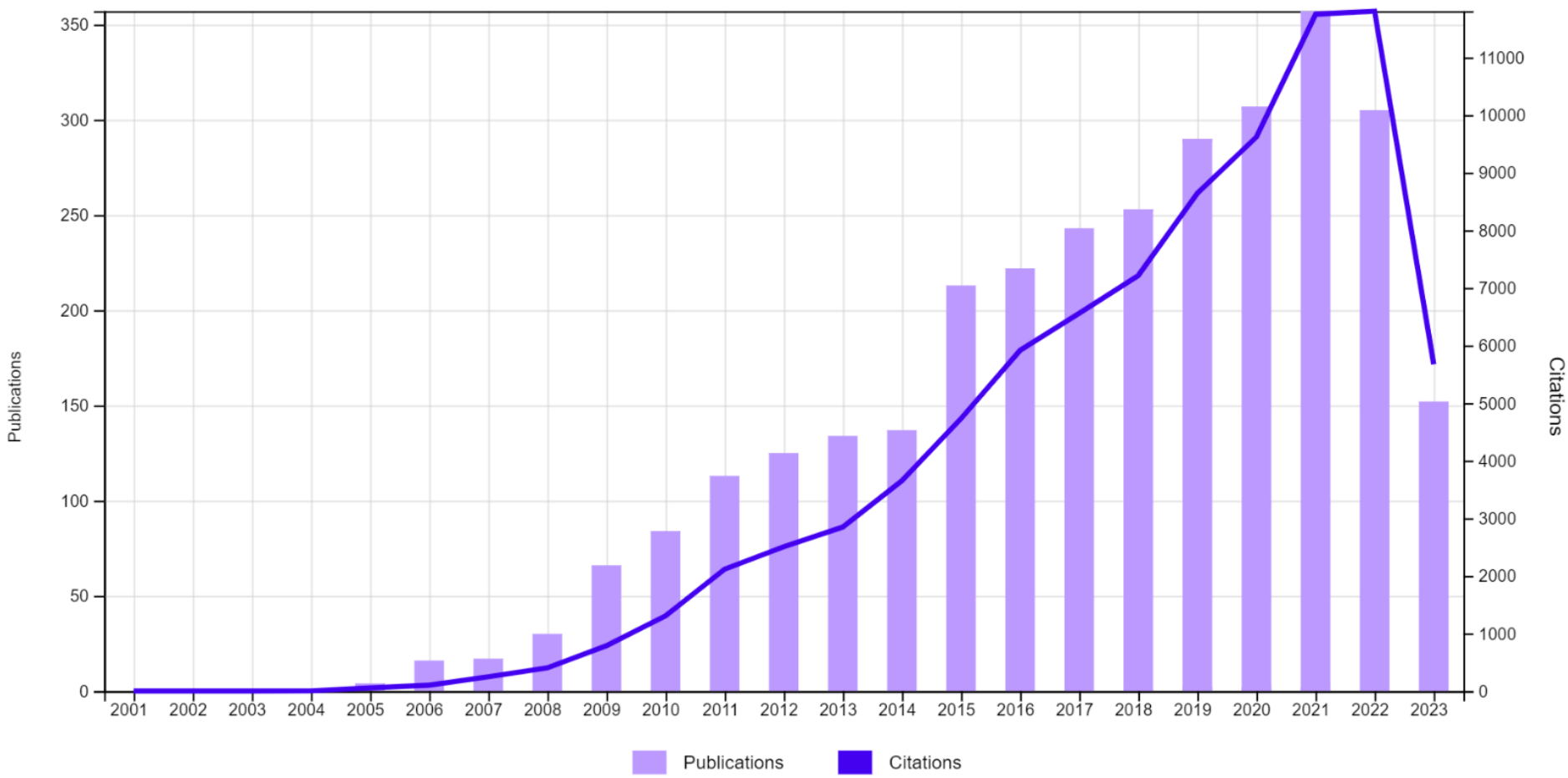


Figure 1.1. Result of DNA Plant Barcoding for publication years between 2004 to August 2023.

Several studies show successful discrimination rates in plants with rbcL, matK, and complementary ITS2 gene regions. In one study, Jones et. al (2021), sequenced barcodes and tried to identify 634 native UK plant species using the barcoding approach. For rbcL, 64% of species had matches at the species level, 98% had matches at the genus level, and 100% had matches at the family level. With 77% of species returned at the species level, 99% at the genus, and 100% at the family level for matK, there was a greater degree of species discrimination. Another example study by Braukmann et. Al (2017) used a DNA barcode reference library with sequence records for 96% of the 5108 vascular plant species known from Canada to test the accuracy of identification. Coverage ranged from 94% for rbcL, 60% for ITS2, and 39% for matK for species level of discrimination.

Furthermore, DNA Barcoding is not only an effective method for research but also an educational tool with a straightforward procedure and comparatively low cost. The method provides an opportunity to bring formal scientists, students, and teachers as citizen scientists all together from an educational perspective. For instance, DNA barcoding has been applied across a range of age groups to species identification in the USA for marine animals (Henter et al., 2016; Marizzi et al., 2018; Santschi et al., 2013), Canada for insects (Steinke et al., 2017) and the UK for plant species (Tosh et al., 2016) as an educational tool and citizen science project.

The keywords of ‘(DNA Barcoding) AND (Citizen Science)’ topic search reached 65 articles under the Web of Science core collection category. The search spanned from 2010 until August 2023. The ‘(Plant DNA Barcoding) AND (Citizen Science)’ keywords provided 12 articles in the same category from 2010 to August 2023.

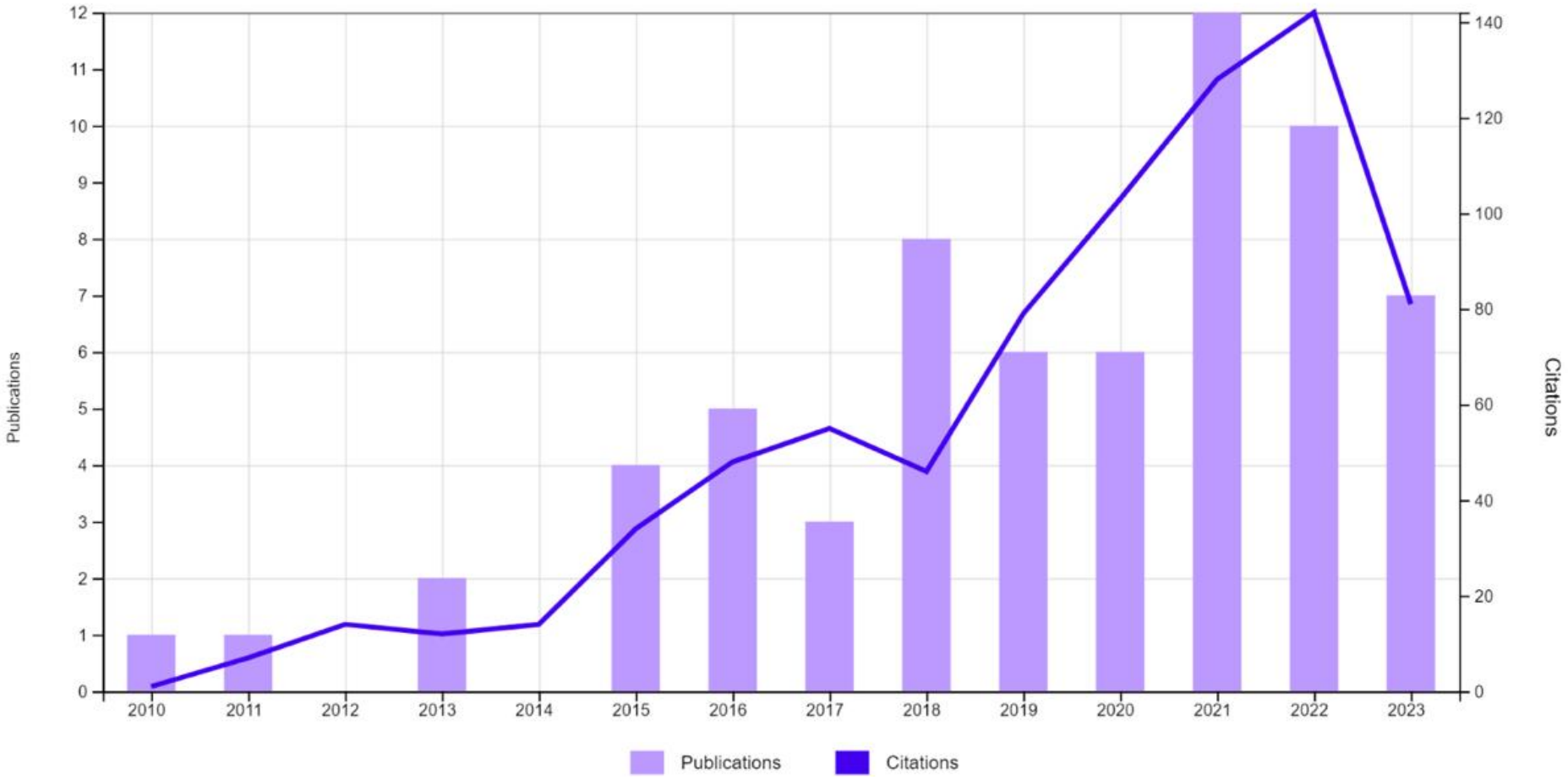


Figure 1.2. Result of DNA Barcoding and Citizen Science for publication years between 2010 to August 2023.

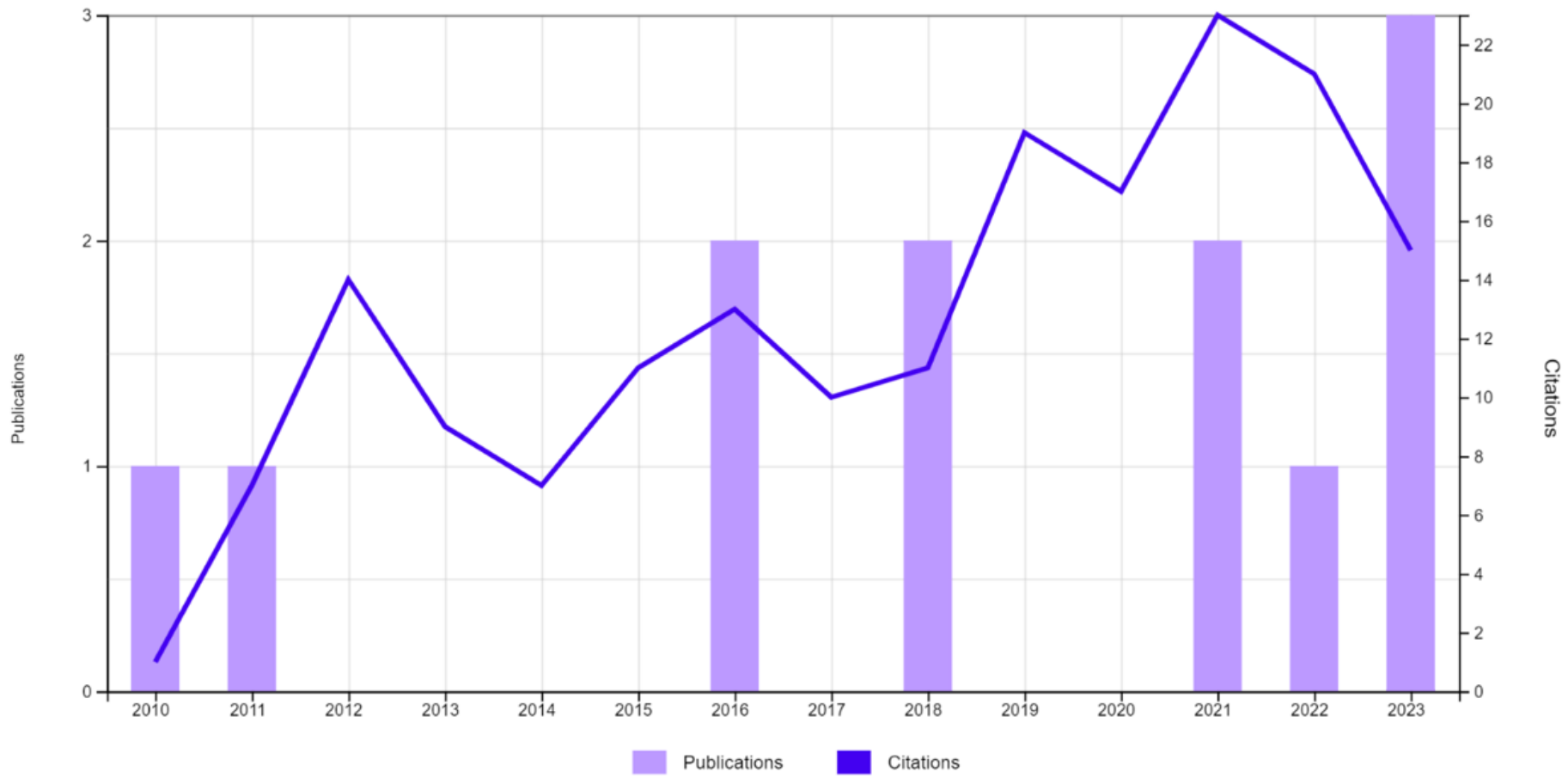


Figure 1.3. Result of Plant DNA Barcoding and Citizen Science for publication years between 2010 to August 2023.

The private high school Robert College offers a great opportunity for high-school students to learn about the environment since it is in a relatively protected woodland area in Istanbul, Turkey. The Robert College Campus includes about 24 ha (hectares) of woods on the Bosporus hills with approximately 60 species of trees recorded (Büyükkılıç 1980). Using the DNA Barcoding method, this study aims to identify the tree species on the campus of Robert College and investigate their evolutionary relationships, with the participation of students using a citizen science approach.

2. LITERATURE REVIEW

The definition of urbanization is the ecological forcing mechanism driven by the expansion of urban areas and the associated human activities (Mcdonnell & Pickett, 1990). Numerous studies have examined the effects of urbanization on the richness, variety, and species abundance of terrestrial animals to understand the dynamics of biodiversity changes in cities. Plant communities, either living or non-living, have a significant role in urban ecosystems that ultimately define a 'habitat' in the studies of urban biodiversity (Faeth et al., 2011). Urban biodiversity richness also depends on the size, quantity, and quality of urban green spaces (Ives et al., 2016; Lepczyk et al., 2017). In places with extreme urbanization, species richness tends to decrease for many groups, such as mammals, reptiles, amphibians, invertebrates, and plants. However, there are considerable differences among the groups in how suburban regions with moderate levels of urbanization affect them (Faeth et al., 2011). For instance, according to a review by McKinney (2008), among the 105 studies, 65% of plant studies show increased species richness with moderate urbanization. On the other hand, about 30% of invertebrate studies and a very small minority of non-avian vertebrate studies (about 12%) indicate increasing species richness with moderate urbanization.

The widespread positive relationship between human presence and species richness shows that people have either chosen to reside and generally thrive in biodiverse regions or have contributed to it by introducing new species and diversifying habitats. Istanbul is a great example of one of the most crowded metropolitan cities with diverse native and immigrant species. The scale-dependent relationship between the presence of humans and biodiversity underlines the significance of keeping green spaces in heavily populated areas. However, at smaller scales, people have chosen to live in regions with high biodiversity, have furthered beta diversity by the import of numerous species, and have increased habitat heterogeneity through complex land use patterns (Pautasso, 2007).

For rapid and accurate assessment of plant diversity, DNA Barcoding is a great tool. Paul Hebert proposed the DNA Barcoding approach in 2003 to classify the species based on a standardized small region of the genome (Hebert et al., 2003). The Barcode of Life Data System (BOLD) Project was subsequently designed to create a global barcode reference library from known species (Ratnasingham & Hebert, 2007) and at the time of publication, BOLD had identified more than 323,060 species using their unique barcodes. As of February 2023, the number of species in the BOLD system is 343.139.

DNA–based identification can only be achieved by the robust library of valid DNA sequence information of individuals. The verified DNA loci for the classification of the species are variable. The most accurate genetic markers for animal and plant species are suggested by the Consortium for the Barcode of Life (CBOL) (Tosh et al., 2016). The standardized barcoding region for animals is the mitochondrial gene cytochrome c oxidase subunit 1 (COI / cox1), which provides a comparatively high discrimination rate of about 95%. On the other hand, the rbcL + matK loci are the most definitive sequences for plant species, however, their classification accuracy is lower than the equivalent animal level of identification (Fazekas et al., 2009). Although the universal plant barcoding markers are still arguable for scientists, rbcL, matK, trnH-psbA, and ITS are the four most valuable markers considering recent technological developments (Yu et al., 2021).

Moreover, for several reasons, DNA barcoding is a convenient method for formal scientists, informal science education participants, as well as citizen scientists (Henter et al., 2016). The technique provides rapid and precise identification with the most miniature tissue specimens of the organism (Marizzi et al., 2018). Moreover, the concept and the steps of the method are simpler in application when compared to other genetics and molecular biology research methods. Therefore, students and citizen scientists can contribute to closing the biodiversity knowledge gap with DNA barcoding to document the diversity of nature at the same time (Henter et al., 2016).

DNA Barcoding is not only a powerful research method but also an important educational tool (Marizzi et al., 2018). DNA barcoding has been used to identify marine animals, insects, and plant species, in the USA (Santschi et al., 2013), Canada (Steinke et al., 2017), the UK (Tosh et al., 2016, and other countries with a variety of age groups as an educational tool. For instance, Barcoding Life's Matrix, a multidisciplinary program of research education for students aged between 5 to 18 was created by Coastal Marine Biolab and collaboration with The Next Generation Science Standards (NGSS) in 2011 (Henter et al., 2016). The program aims to create a global DNA Barcoding reference library to improve interdisciplinary scientific education for students and instructors in many states across the United States (Santschi et al., 2013). The project is supported by the Barcode of Life Data Systems Student Data Portal (BOLD-SDP). The activity provides an opportunity for students and teachers to study big data sets to apply data mining, analysis, and publication with a combination of qualitative and quantitative research methods. Over three years, more than 1600 barcode records of marine animal species were uploaded to the BOLD system. Those data are contributed to inventories conducted by the Alaska Fisheries Science Centre and the Channel Island National Park (Henter et

al., 2016) Barcoding Life Matrix is a solid example to picture the importance of collaborative effort between the student and formal scientists.

Another collaborative educational effort, the School Malaise Trap Program (SMTP), included 15,000 students from 350 primary and secondary schools across Canada between 2013 and 2017 (Steinke et al., 2017). Hundreds of insect species were collected from school backyards with Malaise traps (Malaise, 1937)and identified with the DNA Barcoding method. Until 2017, more than 80,000 insect specimens had been sequenced, and 7,999 insect species were identified. Foremost, the first DNA Barcode of 1.288 Canadian species was added to the BOLD system. Following this experiment, a scientist from the University of Guelph performed an online poll to better understand teacher and student involvement and satisfaction. Most students, 95%, strongly suggested the program to their peers. Moreover, 91% of teachers stated that the students were highly committed to the program and indicated positive feedback to this research/educational approach (Steinke et al., 2017).

Another research project entitled "DNA Barcoding Brooklyn (New York)" was the first design to estimate the biodiversity of Marine Park in Brooklyn with the involvement of citizen scientists. High school teachers were trained by professionals from the NYC Department of Education (NYCDOE) Genovese Environmental Study Centre (GESC) and collaborated with taxonomists from Rockefeller University (Marizzi et al., 2018). High school students and teachers used the BioBlitz method as an accelerated assessment of living organisms in a specific area within a limited time, to collect specimens from Marine Park (Marizzi et al., 2018; Ruch et al., 2010). As a result of the project, one fungus, 70 animals, and 116 plant species were identified, and 12 novel DNA barcodes were added to the genetic data bank GenBank. The COI barcoding marker could discriminate 70% of animal species in this project. However, only 42% of plants were identified at the species level by the COI genetic barcode (Marizzi et al., 2018).

The 'Tree Project' from the UK is another great example of conducting plant identification with a collaborative effort to use the DNA Barcoding method as a tool. Cothill Educational Trust and Natural History Museum established the research education-based project to engage secondary school students in natural sciences via tree species inventory of their environments. Students mainly undertook fieldwork, which required plant sample collection. The project focused on the identically well-known tree species that plant scientists and ecologists could identify. A floristic study that includes native and non-native woodlands and shrub species from Old Malthouse Residential Field Centre, Dorsey, was used to choose tree species in that area. The most available DNA Barcode markers for plants (namely rbcL and matK loci) were chosen. The scientists emphasized the accuracy of sequences and the difficulties of studying with students. The educational result of their study, using

a mix of formal and informal educational approaches such as a combination of designed games to explain scientific terminologies and hands-on training with lectures, was effective. Students chose their tree species and collected leaves from them (Tosh et al., 2016). The project's results showed the identification limits of two markers rbcL+ matK loci at the species scale to compare the use of CO1 gene in animals. However, these markers provided 100% accuracy on the genus scale (Tosh et al., 2016).

Robert College, a private high school, has been in a relatively protected forest area in Istanbul for more than 100 years, providing excellent environmental education opportunities for high school students. The campus area was started to be used by the school in 1914 and is a protected area with high urban biodiversity. Parts of the forest were planted in those years and the plant diversity of the campus varied through the years. The campus underwent a systematic identification study for its trees in 1980, identifying 60 species; however, there is still a wide knowledge gap in biodiversity assessment and evaluation of the area (Büyükkılıç, 1980). This thesis aims to identify the tree species of Robert College campuses with students using the recent educational and scientific tool DNA Barcoding approach. The school campus offers good opportunities for research education and research into the Bosporus's plant richness as a protected urban area. This thesis will also help to introduce the students to recent scientific tools to investigate their environment plant diversity and the evolutionary relationship of plant species, using a citizen science perspective.

3. MATERIALS AND METHODS

3.1. Sampling

We collected 60 tree species for morphological taxonomical identification and molecular studies from the Robert College Campus in Istanbul. Robert College Campus comprises approximately 24 ha. (hectare) of woodland on the hillsides of the Bosporus. Campus settlement area altitude is between 90 – 140 m (Büyükkılıç, 1980).

3.2. Morphological Identification

The PlantNet application, the image recognition system, was used for taxonomical identification. PlantNet is a platform for citizen science research that uses artificial intelligence (AI) to make plant species identification and inventory easier. With millions of contributors from more than 200 nations, it is among the biggest biodiversity observatories in the world (Camille Garcin, n.d.).

3.3 Molecular Studies

The DNeasy Plant Pro Kit (Qiagen, Germany) was used for total DNA purification from the leaf samples. Using a bead-beating technology, the DNeasy Plant Pro Kit recovers high-quality DNA from grapevine leaves, pine needles, and numerous seed types. PCR inhibitors, such as secondary metabolites from plant extracts, are removed during the isolation process isolation using the second generation of QIAGEN's patented Inhibitor Removal Technology® (IRT). High yields of inhibitor-free DNA are produced using IRT in conjunction with improved bead beating and lysis chemistry, making it directly usable in downstream applications like PCR. Qubit 3.0 fluorometer (Fisher Scientific, n.d.) was used to measure the quantification of the concentration of samples' DNA.

3.4 PCR Amplification

The polymerase chain reaction (PCR) will be carried out following the steps provided by (Tosh et al., 2016) with some adjustments after optimizations to amplify the rbcL + matK region of the tree species previously indicated. The primers rbcLa f and rbcLa rev will be used to amplify rbcL. (Kress & Erickson, 2007). The amplification of matK will involve the use of two sets of primers; Gym R1B

and Gym F1B were utilized for gymnosperms, while 3F KIM and 1R KIM were employed for angiosperms (Tosh et al., 2016).

The reaction mixture included 12.5 µL of Taq premix,1.25 µL forward primer cocktail, 1.25 µL reverse primer cocktail, 7 µL ddH2O, and 3 µL DNA in a total volume of 25 µL. PCRs were run under the 35 following thermal cycle conditions: 5 min at 94°C; 30 sec. at 94 C; 1 min at 50°C; 1 min at 72°C; 10 min at 72 °C (Tosh et al., 2016)

Table 3.1. Primers used to amplify plant DNA barcodes (Tosh et al., 2016).

Region	**Primer**	**Primer sequence (5'-3')**
rbcL	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC
rbcL	rbcLa_rev	GTAAAATCAAGTCCACCRCG
matK(angiosperms)	3F_KIM	CGTACAGTACTTTTGTGTTTACGAG
matK(angiosperms)	1R_KIM	ACCCAGTCCATCTGGAAATCTTGGTTC
matK (gymnosperms)	Gym_R1B	TCATCCRGAAATTTTGGTKCG
matK (gymnosperms)	Gym_F1B	ATMGTACTTTTATGTTTACARGC

3.5. Gel Electrophoresis

To examine the PCR results, 1% agarose gels in 1X TAE (40 mM Tris-acetate and 1 mM EDTA) buffer were produced red-safe nucleic acid gel stain. Each reaction was loaded onto the gel with 5 µL of the PCR product and 1 µL of the red DNA loading dye. After 45 minutes of running samples at 90 V, the PCR Product bands on the gel were imaged and visualized under ultraviolet light

3.6. Sequencing and Species Identification

For the sequencing step, DNA samples were sequenced by a company. The obtained nucleotide sequences, which are obtained, was cleaned, and aligned using the software Sequencer program. The results were loaded in the BLAST (The Basic Local Alignment Search Tool for nucleotide) system as a fasta format to identify from a valid gene library.

4. RESULTS AND DISCUSSION

Sample codes were given in the process of sequencing each sample based on the field identification (ID) names. PlantNet, an application tool for plant identification in the field, was used for morphological identification for Field ID. Barcode ID was described by rbcL and matK gene sequences. Twenty-two clean sequences were obtained for the rbcL gene in fifty-four samples, and five clean sequences were gathered for the matK sequences. In fifty-four samples, twenty-two clean sequences were obtained for the rbcL gene, and five clean sequences gathered for the matK sequences. These corresponded to seven families and thirteen genera in the table under the Barcode ID. In the Basic Local Alignment Search Tool for nucleotide (BLAST) method, the first five-hit species with the highest percent identity were considered to make the genetic identification. The number of species beneath the top first five hits in BLAST is represented by the number of species on the table. For the rbcL locus, eight of our samples had a 100% percent identity match rate, and the remaining samples had percent identity matches over 96.22%. Throughout the twenty-two clean sequences for the rbcL marker, query length was between 139 - 462. For the matK locus, two of our samples had over 96.81% percent identity match, and four of them had a percent identity cover of over 87.10%. The query length was between 94 – 495 throughout the six samples for the matK marker.

Table 4.1. BLAST results of the rbcL gene region

Sample code for the rbcL	Field ID	Barcode ID	Query ID	Query Length	Query Cover	E- value	Percent Identity	Number of hit species
42	*Tilia sp.*	*Cercis* genus	lcl|Query_7008669	451	100%	0.0	>99.78%	3
3	*Fraxinus angustifolia*	*Fraxinus* genus	lcl|Query_5769233	394	100%	0.0	99.49%	5
4	*Pistacia atlantica*	*Pistacia* genus	lcl|Query_6210215	319	100%	9e-164	100%	4
5	*Celtis occidentalis*	Celtis genus	lcl|Query_8032555	283	100%	83-144	100%	2
7	*Ficus carica*	*Ficus* genus	lcl|Query_5694839	371	100%	0.0	100%	2
8	*Morus alba*	*Morus genus*	lcl|Query_3837563	343	100%	8e-160	97.08%	5
9	*Paulownia tomentosa*	*Paulownia* genus	lcl|Query_ 921557	246	100%	3e-123	100%	5
10	*Cinnamomum camphora*	*Cinnamomum* genus	lcl|Query_6313065	409	100%	0.0	99.76%	3
11	*Juglans regia*	Juglandeceae family	lcl|Query_313673	370	100%	0.0	100%	4
12	*Quercus robur*	*Fagaceae* family	lcl|Query_3412037	208	100%	5e-100	99.52%	4
13	*Quercus coccifera*	*Fagaceae* family	lcl|Query_450545	359	100%	3e-179	98.89%	3
15	*Aucuba japonica*	*Aucuba* genus	lcl|Query_6359639	139	100%	4e-64	100%	3
16	*Carpinus orientalis*	*Carpinus* genus	lcl|Query_2524277	278	100%	8e-129	97.48%	3
18	*Rubus pruinosus*	*Rubus* genus	lcl|Query_3491267	318	100%	3e-163	100%	5
19	*Nerium oleander*	*Apocynaceae* family	Icl|Query_5953909	373	100%	0.0	99.20%	5

Sample code for the rbcL	Field ID	Barcode ID	Query ID	Query Length	Query Cover	E- value	Percent Identity	Number of hit species
29	*Juniperus virginiana*	*Cupresseae* family	lcl|Query_4298241	393	100%	0.0	99.49%	3
31	*Olea europaea*	*Oleaceae* family	lcl|Query_5758285	340	100%	4e-163	97.94%	2
32	*Laurus nobilis*	*Lauraceae* family	lcl|Query_3318789	441	100%	0.0	96.83%	5
46	*Tilia sp.*	*Malvaceae* family	lcl|Query_7627403	343	100%	1e-153	96.22%	5
40	*Quercus coccifera*	*Quercus* genus	lcl|Query_6744123	443	100%	0.0	100%	5
43	*Viburnum tinus*	*Viburnum* genus	lcl|Query_7639431	462	100%	0.0	99.78%	3
33	Unknown	*Lauraceae* family	lcl|Query_1764913	388	100%	0.0	99.74%	5

Table 4.2. Results of the matK gene region

Sample code for the matK	Field ID	Barcode ID	Query ID	Query Length	Query Cover	E- value	Percent Identity	Number of hit species
5	*Celtis occidentalis*	Celtis genus	lcl|Query_7407177	495	99%	0.0	91.06%	3
7	*Ficus carica*	*Ficus* genus	lcl|Query_3950661	241	100%	4e-92	92.98%	2
11	*Juglans regia*	*Juglandaceae* family	lcl|Query_1065257	244	100%	2e-95	92.21%	1
13	*Platanus hispanica*	*Fagaceae* family	lcl|Query_6755629	216	100%	3e-57	87.10%	5
15	*Citrus paradisi*	*Aucuba* genus	lcl|Query_5847791	104	100%	2e-41	98.08%	1
18	*Rubus pruinosus*	Rubus genus	lcl|Query_8179109	94	100%	2e-35	96.81%	5

4.1. *Cercis*

One sample with code 42 was identified to belong to the genus *Cercis* by The Basic Local Alignment Search Tool for nucleotide (BLAST) method. The query cover for the samples was 100%. Sample 42 had a percent identity match of >99.78% with three different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.1. below shows the tree generated for the tree-based identification of the sequence. Our unknown sample did not cluster with a particular species but with different species within the genus *Cercis*.

Using the PlantNet identification application in the field, our sample was morphologically identified as a species from the *Tilia* genus. Although the morphological identification and barcoding identification pointed to a different genus, the results of BLAST, the *Cercis* genus revealed a percent identity match of >99.78% for the first sixteen hits. The first hit *Cercis siliquastrum* subsp. *siliquastrum* had a percentage identity match of 100%. It is a native species of Turkey and is distributed in Istanbul (Allmer et al., n.d.). Therefore, based on the BLAST results and distribution data, our sample could be *Cercis siliquastrum* subsp. *siliquastrum*.

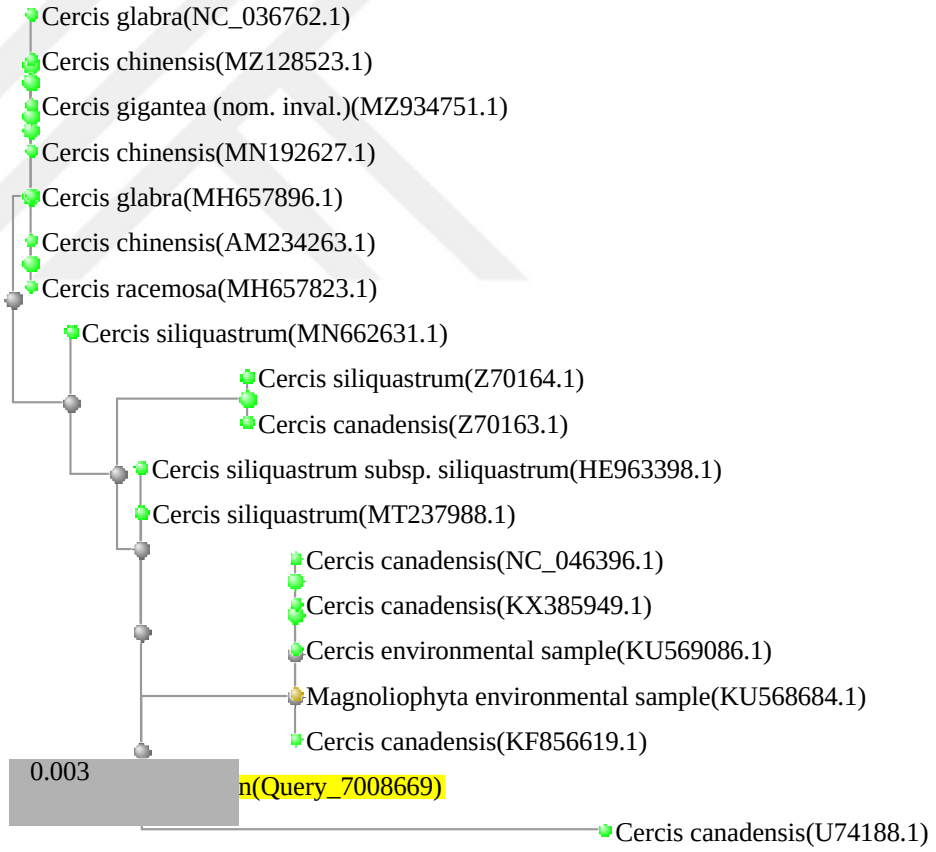


Figure 4.1. The tree-based identification of sample 42. Sample 42 is highlighted in yellow.

4.2. *Fraxinus*

One sample with code 3 was identified to belong to the genus *Fraxinus* by BLAST method. The query cover for the samples were 100%. Sample 42 had a percent identity match of 99.49% with five different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.2. below shows the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification software in the field, our sample was morphologically identified as a *Fraxinus angustifolia* species. Accordingly, the identification of the morphological species fits under the *Fraxinus* genus. *Fraxinus angustifolia* is found throughout central-southern Europe, northwest Africa, and the Caucasus region (Caudullo & Durrant, n.d.). Similar to the first five results in the BLAST, the species *Fraxinus angustifolia* exhibited a percent identity match of 99.49%. Due to high degree of morphological variability, three subspecies of *Fraxinus angustifolia* have been identified based on molecular and morphological data: *Fraxinus angustifolia* subsp. *angustifolia* which is found in southwest Europe and northwest Africa; the *Fraxinus angustifolia* subsp. *oxycarpa* (M. Bieb. Ex Willd.) Franco & Rocha Afonso] in central Europe, the Balkans, and the Black Sea region; and the *Fraxinus angustifolia* subsp. *syriaca* (Boiss.) Yalt.) in southeast Anatolia and the Middle East up to Iran (Caudullo & Houston Durrant, 2016). All three subspecies of *Fraxinus angustifolia* are distributed in Turkey, two of them *Fraxinus angustifolia* subsp. *syriaca* and *Fraxinus angustifoli* subsp. *oxycarpa* were identified in Istanbul (Allmer et al., n.d.). Moreover, *Fraxinus angustifolia* species is more commonly employed in rapidly expanding plantations in marshy lowlands in northwest Turkey (Çiçek & Yılmaz, 2002). Therefore, in BLAST, our percent identity got 99.49% with other species under the *Fraxinus* genus, the species could be *Fraxinus angustifolia* subsp. *oxycarpa* or *Fraxinus angustifolia* subsp. *syriaca*.

4.3. *Pistacia*

One sample with code 4 was identified to belong to the genus *Pistacia* by BLAST method. The query cover for the samples were 100%. Sample 4 had a percent identity match of 100% with four different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.3. below shows the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Pistacia atlantica* species. Accordingly, the identification of the morphological species fits under the Pistacia genus. Similar to the first five results in the BLAST, the species *Pistacia atlantica* exhibited a percent identity match of 100%. *Pistacia atlantica* is distributed as a native species in Turkey (Kew science, 2024) and was identified in Istanbul (Allmer et al., n.d.). Therefore, in BLAST, our percent identity was 100% with other species under the Pistacia genus, the species could be *Pistacia atlantica.*



Figure 4.2. The tree-based identification of sample 3. Sample 3 is highlighted in yellow.

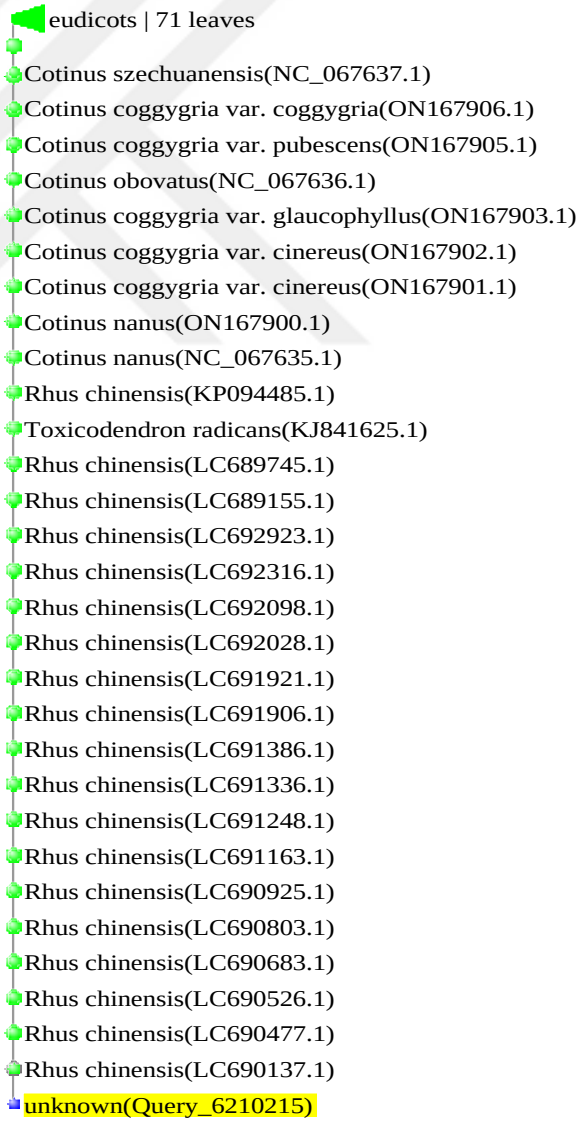


Figure 4.3. The tree-based identification of sample 4. Sample 4 is highlighted in yellow.

4.4. *Celtis*

The sample with code 5 was identified to belong to the genus *Celtis* by BLAST method with both rbcL and matK regions. The matK region query cover was 99%. It had a 91.06 % percent identity match with three species (*Celtis australis, Celtis sinensis* and *Celtis timorensis)* in the first five results. The rbcL sequence query cover was 100%. Sample 5 had a percent identity match of 100% with three different species (*Celtis australis*, *Celtis tournefortii*, and *Celtis glabrata)* in the first five hits. Therefore, a species-level identification was not possible for both regions, and the identification was left at the genus level. Figure 4.4. and figure 4.5 below show the tree generated for the tree-based identification of the sequences.

Our sample was identified morphologically as a species *Celtis occidentalis* using the PlantNet identification tool in the field. Accordingly, the identification of the morphological species fits under the *Celtis* genus. The species *Celtis occidentalis* exhibited a percent identity match of 100% for the rbcL and 90.65% for the matK region. However, according to literature, *Celtis occidentalis* is not found in Turkey (Allmer et al., n.d.) and s natively distributed in North America ((Kew science, 2024). There are four species under the *Celtis* genus recorded in Turkey, and three of them matched with our sample in BLAST. *Celtis australis*, *Celtis tournefortii*, and *Celtis glabrata* had a percent identity match of 100% for rbcL. Within these three species, only *Celtis australis* matched for matK region with the highest result 91.06% percent identity in BLAST result. Moreover, only *Celtis australis* has been previously documented in Istanbul among these three species (Allmer et al., n.d.) Therefore, considering distribution records in Istanbul and BLAST data, and because *Celtis australis* had a percent identity got a 100% similarity under the *Celtis* genus, the species could be *Celtis australis.*

4.5. *Ficus*

The sample with code 7 was identified with rbcL and matK gene regions to belong to the genus *Ficus* by the BLAST method. The query cover for the rbcL marker of the samples was 100%. Sample 7 had a percent identity match of 100% with five different species in the first five hits of rbcL analysis. The matK gene region had 100% query cover in BLAST and 92.98% percent identity for the same four species. Additionally, *Ficus carica* had the highest number of hits (24) among the other species. Figure 4.6. and Figure 4.7. below are the trees generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as *Ficus carica*. Accordingly, the identification of the morphological species fits under the *Ficus* genus. Complementary with the first five results in the BLAST, the species *Ficus carica* exhibited a percent identity match of 100% with rbcL marker and 92.98% with matK marker. The rbcL region is thought to be an accurate barcode locus for a variety of plant species since it could resolve diverse species that belonged to the *Ficus* genus (Wei et al., 2024). *Ficus carica* is natively distributed in Turkey ((Kew science, 2024).) and was identified in Istanbul (Allmer et al., n.d.). Therefore, in BLAST, our percent identity results in two gene regions with other species under the *Ficus* genus, the species could be *Ficus carica*.

4.6. *Morus*

One sample with code 8 was identified to belong to the genus *Morus* by BLAST method. The query cover for the samples was 100%. Sample 8 had a percent identity match of 97.08% with five different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.8. below shows the tree generated for the tree-based identification of the sequence.

Our sample was identified morphologically as a *Morus alba* species using the PlantNet identification tool in the field. Accordingly, the identification of the morphological species fits under the *Morus* genus. Complementary to the first five results in the BLAST, the species *Morus alba* exhibited a percent identity match of 100%. with other four species. Hussain et al. searched the sequence cover of the rbcL and matK for *Morus* species including *Morus alba.* The rbcL alignment showed that at 91.4% sequence coverage, all 12 sequences had at least 91.6% identity. While the matK findings shared at least 69.3% identity and 67% sequence coverage, they differed slightly from one another. The results demonstrate the effective species discrimination capabilities of the rbcL and matK markers with the shortest sequence of 476 bp (Hussain et al., 2023) In our study, Sample 8 had a 343 bp query length.

There are over 1,100 species in the *Moraceae* family, which is found in temperate, tropical, and subtropical climates worldwide in both wild and cultivated forms (Clement & Weiblen, 2009). The family comprises 37 genera. The species of the genus *Morus* are found in parts of Africa, the Northwest, South and South America, and Southern Europe. Another name for them is mulberries. Morus species may adapt quite well to a wide range of ecological circumstances. (Hussain et al., 2023)*Morus alba* was distributed in Turkey as an introduced species ((Kew science, 2024).) and was

identified in Istanbul (ref. TÜBİVES). Therefore, in BLAST, our percent identity got 100% with other species under the *Morus* genus, the species could be *Morus alba*.

4.7. *Paulownia*

One sample with code 9 was identified to belong to the genus *Paulownia* by BLAST method. The query cover for the samples was 100%. Sample 9 had a percent identity match of 100% with five different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.9. below is the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Paulownia tomentosa*_species. Accordingly, the identification of the morphological species fits under the *Paulownia* genus. The species *Paulownia tomentosa* showed a percent identity match of 100%, which is comparable to the first five hits in the BLAST. *Paulownia tomentosa* is widely distributed in China, Korea, and Japan ((Kew science, 2024).). The Paulownia tree can grow at very high rates when conditions are right. Fast-growing *Paulownia tomentosa* was brought to Turkey from China, where it is a commonly distributed species (Hakan Akyildiz & Kol, 2010). Therefore, in BLAST, our percent identity was 100% with other species under the *Paulownia* genus, the species could be *Paulownia tomentosa*.

4.8. *Cinnamomum*

One sample with code 10 was identified to belong to the genus *Cinnamomum* by BLAST method. The query cover for the samples were 100%. Sample 10 had a percent identity match of 99.76% with three different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.10. below shows the tree generated for the tree-based identification of the sequence.

Our sample was identified morphologically as a *Cinnamomum camphora* species using the PlantNet identification tool in the field. Accordingly, the identification of the morphological species fits under the *Cinnamomum* genus. Like the first five results in the BLAST, the species *Cinnamomum camphora* exhibited a percent identity match of 99.76%. *Cinnamomum camphora* resident for China, Korea, Japan, and Vietnam ((Kew science, 2024). Worldwide, the genus *Cinnamomum* has approximately 250 species. In the sudy of Bhagya Chandrasekara et al., eight species of *Cinnamomum* in Sri Lanka was evaluated using matK, rbcL, trnH-psbA, and universal barcoding areas. Overall, the

DNA barcoding regions examined did not allow for a comprehensive resolution of the phylogeny of the *Cinnamomum* species identified in Sri Lanka (Bhagya Chandrasekara et al., 2021). *Cinnamomum camphora* was brought to Europe in the 17th century, and gardens there are now often grown with it. It is a species with a very long lifespan that can adapt to a variety of climates and soil types (Bottoni et al., 2021). Therefore, in BLAST, our percent identity was 99.76% with other species under the *Cinnamomum* genus, the species could be *Cinnamomum camphora*.

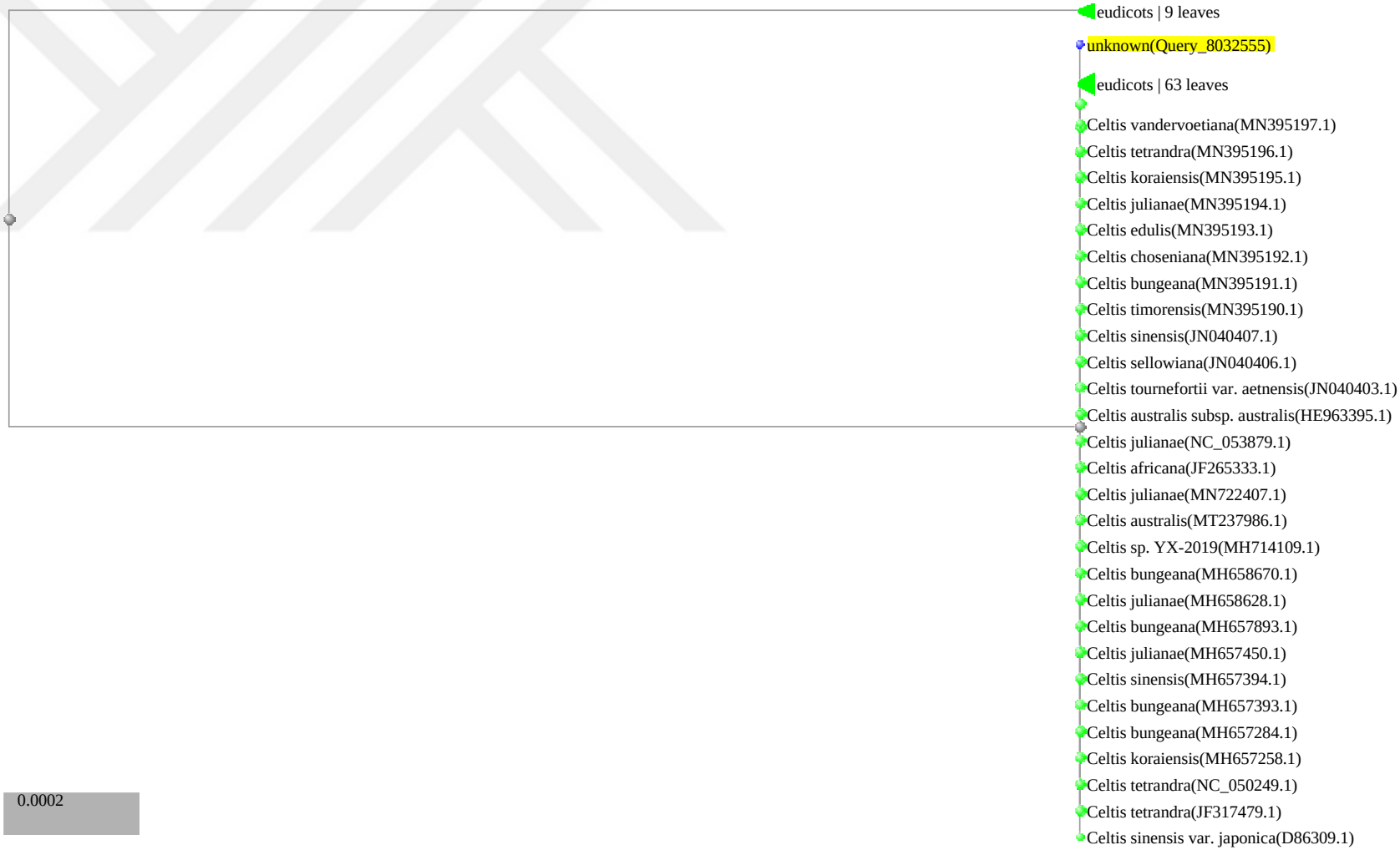


Figure 4.4. The tree-based identification of sample 5 for the rbcL marker. Sample 5 is highlighted in yellow.



Figure 4.5. The tree-based identification of sample 5 for the matK marker. Sample 5 is highlighted in yellow.

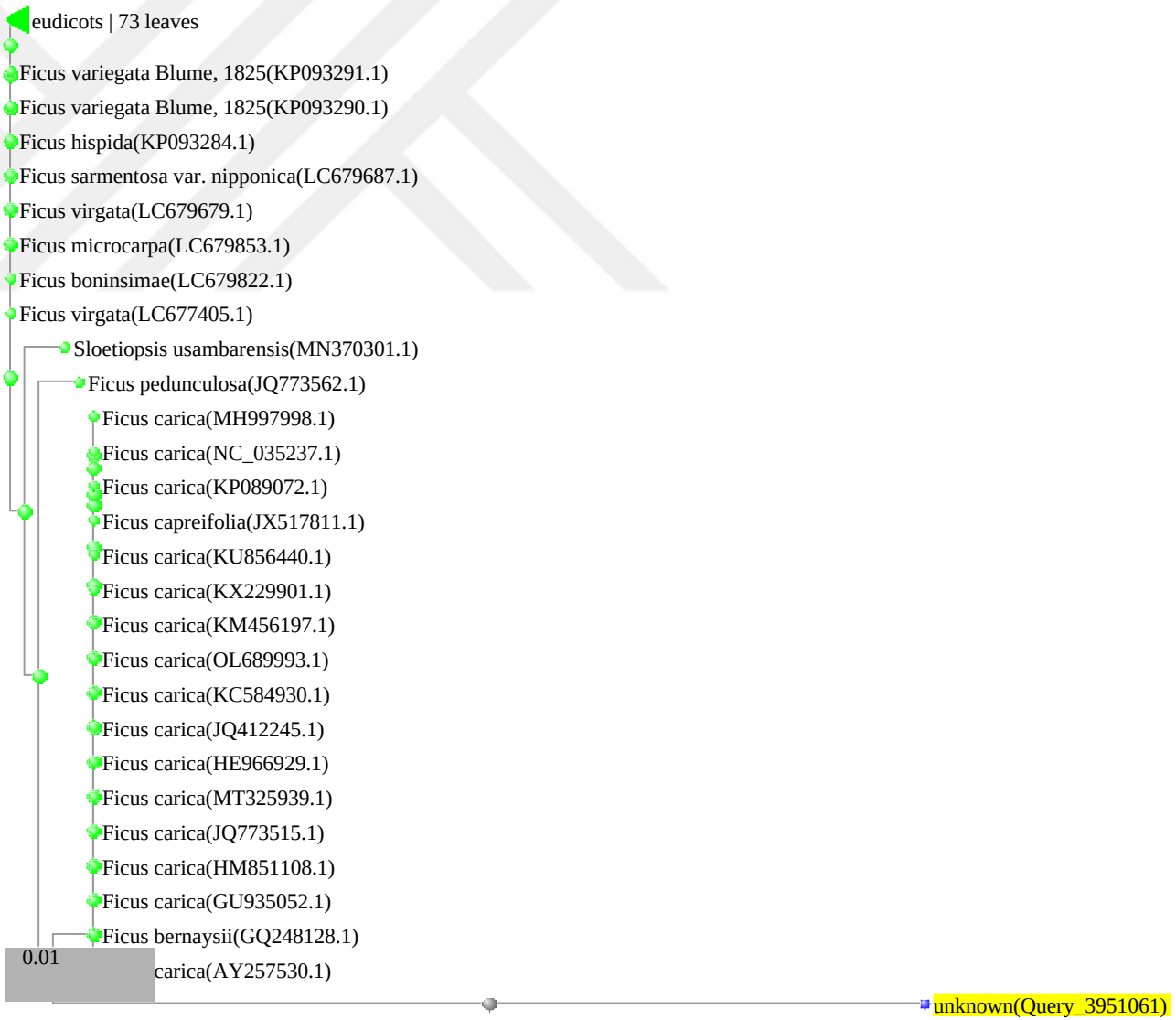


Figure 4.6. The tree-based identification of sample 7 for the matK marker. Sample 7 is highlighted in yellow.

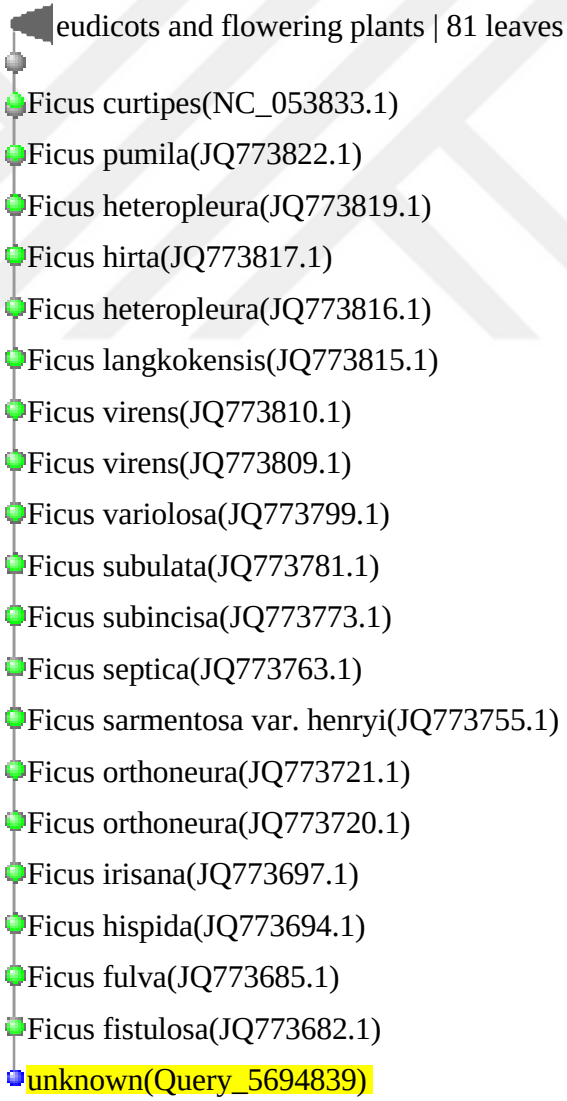


Figure 4.7. The tree-based identification of sample 7 for the rbcL marker. Sample 7 is highlighted in yellow.

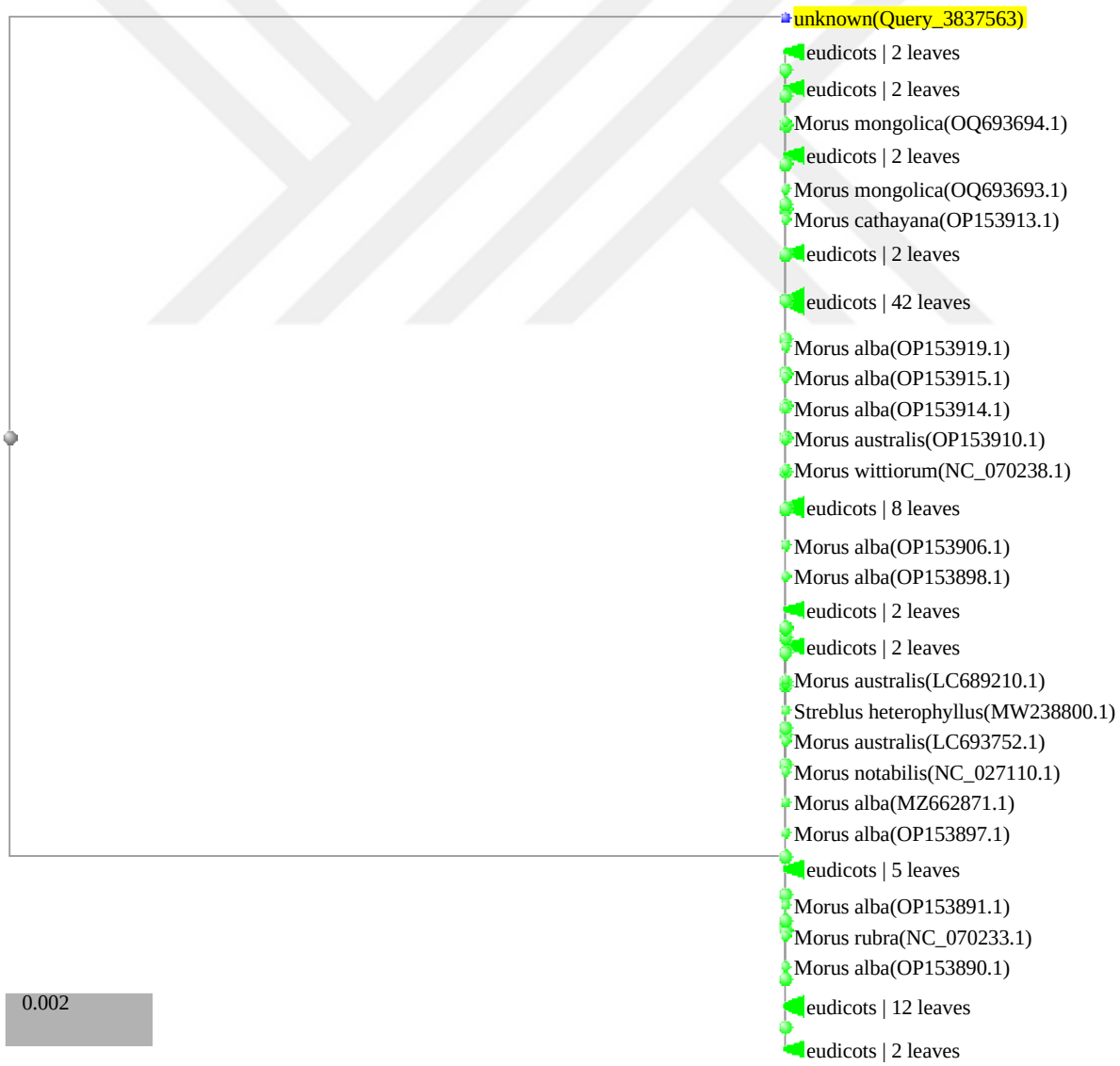


Figure 4.8. The tree-based identification of sample 8 for the rbcL gene region. Sample 8 is highlighted in yellow.



Figure 4.9. The tree-based identification of sample 9 for the rbcL gene sequences. Sample 9 is highlighted in yellow.

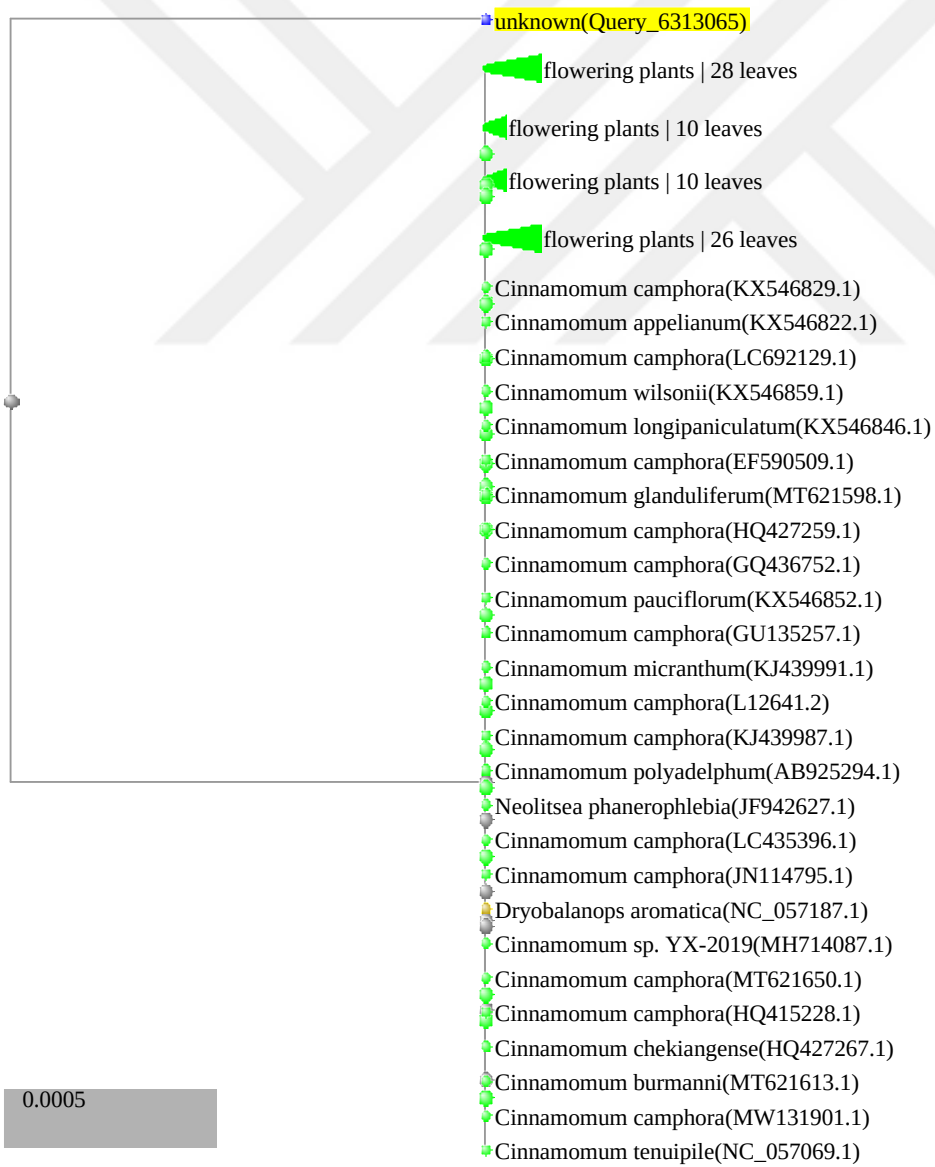


Figure 4.10. The tree-based identification of sample 10 for the rbcL gene region. Sample 10 is highlighted i

4.9. *Juglandaceae*

One sample (code 11) was identified to belong to the family *Juglandaceae* by the BLAST method. The query cover for the samples was 100% for both rbcL and matK gene regions. Xiang et al (2011) conducted research about *Juglandaceae* family for nine genera and fifty-four species with barcoding loci rbcL, matK, trnH-psbA, and ITS. Whereas rbcL has the least variable information among the four analyzed DNA regions, ITS has the most. The matK can effectively identify seven out of nine genera at the generic level. For the rbcL marker, sample 11 had a percent identity match of 100% with four different species in the first five hits for the rbcL marker. Accordingly, a species or genus-level identification was unobtainable and left at the family level. Figure 4.11. below shows the tree generated for the tree-based identification of the rbcL sequence. For the matK gene region, sample 11 had a percent identity match of 92.21% with one species in the first five hits. Figure 4.12. below is the matK marker tree generated for the tree-based identification.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Juglans regia*. Accordingly, the identification of the morphological species fits under the *Juglandacea* family. Similar to the first five results in the BLAST, the species *Juglans regia* exhibited a percent identity match of 100% for rbcL gene region. Moreover, *Juglans regia* was the only species in the first five hits with a percent identity of 92.21% for matK marker. .*Juglans regia* is grown locally in the Balkans, Iran, Turkey, Central Asia, the Himalayas, and China in large-scale agriculture (Gunn et al., 2010). It is natively distributed in Turkey (Kew science, 2024) and was identified in Istanbul (Allmer et al., n.d.). Therefore, considering the BLAST percent identity of 100% with other species under the *Juglandaceae* family, the species could be *Juglans regia.*

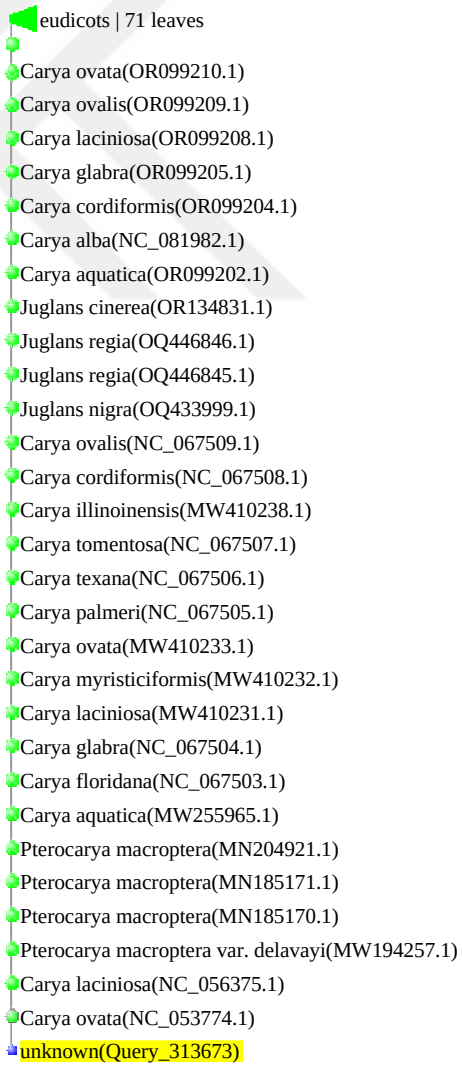


Figure 4.11. The tree-based identification of sample 11 for the rbcL sequence. Sample 11 is highlighted in yellow.



Figure 4.12. The tree-based identification of sample 11 for the matK sequence. Sample 11 is highlighted in yellow.

4.10. *Fagaceae*

Samples 12 and 13 were identified to belong to the family *Fagaceae* by BLAST method. Sample 12 had a percent identity match of 99.52 % and query cover of 100% with four different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.13. below shows the tree generated for the tree-based identification of the sequence. Sample 12 was morphologically identified as a *Quercus robur* species by using the PlantNet identification application in the field. Accordingly, the identification of the morphological species fits under the *Fagaceae* family. However, *Quercus robur* did not match as a species in BLAST. Four different genera *Quercus, Castanea, Castanopsis,* and *Lithocarpus* under the *Fagaceae* family exhibited a percent identity match of 99.04%. There were also forty different species from the *Quercus* genus that exhibited a percent identity match of 99.04% with all the species. Furthermore, *Quercus pubescens* and *Quercus macranthera* are distributed in Turkey; however, only *Quercus pubescens* has been previously recorded in Istanbul (Allmer et al., n.d.). Therefore, considering both genetic and morphological identification results with distribution data, *Quercus pubescens* had a percent identity got 99.04% with other species under the *Fagaceae* family, the species could be *Quercus pubescens.*

Sample 13 was also identified with two gene regions, rbcL and matK. The rbcL gene region for sample 13 had 100% query cover and a percent identity match of 98.89% with three different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Two species from the *Quercus* genus, namely *Quercus pubescens* and *Quercus congesta* had the highest percent identity match of 98.89% and only *Quercus pubescens* is recorded in Istanbul (Allmer et al., n.d.). The distribution of *Quercus congesta* is Italy(Carrero et al., 2020) Figure 4.14. below is the tree generated for the tree-based identification of the sequence. The matK marker results from BLAST matched with *Quercus pubescens, Quercus helferiana,* and *Quercus coccifera.* The species *Quercus helferiana* had a 100% query cover and 86.88% percent identity overlay in BLAST, however, the distribution of species recorded only includes Asia (Carrero. 2019) *Quercus pubescens,* and *Quercus coccifera* had a query cover 97% and a percent identity match 87.10% for the matK marker. Figure 4.15. below is the tree generated for the tree-based identification of the sequence. Sample code with 13 was morphologically identified as a *Quercus coccifera* by using the PlantNet identification application in the field. Accordingly, the identification of the morphological species fits under the *Fagaceae* family. Although *Quercus coccifera* is distributed in Istanbul (Allmer et al., n.d.), it did not hit a species level in BLAST for the rbcL marker. Therefore, considering both genetic marker identification results with distribution data,

Quercus pubescens had a percent identity got 98.89% in the rbcL marker and 87.10% for the matK marker with the other species under the *Fagaceae* family, the species could be *Quercus pubescens.*

4.11. Aucuba

One sample (code 15) was identified to belong to the genus *Aucuba* by the BLAST method. The query cover for the samples was 100% for both marker rbcL and matK. For the rbcL gene region, sample 15 had a percent identity match of 100% with three different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.16. below shows the tree generated for the tree-based identification of the rbcL sequence. For the matK marker, sample 15 matched with *Aucuba japonica* with a percent identity of 98.08% in the first five hits. Figure 4.17. below is the tree generated for the tree-based identification of the matK sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as an *Aucuba japonica* species. Accordingly, the identification of the morphological species fits under the *Aucuba* genus. Similar to the first five results for the BLAST with both genetic markers, the species *Aucuba japonica* exhibited the highest percent identity. *Aucuba japonica* was distributed in China and Eastern Asia ((Kew science, 2024), introducing the Mediterranean region as a potted plant. Therefore, in BLAST, our percent identity was 100% for the rbcL marker and 98.08% for the matK region under the *Aucuba* genus, the species is likely to be *Aucuba japonica.*

4.12. *Carpinus*

Sample 16 was identified to belong to the genus *Carpinus* by BLAST method. For the rbcL gene region, the query cover was 100%. With three distinct species matched in the first five hits of sample 16, the percent identity was 97.48%. Consequently, a species-level identification was not achievable, and the identification was left at the genus level. Figure 4.18. below is the tree generated for the tree-based identification of the rbcL sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Carpinus orientalis* species. Accordingly, the identification of the morphological species fits under the *Carpinus* genus. Like the first five results in the BLAST, the species *Carpinus orientalis* exhibited a percent identity match of 97.48%. *Carpinus orientalis* was distributed as a native species in Turkey ((Kew science, 2024).) and was identified in Istanbul (Allmer et al., n.d.

Therefore, in BLAST, our percent identity was 97.48% with other species under the *Carpinus* family, the species could be *Carpinus orientalis.*

4.13. *Rubus*

One sample with code 18 was identified to belong to the genus *Rubus* by the BLAST method. The query cover for the samples was 100% for both the rbcL and matK gene regions. Sample 18 had a percent identity match of 100% with five different species in the first five hits for the rbcL gene region, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.19. below is the tree generated for the tree-based identification of the rbcL sequence. For the matK marker, sample 18 had a 96.81% percent identity match with the first five-hit species in the results.

Using the PlantNet identification application, our sample was morphologically identified as a *Rubus pruinosus* species in the field. Accordingly, the identification of the morphological species fits under the *Rubus* genus. However, *Rubus pruinosus* did not match as a species in BLAST for both genetic markers. Moreover, *Rubus pruinosus* is distributed in Europe, but not recorded in Turkey ((Kew science, 2024) Two species from the genus *Rubus* natively distributed in Istanbul, are *Rubus idaeus* and *Rubus caesius*. All three species had the highest percentage identity match of 100% in BLAST for the rbcL marker. However, *Rubus idaeus* and *Rubus caesius* didn't have a match for the matK region. Therefore, the species discrimination was left on the genus level as *Rubus* genus for the rbcL marker.

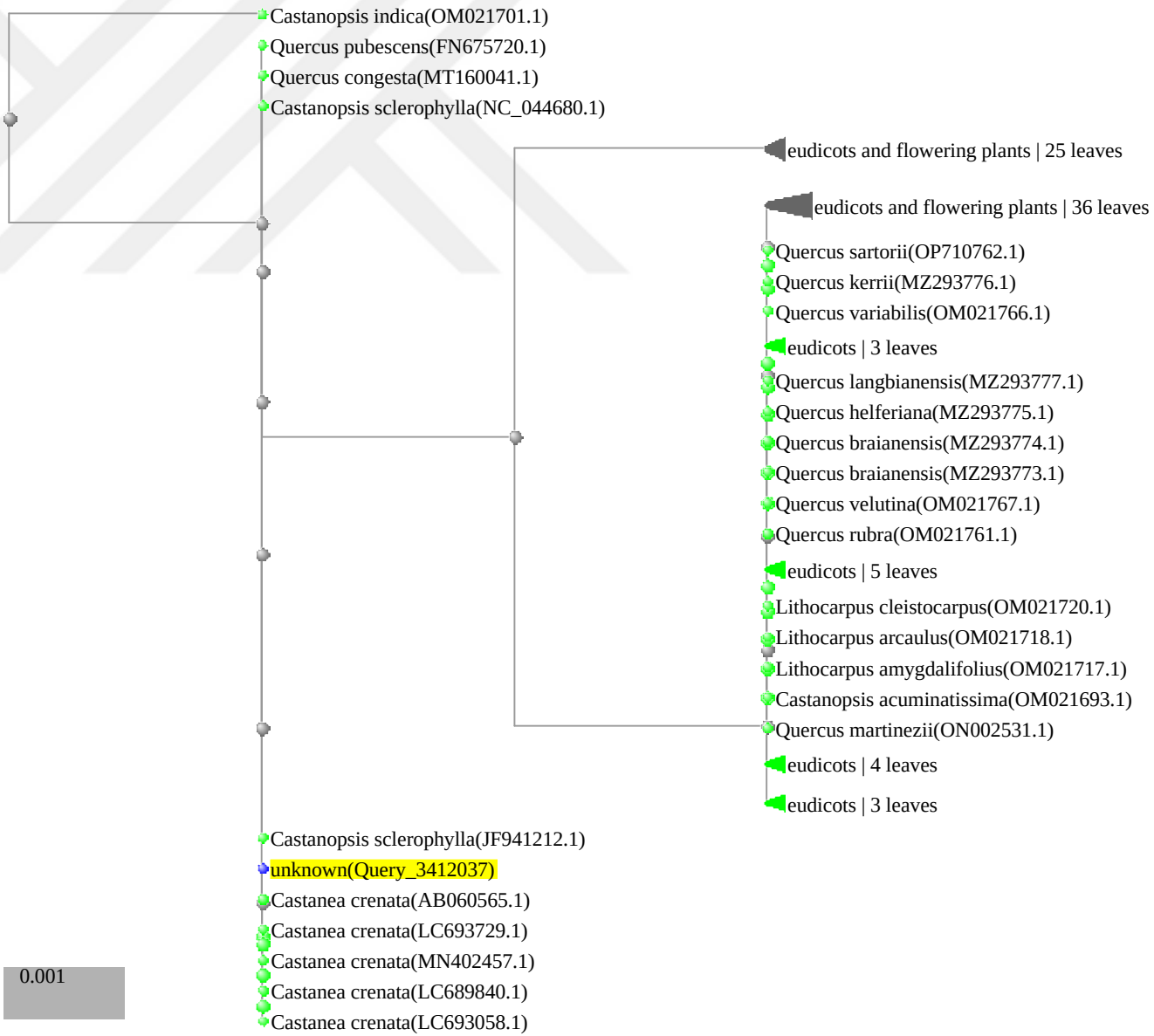


Figure 4.13. The tree-based identification of sample 12 for the rbcL marker. Sample 12 is highlighted in yellow.



Figure 4.14. The tree-based identification of sample 13 for the rbcL marker. Sample 13 is highlighted in yellow.

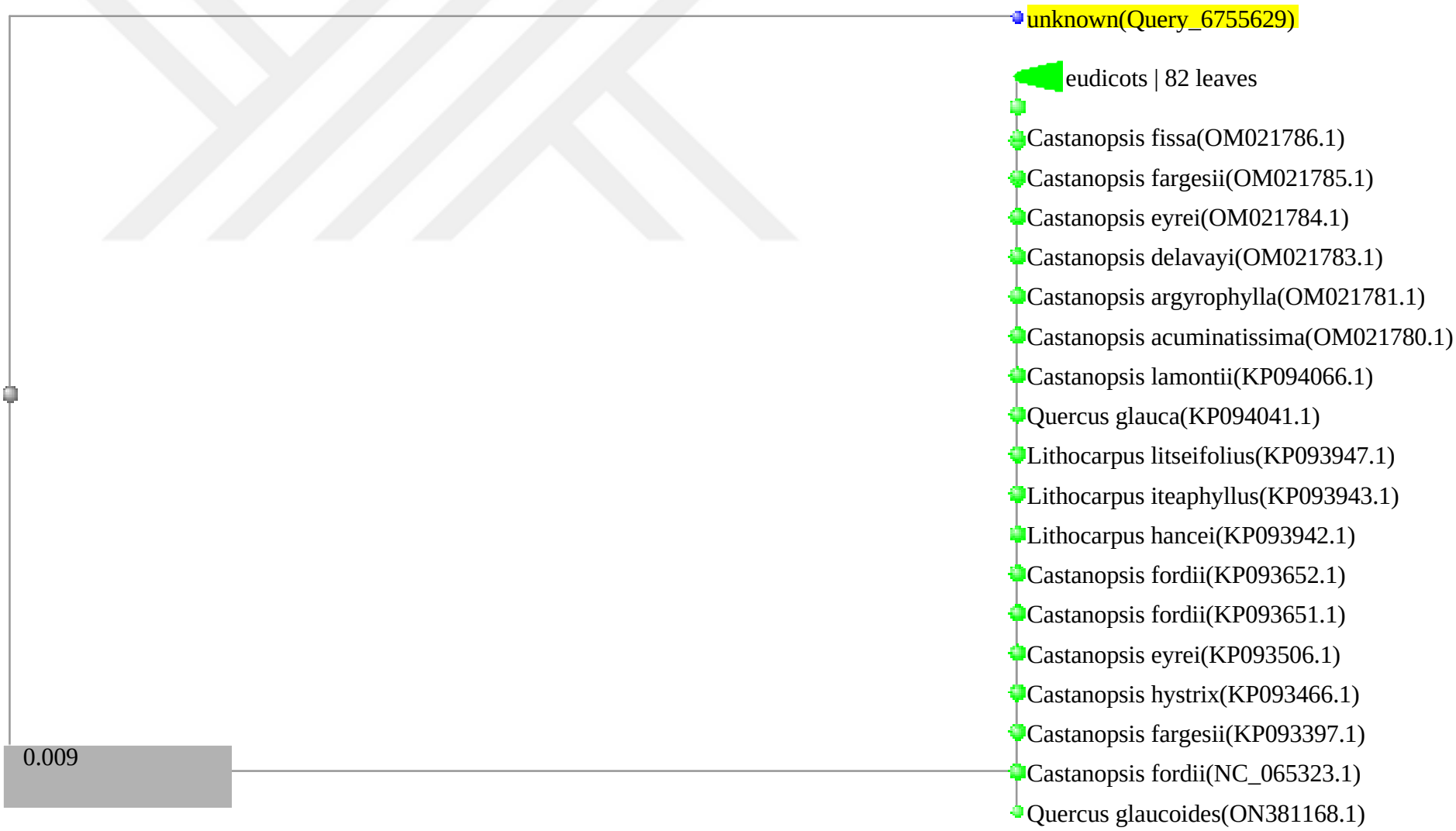


Figure 4.15. The tree-based identification of sample 13 for the matK marker. Sample 13 is highlighted in yellow.

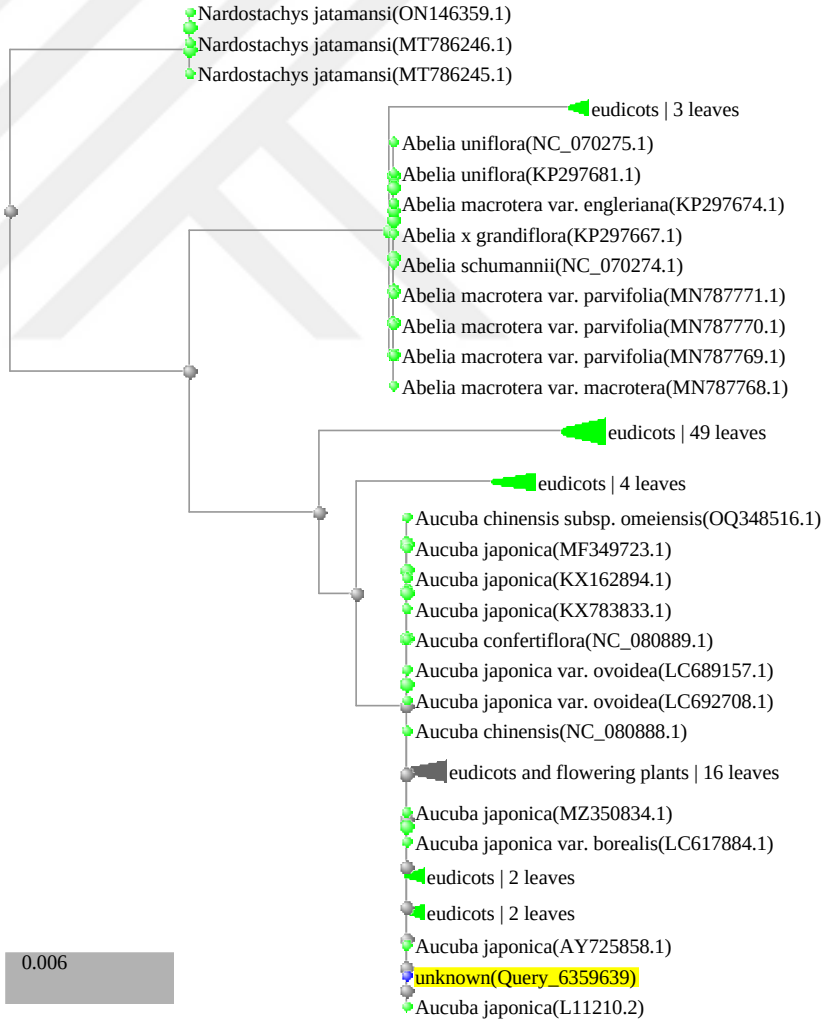


Figure 4.16. The tree-based identification of sample 15 for the rbcL gene region. Sample 15 is highlighted in yellow.

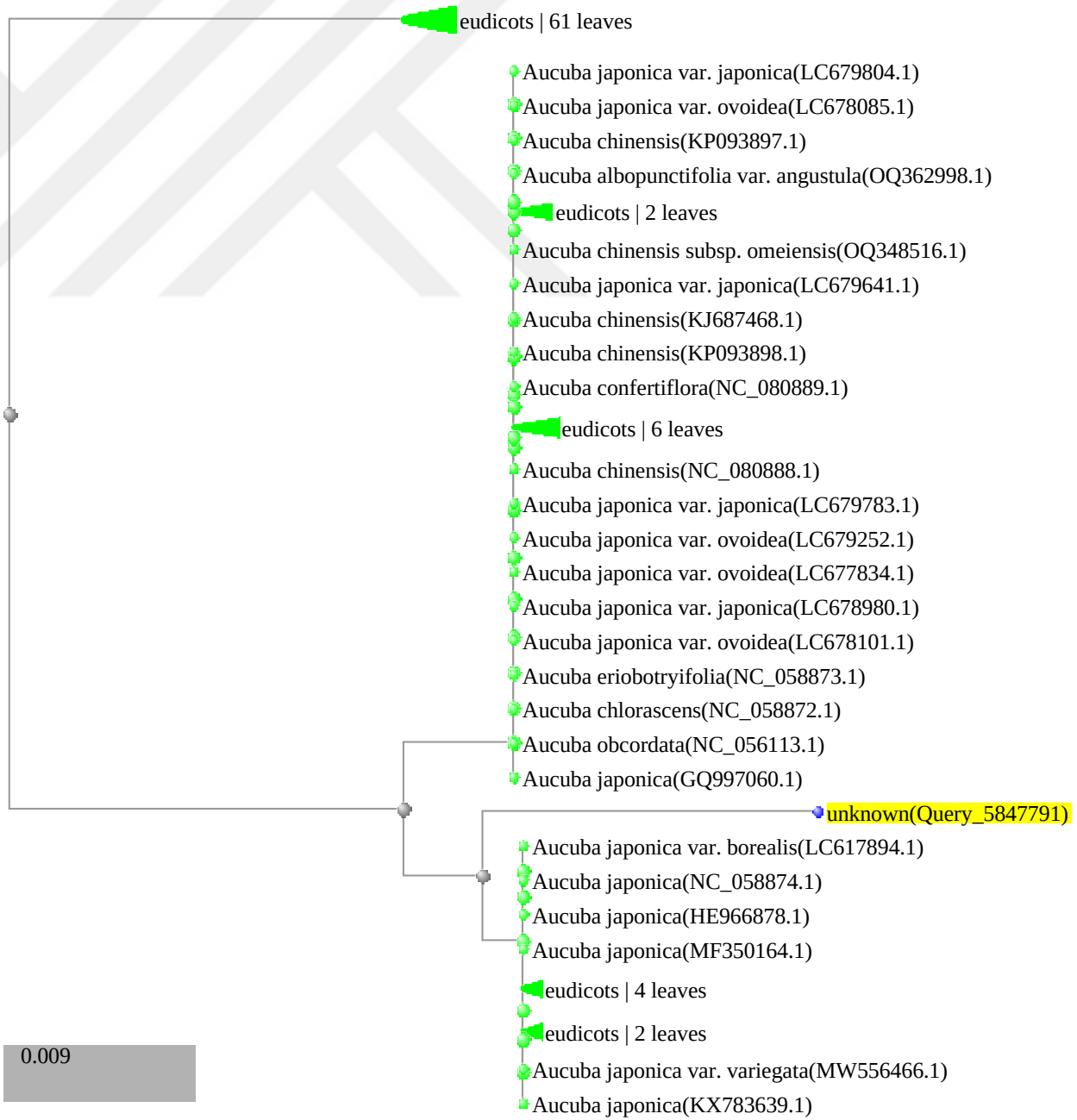

eudicots | 61 leaves
Aucuba japonica var. japonica(LC679804.1)
Aucuba japonica var. ovoidea(LC678085.1)
Aucuba chinensis(KP093897.1)
Aucuba albopunctifolia var. angustula(OQ362998.1)
eudicots | 2 leaves
Aucuba chinensis subsp. omeiensis(OQ348516.1)
Aucuba japonica var. japonica(LC679641.1)
Aucuba chinensis(KJ687468.1)
Aucuba chinensis(KP093898.1)
Aucuba confertiflora(NC_080889.1)
eudicots | 6 leaves
Aucuba chinensis(NC_080888.1)
Aucuba japonica var. japonica(LC679783.1)
Aucuba japonica var. ovoidea(LC679252.1)
Aucuba japonica var. ovoidea(LC677834.1)
Aucuba japonica var. japonica(LC678980.1)
Aucuba japonica var. ovoidea(LC678101.1)
Aucuba eriobotryifolia(NC_058873.1)
Aucuba chlorascens(NC_058872.1)
Aucuba obcordata(NC_056113.1)
Aucuba japonica(GQ997060.1)
unknown(Query_5847791)
Aucuba japonica var. borealis(LC617894.1)
Aucuba japonica(NC_058874.1)
Aucuba japonica(HE966878.1)
Aucuba japonica(MF350164.1)
eudicots | 4 leaves
eudicots | 2 leaves
Aucuba japonica var. variegata(MW556466.1)
Aucuba japonica(KX783639.1)
0.009

Figure 4.17. The tree-based identification of sample 15 for the matK gene region. Sample 15 is highlighted in yellow.

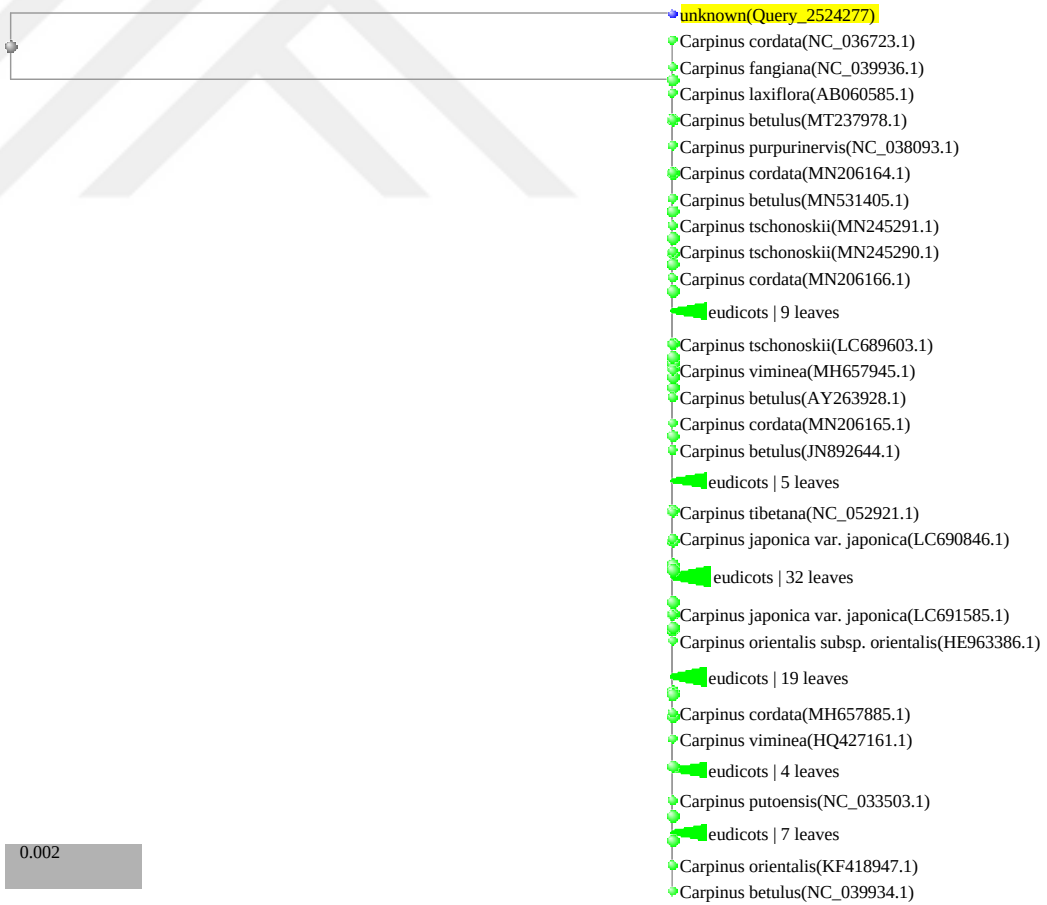


Figure 4.18. The tree-based identification of sample 16 for the rbcL marker. Sample 16 is highlighted in yellow.

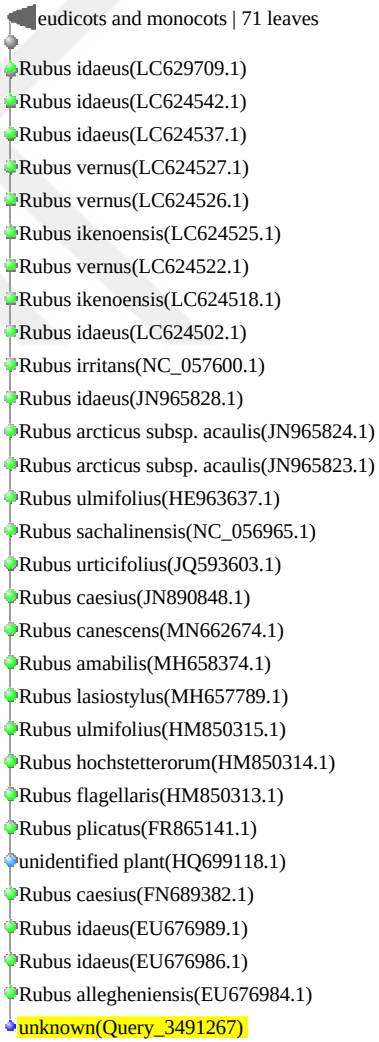


Figure 4.19. The tree-based identification of sample 18 for the rbcL gene sequences. Sample 18 is highlighted in yellow.

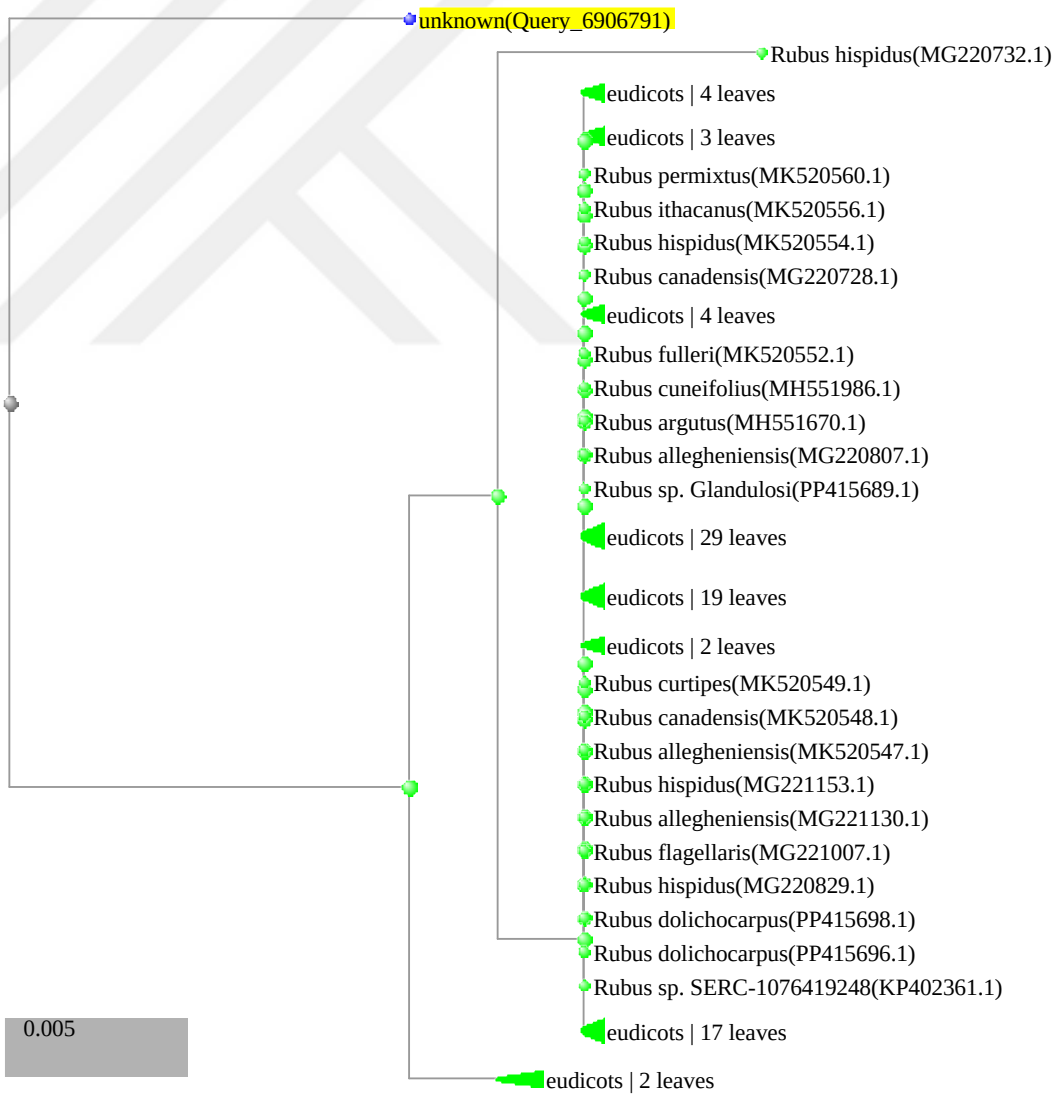


Figure 4.20. The tree-based identification of sample 18 for the matK gene sequences. Sample 18 is highlighted in yellow.

4.14. *Apocynaceae*

One sample with code 19 was identified to belong to the family *Apocynaceae* by BLAST method. The query cover for the samples was 100%. Sample 11 had a percent identity match of >98.93 % with five different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.21. below is the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Nerium oleander* species. Accordingly, the identification of the morphological species fits under the *Apocynaceae* family. Including the first five results in the BLAST, the species *Nerium oleander* exhibited a percent identity match of 100%. *Nerium oleander* is distributed in Turkey, especially in the Mediterranean region (Allmer et al., n.d.; (Kew science, 2024)). Moreover, *Nerium oleander* is widely used for landscape restoration as a flowering pot plant, due to its high adaptability (Ochoa et al., 2009). Therefore, in BLAST, our percent identity got 100% with other species under the *Apocynaceae* family, the species could be *Nerium oleander*.

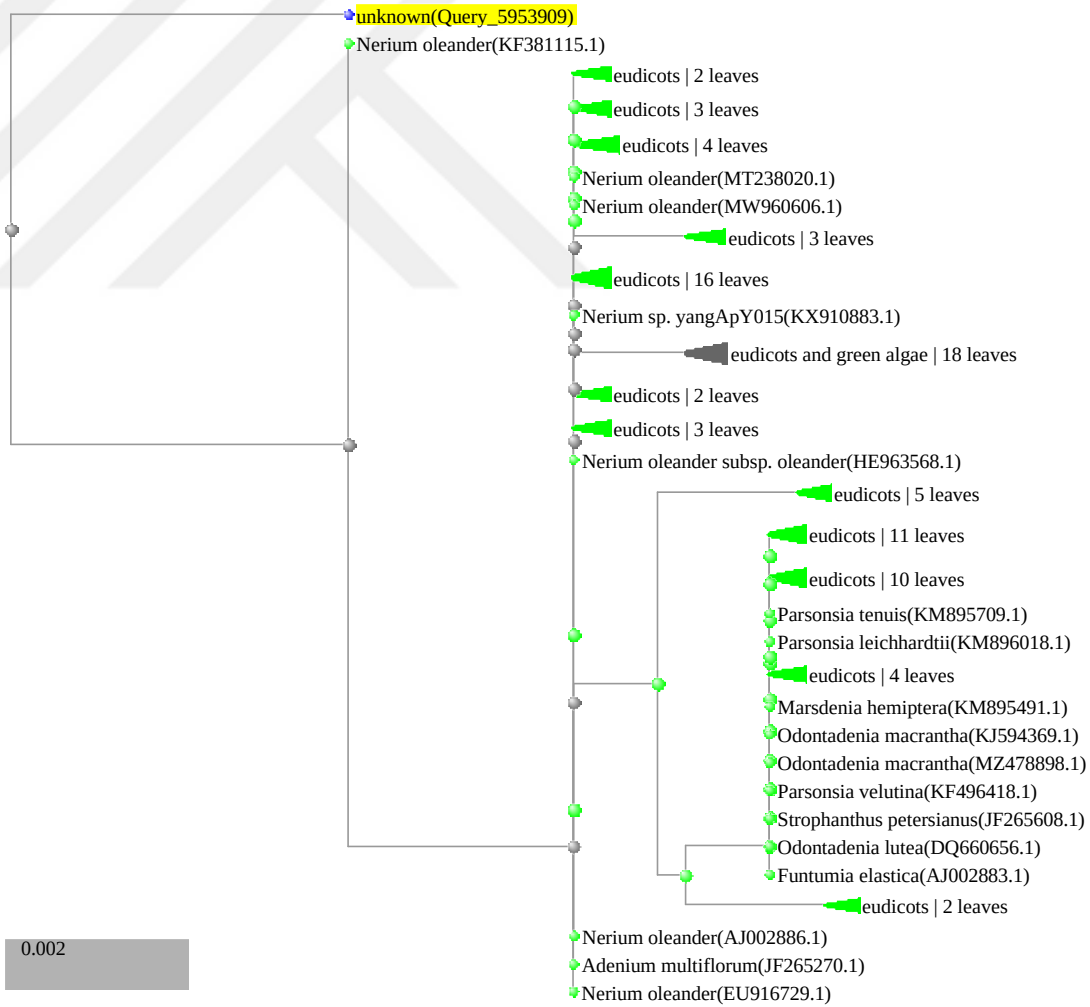


Figure 4.21. The tree-based identification of sample 19. Sample 19 is highlighted in yellow.

4.15. *Cupressaceae*

Sample 29 was identified to belong to the family *Cupressaceae* by BLAST method. The query cover for the samples were 100%. Sample 29 had a percent identity match of 99.49% with three different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.22. below shows the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Juniperus virginiana* species. Accordingly, the identification of the morphological species fits under the *Cupressaceae* family. The first sixteen species hit from five different genera in BLAST exhibited a percent identity match of 99.49%. None of them were recorded in Turkey. The seventeen hit species is *Juniperus communis* had 98.73% and was distributed in Turkey, Marmara region (Allmer et al., n.d.). *Juniperus communis* species is under the *Juniperus* genus with *Juniperus virginiana*. Although the percent identity is lower than the first sixteen hits in BLAST, a match of >95% is acceptable to identify species for the rbcL gene (Bafeel et al., 2012).

4.16. *Oleaceae*

Sample 31 was identified to belong to the family *Oleaceae* by BLAST method. The query cover for the samples were 100%. Sample 31 had a percent identity match of 97.94% with two different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.23. below shows the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as an *Olea europaea* species. Accordingly, the identification of the morphological species fits under the *Oleaceae* family. Similar to the first five results in the BLAST, the species *Olea europaea* exhibited a percent identity match of 97.94%. *Olea europaea* was distributed in Turkey and was identified in Istanbul (Allmer et al., n.d.). Therefore, in BLAST, our percent identity was 97.94% with other species under the *Oleaceae* family, the species could be *Olea europaea*.

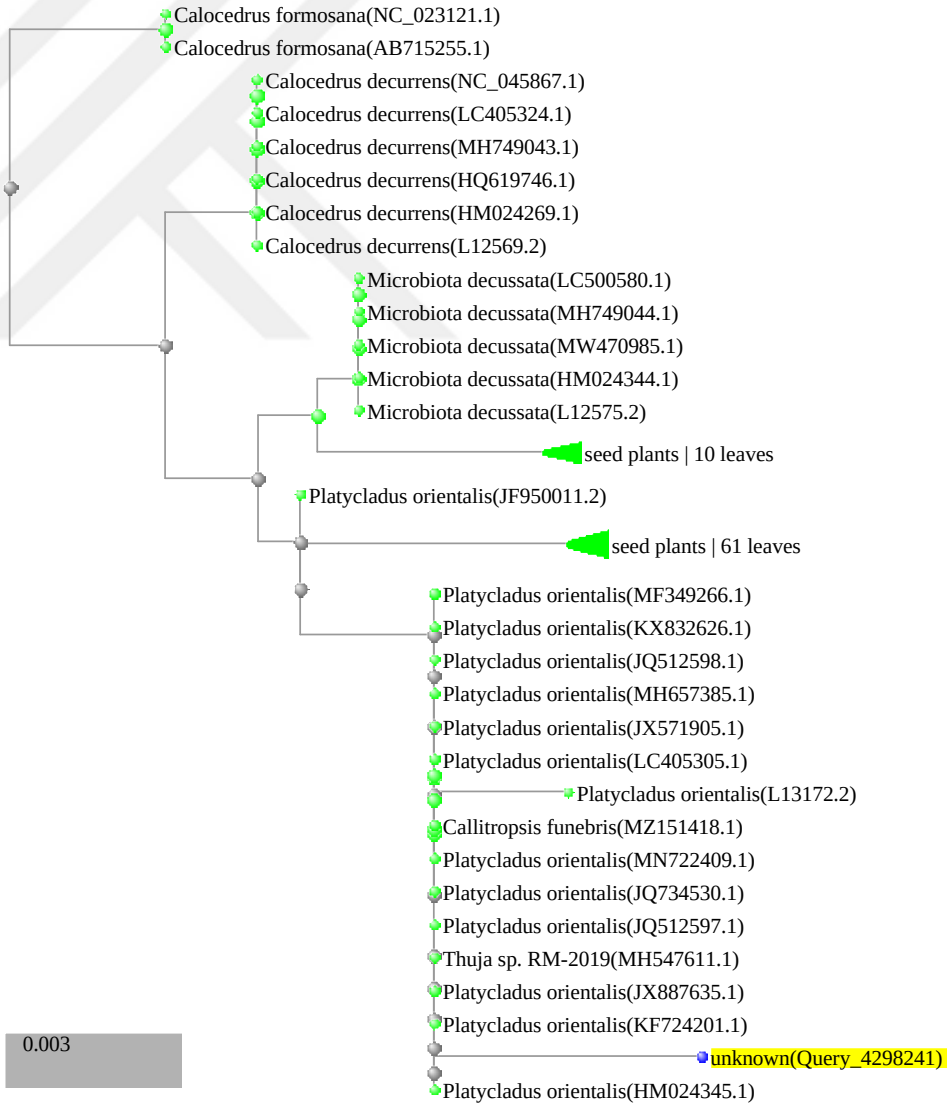


Figure 4.22. The tree-based identification of sample 29 for the rbcL gene region. Sample 29 is highlighted in yellow.

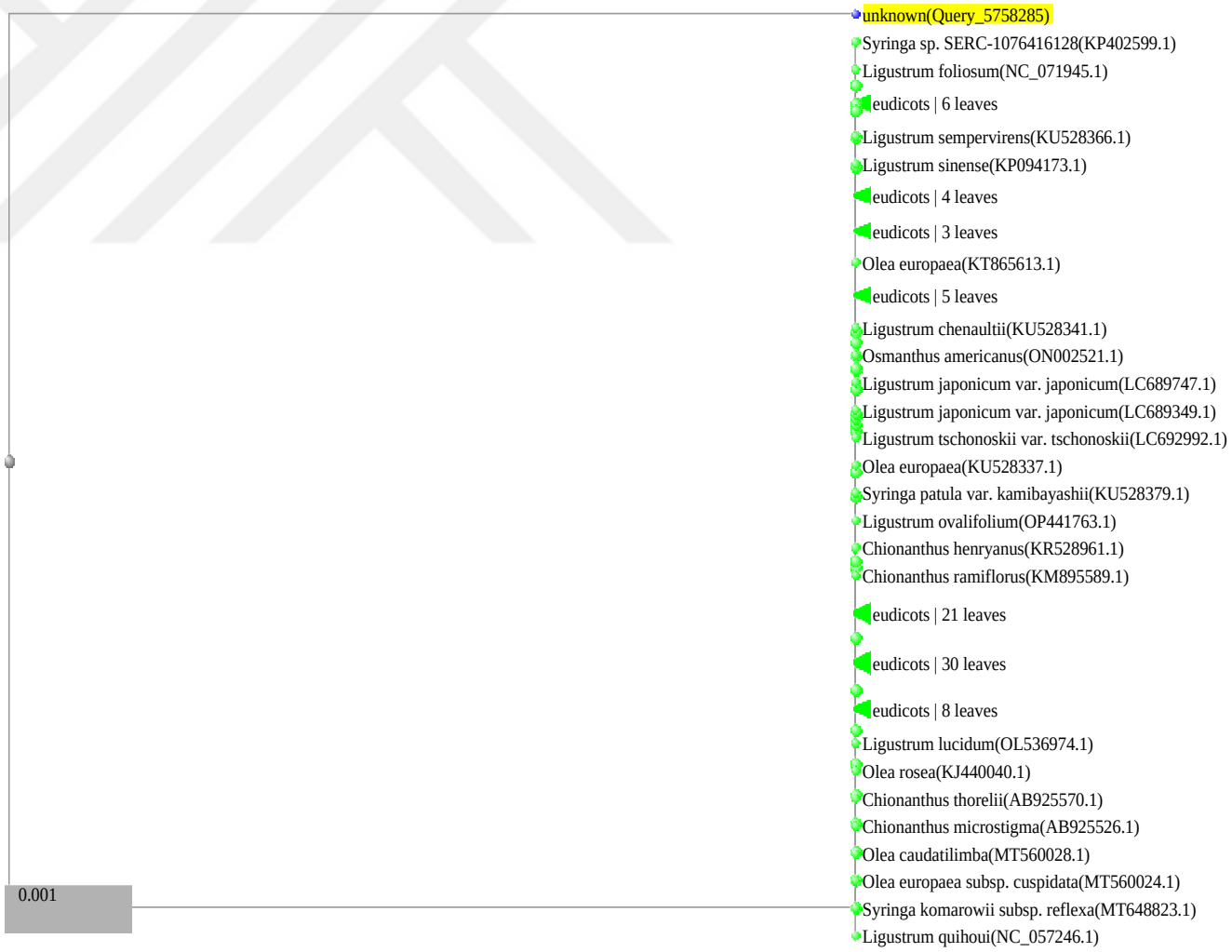


Figure 4.23. The tree-based identification of sample 31 for the rbcL marker. Sample 31 is highlighted in yellow.

4.17. *Lauraceae*

Samples 32 and 33 were identified to belong to the family *Lauraceae* by BLAST method. The query cover for the samples were 100%.

Sample 32 had a percent identity match of 96.83% with five different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.24. below shows the tree generated for the tree-based identification of the sequence. Sample 32 was morphologically identified as a *Laurus nobilis* by using the PlantNet identification application in the field. According to BLAST, Sample 32 exhibited a percent identity match of 96.83%. with eight different genera. Along with all the species only *Laurus nobilis* had a distribution in Istanbul (Allmer et al., n.d.). Similar to the first five results in the BLAST, the species *Laurus nobilis* exhibited a percent identity match of 96.83%. Therefore, considering both genetic identification results with distribution data, *Laurus nobilis* had a percent identity got 96.83% with other species under the *Lauraceae* family, the species could be *Laurus nobilis.*

Sample 33 had a percent identity match of 99.74% with five different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.25. below shows the tree generated for the tree-based identification of the sequence. Sample 46 was morphologically not identified by using the PlantNet identification application in the field, labelled as an unknown. According to BLAST, Sample 33 exhibited a percent identity match of 99.74%. with ten different genera. Along with all the species only *Laurus nobilis* had a distribution in Istanbul (Allmer et al., n.d.) and a native species for Turkey ((Kew science, 2024). *Laurus nobilis* exhibited a percent identity match of 99.74%. Therefore, considering both genetic identification results with distribution data, *Laurus nobilis* had a percent identity got 99.74% with other species under the *Lauraceae* family, the species could be *Laurus nobilis.*

4.18. *Malvaceae*

One sample with code 46 was identified to belong to the family Malvaceae by BLASTmethod. The query cover for the samples were 100%. Sample 46 had a percent identity match of 96.22% with five different species in the first five hits, therefore a species or genus level identification was not possible, and the identification was left at the family level. Figure 4.20. below shows the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Tilia sp.* species. Accordingly, the identification of the morphological species fits under the *Malvaceae* family. Like the first five results in the BLAST, the species *Tilia cordata* exhibited a percent identity match of 96.22%. *Tilia cordata* is natively distributed in Turkey and were identified in Istanbul (Allmer et al., n.d.). Therefore, in BLAST, our percent identity was 96.22% with other species under the *Malvaceae* family, the species could be *Tilia cordata*.

4.19. *Quercus*

Sample 40 was identified to belong to the genus *Quercus* by BLAST method. The query cover for the samples were 100%. Sample 40 had a percent identity match of 100% with five different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.26. below is the tree generated for the tree-based identification of the sequence.

Using the PlantNet identification application in the field, our sample was morphologically identified as a *Quercus coccifera* species. Accordingly, the identification of the morphological species fits under the *Quercus* genus. Although *Quercus coccifera* was distributed in Istanbul (ref. TÜBİVES), it did not match as a species in BLAST. Two species from the genus *Quercus*, *Quercus robur* subsp. *robur* and *Quercus ilex* had the highest percent identity match of 100% and both were natively distributed in Istanbul (Allmer et al., n.d.). Therefore, in BLAST, our percent identity got 100% with other species under the *Quercus* genus, the species could be *Quercus robur* subsp. *robur* or *Quercus ilex.*

4.20. *Viburnum*

One sample with code 43 was identified to belong to the genus *Viburnum* by BLAST method. The query cover for the samples were 100%. Sample 43 had a percent identity match of 99.78% with three different species in the first five hits, therefore a species-level identification was not possible, and the identification was left at the genus level. Figure 4.27. below shows the tree generated for the tree-based identification of the sequence.

Sample 43 was identified to belong to the genus *Viburnum.* Using the PlantNet identification application in the field, our sample was morphologically identified as a *Viburnum tinus* species. Accordingly, the identification of the morphological species fits under the *Viburnum* genus. Similar to the first five results in the BLAST, the species *Viburnum tinus* exhibited a percent identity match

of 99.78%. *Viburnum tinus* was distributed in Turkey (Allmer et al., n.d.) and was identified in cemeteries of Istanbul (Servet & Aktağ, 2019). Therefore, in BLAST, our percent identity was 99.78% with other species under the *Viburnum* family, the species could be *Viburnum tinus*.



Figure 4.24. The tree-based identification of sample 32 for the rbcL sequences. Sample 32 is highlighted in yellow.

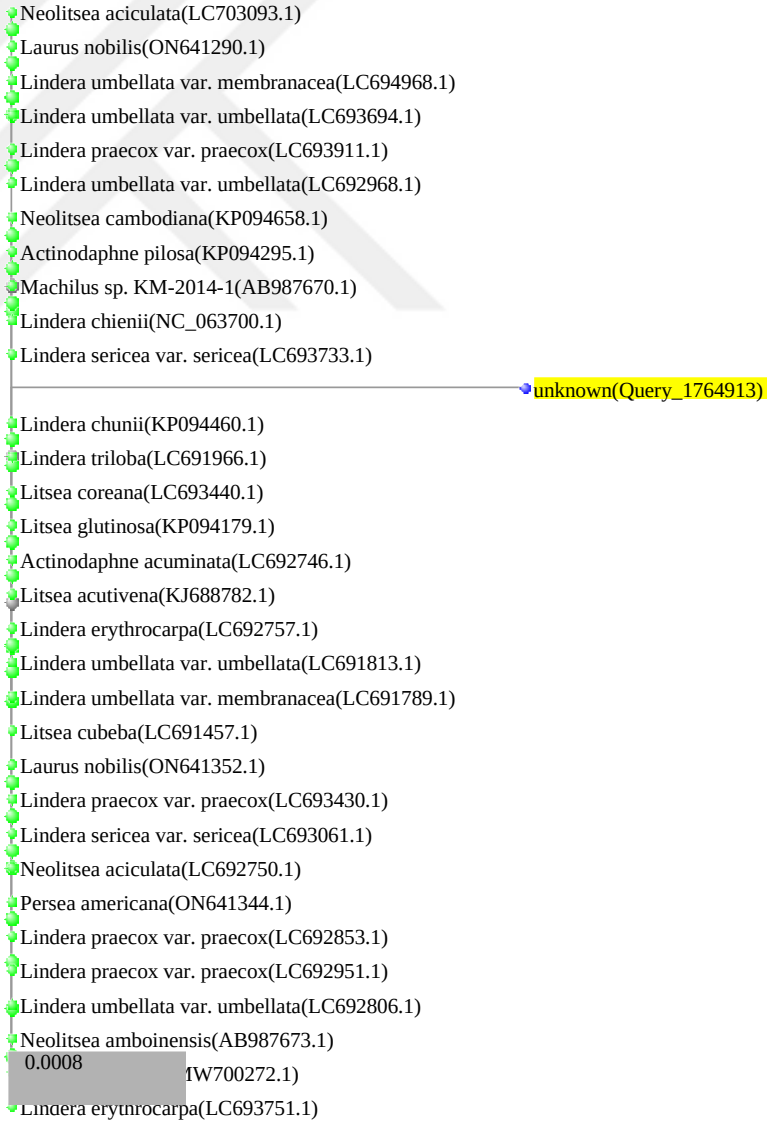


Figure 4.25. The tree-based identification of sample 46 for the rbcL sequences. Sample 46 is highlighted in yellow.



Figure 4.26. The tree-based identification of sample 33 for the rbcL gene region. Sample 33 is highlighted in yellow.

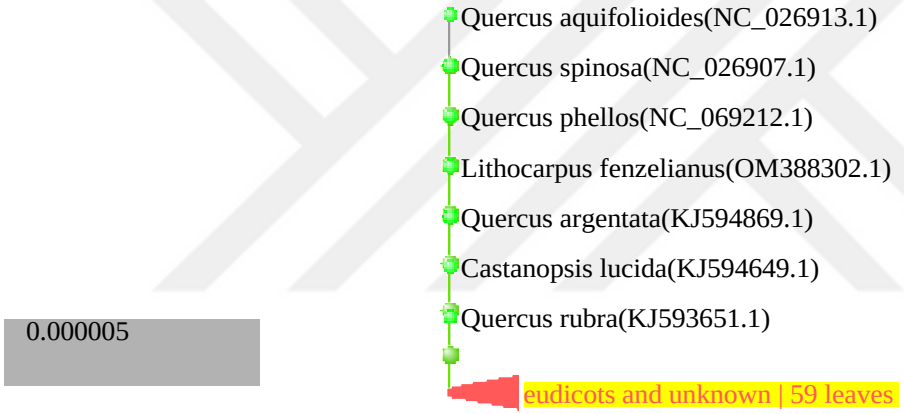


Figure 4.27. The tree-based identification of sample 40 for the rbcL gene region. Sample 40 is highlighted in yellow.

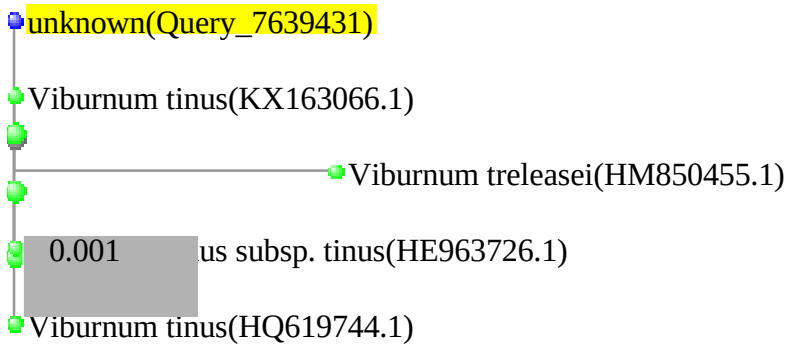


Figure 4.28. The tree-based identification of sample 43 for the rbcL marker. Sample 43 is highlighted in yellow.

5. CONCLUSION AND RECOMMENDATIONS

As a part of the barcoding effort for this project, twenty-two clean sequences for the rbcL gene region and six sequences for the matK loci were received. These were identified down to seven families and thirteen genera. Genetic identification and taxonomical identification data were evaluated and filtered for distribution data for Istanbul in the literature. In twenty-two samples, seventeen native and five introduced taxa were detected for Turkey. All native species were previously recorded in Istanbul and likely represent individuals from the remnant forest areas within the Robert College Campus. Seven family and thirteen genera that were identified belong to the *Juglandaceae, Fagaceae, Apocynaceae, Cupressaceae, Oleaceae, Lauraceae, Malvaceae* families and *Cercis, Fraxinus*, *Pistacia, Celtis, Ficus, Morus, Paulownia, Cinnamomum, Aucuba, Carpinus, Rubus, Quercus, Viburnum* genera.

We used the rbcL and the matK markers for DNA sequencing, amplification, and species discrimination for the tree and shrub species that were the subject of this investigation. In this study, rbcL + matK genes were insufficient to discriminate samples at the species level but provided discrimination at the genus and family levels. For future studies, we recommend using another marker, such as ITS, or whole mitogenome sequencing using next-generation sequencing approaches for potential species-level identification.

As a citizen science initiative, 52 high school students were involved in the data collection step, and 36 of them joined the DNA extraction, PCR, and gel-electrophoresis steps. Our project requires participants to have a high degree of expertise, a lot of training and assistance, and higher degrees of involvement. We employed informal and formal science education methods in this project to help their comprehension. This project empowered the high school students to take part in ongoing research and contribute the global scientific data. We recommend using the citizen science projects implemented for high school students for data collection in DNA Barcoding studies.

6. REFERENCES

Allmer, J., Annual IEEE Computer Conference, International Symposium on Health Informatics and Bioinformatics 6 2011.05.02-05 Izmir, & HIBIT 6 2011.05.02-05 Izmir. (n.d.). *2011 6th International Symposium on Health Informatics and Bioinformatics (HIBIT) 2-5 May 2011, Izmir, Turkey*.

Bafeel, S. O., Arif, I. A., Bakir, M. A., Al Homaidan, A. A., Al Farhan, A. H., & Khan, H. A. (2012). DNA barcoding of arid wild plants using rbcL gene sequences. *Genetics and Molecular Research : GMR*, *11*(3), 1934–1941. https://doi.org/10.4238/2012.July.19.12

Bhagya Chandrasekara, C. H. W. M. R., Naranpanawa, D. N. U., Bandusekara, B. S., Pushpakumara, D. K. N. G., Wijesundera, D. S. A., & Bandaranayake, P. C. G. (2021). Universal barcoding regions, rbcL, matK and trnH-psbA do not discriminate Cinnamomum species in Sri Lanka. *PLoS ONE*, *16*(2 February). https://doi.org/10.1371/journal.pone.0245592

Bottoni, M., Milani, F., Mozzo, M., Kolloffel, D. A. R., Papini, A., Fratini, F., Maggi, F., & Santagostini, L. (2021). Sub-tissue localization of phytochemicals in cinnamomum camphora (L.) j. presl. growing in northern Italy. *Plants*, *10*(5). https://doi.org/10.3390/plants10051008

Braukmann, T. W. A., Kuzmina, M. L., Sills, J., Zakharov, E. V., & Hebert, P. D. N. (2017). Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada. *PLoS ONE*, *12*(1). https://doi.org/10.1371/journal.pone.0169515N

Büyükkılıç, N. (1980) *Robert Kolej.pdf*. (n.d.).

Camille Garcin, A. J. P. B. A. A. J.-C. L. M. C. M. S. T. L. & J. Salmon. (2021). P.-300K image dataset (1. 1) [Data set]. Zenodo. (n.d.). *PlantNet*. Https://Doi.Org/10.5281/Zenodo.5645731.

Carrero, C., Jerome, D., Beckman, E., Byrne, A., Coombes, A. J., Deng, M., González Rodríguez, A., Van Sam, H., Khoo, E., Nguyen, N., Robiansyah, I., Rodríguez Correa, H., Sang, J., Song, Y.-G., Strijk, J., Sugau, J., Sun, W., Valencia-Ávalos, S., & Westwood, M. (2020). *Oaks 2020*. www.seascapedesign.co.uk

Caudullo, G., & Durrant, T. H. (n.d.). *Fraxinus angustifolia*. https://w3id.org/mtv/FISE-Comm/v01/e0101d2.

Çiçek, E., & Yılmaz, M. (2002). The importance of Fraxinus angustifolia subsp. oxycarpa as a fast growing tree for Turkey. *IUFRO Meeting on Management of Fast Growing Plantations, September*, 11–13.

Faeth, S. H., Bang, C., & Saari, S. (2011). Urban biodiversity: Patterns and mechanisms. *Annals of the New York Academy of Sciences*, *1223*(1), 69–81. https://doi.org/10.1111/j.1749-6632.2010.05925.x

Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C. H., Newmaster, S. G., Hajibabaei, M., & Husband, B. C. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, *9*(SUPPL. 1), 130–139. https://doi.org/10.1111/j.1755-0998.2009.02652.x

Fisher Scientific, T. (n.d.). *Qubit ® 3.0 Fluorometer Catalog Number Q33216*.

Gunn, B. F., Aradhya, M., Salick, J. M., Miller, A. J., Yongping, Y., Lin, L., & Xian, H. (2010). Genetic variation in walnuts (Juglans regia and J. sigillata; Juglandaceae): Species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *American Journal of Botany*, *97*(4), 660–671. https://doi.org/10.3732/ajb.0900114

Hakan Akyildiz, M., & Kol, H. S. (2010). Some technological properties and uses of paulownia (Paulownia tomentosa Steud.) wood. In *Journal of Environmental Biology*.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Henter, H. J., Imondi, R., James, K., Spencer, D., & Steinke, D. (2016). DNA barcoding in diverse educational settings: Five case studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 1–9. https://doi.org/10.1098/rstb.2015.0340

Carrero, C. 2019. Quercus helferiana. The IUCN Red List of Threatened Species 2019 e.T194136A2302315. http://dx.doi.org/10.2305/IUCN.UK.2019 1.RLTS.T194136A2302315.en

Ives, C. D., Lentini, P. E., Threlfall, C. G., Ikin, K., Shanahan, D. F., Garrard, G. E., Bekessy, S. A., Fuller, R. A., Mumaw, L., Rayner, L., Rowe, R., Valentine, L. E., & Kendal, D. (2016). Cities are hotspots for threatened species. *Global Ecology and Biogeography*, *25*(1), 117–126. https://doi.org/10.1111/geb.12404

Kress, W. J., & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS ONE*, *2*(6). https://doi.org/10.1371/journal.pone.0000508

Lepczyk, C. A., Aronson, M. F. J., Evans, K. L., Goddard, M. A., Lerman, S. B., & Macivor, J. S. (2017). Biodiversity in the City: Fundamental Questions for Understanding the Ecology of Urban Green Spaces for Biodiversity Conservation. In *BioScience* (Vol. 67, Issue 9, pp. 799–807). Oxford University Press. https://doi.org/10.1093/biosci/bix079

Hussain, S., Asghar, R., Ajaib, M., Ali, I., Siddiqui, M. F., Ain, S., & Sufian, M. (2023). DNA Barcode for Phylogenetic Analysis of Genus Morus Species from Azad Jammu and Kashmir. *Pakistan Journal of Botany*, *55*(6), 2211–2220. https://doi.org/10.30848/PJB2023-6(18)

Malaise, R. (1937). A new insect-trap. *Entomologisk Tidskrift*, *58*, 148–160.

Marizzi, C., Florio, A., Lee, M., Khalfan, M., Ghiban, C., Nash, B., Dorey, J., McKenzie, S., Mazza, C., Cellini, F., Baria, C., Bepat, R., Cosentino, L., Dvorak, A., Gacevic, A., Guzman-Moumtzis, C., Heller, F., Holt, N. A., Horenstein, J., … Micklos, D. A. (2018). DNA barcoding Brooklyn (New York): A first assessment of biodiversity in Marine Park by citizen scientists. *PLoS ONE*, *13*(7), 1–18. https://doi.org/10.1371/journal.pone.0199015

Mcdonnell, M. J., & Pickett, S. T. A. (1990). Ecosystem Structure and Function along Urban-Rural Gradients: An Unexploited Opportunity for Ecology. In *Source: Ecology* (Vol. 71, Issue 4).

Ochoa, J., Franco, J. A., Bañón, S., & Fernández, J. A. (2009). Distribution in plant, substrate and leachate of paclobutrazol following application to containerized Nerium oleander L. seedlings. In *INIA) Spanish Journal of Agricultural Research* (Vol. 7, Issue 3). www.inia.es/sjar

Pautasso, M. (2007). Scale dependence of the correlation between human population presence and vertebrate and plant species richness. *Ecology Letters*, *10*(1), 16–24.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System: Barcoding. *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ruch, D. G., Karns, D. R., Mcmurray, P., Moore-palm, J., & Murphy, W. (2010). Results of the Loblolly Marsh Wetland Preserve BioBlitz, Jay County, Indiana. *Proceedings of the Indiana Academy of Science*, *119*(1), 1–3.

Santschi, L., Hanner, R. H., Ratnasingham, S., Riconscente, M., & Imondi, R. (2013). Barcoding Life's Matrix: Translating Biodiversity Genomics into High School Settings to Enhance Life Science Education. *PLoS Biology*, *11*(1). https://doi.org/10.1371/journal.pbio.1001471

Servet, C., & Aktağ, A. (2019). The composition of woody plants in the cemeteries of various religious communities in Istanbul-Turkey. *Urban Forestry and Urban Greening*, *43*. https://doi.org/10.1016/j.ufug.2019.05.012

Steinke, D., Breton, V., Berzitis, E., & Hebert, P. D. N. (2017). The School Malaise Trap Program: Coupling educational outreach with scientific discovery. *PLoS Biology*, *15*(4), 1–8. https://doi.org/10.1371/journal.pbio.2001829

Tosh, J., James, K., Rumsey, F., Crookshank, A., Dyer, R., & Hopkins, D. (2016). Is DNA barcoding child's play? Science education and the utility of DNA barcoding for the discrimination of UK tree species. *Botanical Journal of the Linnean Society*, *181*(4), 711–722. https://doi.org/10.1111/boj.12449

Uchida, K., Blakey, R. V., Burger, J. R., Cooper, D. S., Niesner, C. A., & Blumstein, D. T. (2021). Urban biodiversity and the importance of scale. *Trends in ecology & evolution*, *36*(2), 123-131.

Kew science. Plants of the World Online. (2024). https://powo.science.kew.org/

Xiang, X. G., Zhang, J. B., Lu, A. M., & Li, R. Q. (2011). Molecular identification of species in Juglandaceae: A tiered method. *Journal of Systematics and Evolution*, *49*(3), 252–260. https://doi.org/10.1111/j.1759-6831.2011.00116.x

Yu, J., Wu, X., Liu, C., Newmaster, S., Ragupathy, S., & Kress, W. J. (2021). Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety*, *208*. https://doi.org/10.1016/j.ecoenv.2020.111691