YEDITEPE UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DNA BARCODING FOR THE IDENTIFICATION OF *VERBASCUM* L. FROM TURKEY

A Thesis Submitted
by
İbrahim Sercan Başıbüyük

In Partial Fullfillment
of the Requirements for the Degree of
Master of Sciences
In Biotechnology

Supervisor
Assist. Prof. Dr. Bahar Soğutmaz Özdemir

Istanbul - 2023

DNA BARCODING FOR THE IDENTIFICATION OF *VERBASCUM* L. FROM TURKEY

by

İbrahim Sercan Başıbüyük

APPROVED BY:

Assist. Prof. Dr. Bahar Soğutmaz Özdemir ……
(Thesis Supervisor)
(Yeditepe University)

Assoc. Prof. Dr. Emrah Nikerel ……
(Yeditepe University)

Assist. Prof. Dr. Merve Seven ……
(Bahçeşehir University)

DATE OF APPROVAL: …./…./20…

DECLARATION OF ORIGINALITY

I hereby declare that this thesis is my own work and that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I have fully cited and referenced all material and results as required by these rules and conduct, and this thesis study does not contain any plagiarism. If any material used in the thesis requires copyright, the necessary permissions have been obtained. No material from this thesis has been used for the award of another degree.

I accept all kinds of legal liability that may arise in case contrary to these situations.

Name, Last name İbrahim Sercan Başıbüyük

Signature …..…….……..

ABSTRACT

DNA BARCODING FOR THE IDENTIFICATION OF *VERBASCUM* L. FROM TURKEY

The increasing industrialization and environmental degradation lead to a planetary biodiversity crisis further leading to species extinction, thus the loss of genetic resources and germplasm. By preserving the genetic resources, it is also possible to protect the ecosystem which highly depends on taxonomy and the accurate identification of the species. The climate, topography, and geomorphology of Turkey make it to be one of the centers for plant diversity and endemism. It is estimated that there are approximately 2700 endemic species in Turkey where the *Verbascum* L. genus is occupying an important role with its distinctive biodiversity. The distribution range of the *Verbascum* L. genus is narrow, and the increasing risk of its habitat destruction creates a requirement for a strict conservation policy considering its therapeutic qualities and ecological relevance. Due to the species physical similarity and possibility for hybridization, these endemic *Verbascum* L. species pose difficulties for correct identification and classification. In order to clarify the taxonomic relationships, boundaries and conservation of *Verbascum* L. germplasm, it is crucial to use DNA barcoding technique which utilizes short and conserved DNA sequences with sufficient variations known as DNA barcodes to enable the molecular-level identification and differentiation of species. The objective of this study is the molecular identification of key species of the *Verbascum* L. genus from Turkey by DNA barcoding. For this purpose, the Anatolian endemic species from the *Verbascum* L. genus were selected mainly from the herbarium collections. The recommended barcode regions by CBOL, namely *rbcL*, *matK*, *ITS*, and *trnH-psbA*, which enable detection with greater variation, were utilized. The DNA sequencing and bioinformatic analysis were conducted following the established DNA barcoding standards. The barcode regions, *ITS* and *matK*, were identified in 15 and 11 *Verbascum* species respectively and used for the phylogenetic analysis. This study will enable the addition of *Verbascum* L. to Turkey's plant inventory and make molecular-level adjustments to correct taxonomic problems if necessary. Consequently, the involvement of the DNA barcodes will further contribute to the conservation of the *Verbascum* germplasm.

ÖZET

TÜRKİYE'YE AİT *VERBASCUM* L. TÜRLERİNİN DNA BARKODLAMA YÖNTEMİ İLE KİMLİKLENDİRİLMESİ

Endüstrileşme ve çevresel bozulmanın artması, gezegende biyolojik çeşitlilik krizine yol açarak türlerin yok olmasına, dolayısıyla genetik kaynaklarının kaybedilmesine sebep olmaktadır. Genetik kaynakların korunması ile taksonomiye ve türlerin doğru tespitine büyük ölçüde bağımlı olan ekosistemi de korumak mümkün hale gelmektedir. İklim, topografi ve jeomorfoloji gibi öğeler, Türkiye'yi en yüksek bitki çeşitliliği ve endemizme sahip olan merkezlerden biri haline getirmektedir. Türkiye'de yaklaşık olarak 2700 endemik tür olduğu tahmin edilmektedir ve bu çeşitliliğin önemli bir kısmını *Verbascum* L. cinsi oluşturmaktadır. Yayılım alanı dar olduğundan, habitatının yok olma riski yüksek olan *Verbascum* L. cinsi, terapötik özellikleri ve ekolojik önemi ile ciddi bir koruma politikasına ihtiyaç duymaktadır. Türler arası morfolojik benzerlikleri ve hibridizasyon olasılıkları sebebiyle endemik *Verbascum* L. türlerinin doğru tespiti ve sınıflandırılmasında zorluklar yaşanmaktadır. Taksonomik ilişkileri ve türlerin sınırlarını netleştirmek ve *Verbascum* L. gen kaynaklarını korumak ve tür ayrımı sağlamak için moleküler seviyede tanımlama yapılması büyük önem arz etmektedir. DNA barkodlama yöntemi, DNA barkodları adı verilen kısa ve yeterli varyasyona sahip korunmuş DNA dizilerini kullanarak yüksek seviyede tür ayrımı sağlayabilmektedir. Bu çalışmanın amacı, Türkiye'deki *Verbascum* L. cinsinin anahtar türlerinin DNA barkodlama yöntemiyle moleküler seviyede tanımlanmasıdır. Bu amaçla, herbaryum koleksiyonlarından, *Verbascum* L. cinsine ait Anadolu endemik türleri seçilmiştir. Tür tanımlamasını, DNA barkodlama ile standardize edilmesini destekleyen CBOL tarafından önerilen *rbcL*, *matK*, *ITS* ve *trnH-psbA* barkod bölgeleri kullanılmıştır. DNA sekanslama ve biyoinformatik analizler, kurulmuş olan DNA barkodlama standartlarına göre gerçekleştirilmiştir. *ITS* ve *matK* barkod bölgeleri, sırasıyla 15 ve 11 *Verbascum* L. türünde tanımlanmıştır ve filogenetik analizlerde kullanılmıştır. Bu çalışma, *Verbascum* L.'nin Türkiye bitki envanterine eklenmesine ve gerektiğinde taksonomik sorunları düzeltmek için moleküler düzeyde düzeltmeler yapılmasına ve DNA barkodlarının literatüre dahil edilmesi ile *Verbascum* L. gen kaynaklarının korunmasına katkı sağlayacaktır.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere appreciation to my thesis supervisor, Assist. Prof. Bahar Soğutmaz Özdemir, for her extensive experience and invaluable guidance throughout my Master's Thesis. Her thoughtful advice and the opportunity to work in her laboratory have been crucial to the successful completion of my research. Also special thanks to Assoc. Prof. Emrah Nikerel for his wisdom and help about all the analysis process.

I am grateful to Aylin Kobaneri and Zeynep Işık from the Yeditepe University Plant Tissue Culture and Genetics Laboratory. Their unwavering support, dedication, and mentorship have been instrumental throughout the research process. They have not only provided me with essential resources and guidance within the laboratory but also shared their profound expertise and insights, significantly contributing to the quality and validity of this thesis. Their collaborative spirit and passion for scientific inquiry have been an inspiration to me.

Finally, I would like to express my heartfelt thanks to my beautiful girlfriend Miray Yiğit for all the things that I can't even begin to explain. She was there with me from start to beginning. Also, thanks to My dearest sister İpeksu Başıbüyük for all the motivation and support. My best friends Sinan Bozkurt and Ataberk Keser were there whenever I needed them and of course my parents Fethi Başıbüyük and Seher Başıbüyük, whose constant love, guidance, and support have been the foundation of my accomplishments. They have always believed in my abilities, offering encouragement during times when I doubted myself. Their understanding of the sacrifices I have made to pursue this endeavor has provided me with unwavering strength. Their presence in my life has created a support system that has helped me navigate the challenges of academia while maintaining a balanced outlook. I am sincerely grateful for their patience, understanding, and unwavering support, as they have played a crucial role in fueling my determination to overcome obstacles and successfully complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

°C	Degree centigrade
Bp	Base pair
Ca^{++}	Calcium
Gb	Gigabase
Mbp	Megabase pair
µg	Microgram
µg/ml	Microgram per milliliter
µM	Micromolar
ml	Milliliter
mM	Millimolar
CBOL	Consortium for the Barcode of Life
COXI	Cytochrome c oxidase
ddH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
ITS	Internal Transcribed Spacer
*ITS*1	Internal Transcribed Spacer 1
*ITS*2	Internal Transcribed Spacer 2
matK	Maturase K
$MgCl_2$	Magnesium chloride
NaCl	Sodium chloride
NGBB	Nezahat Gökyiğit Botanik Bahçesi
PCR	Polymerase Chain Reaction
rbcL	Ribulose-1,5-bisphosphate carboxylase/oxygenase
rpl22	ribosomal protein L22
rpoB	RNA polymerase beta subunit

rpoC1	RNA polymerase beta prime subunit
rps19	Ribosomal protein S19
rRNA	Ribosomal RNA
Tris-HCl	Tris hydrochloride
Tm	Melting temperature
UV	Ultraviolet

1. INTRODUCTION

The increasing industrialization and environmental degradation lead to a planetary biodiversity crisis. The loss of habitat due to climate crisis and contaminants brings many animal and plant species to the edge of extinction, thus the loss of genetic resources. Since almost 80 percent of the total biomass on the earth consists of plants, the loss of germplasm results in a significant diminishment in agriculture that further affects the whole ecosystem. By improving the conservation of germplasm, it is also possible to protect the ecosystem which highly depends on the taxonomy and accurate identification of the species. From the viewpoint of genetic diversity, the location of Turkey has a significant position as a meeting point for Euro-Siberian, Mediterranean, and Irano-Turanian phytogeographical regions. The climate, topography, and geomorphology of Turkey make it to be the home for endemic plants and rich in plant biodiversity [1]. It is estimated that there are approximately 2700 endemic species in Turkey where the *Verbascum* L. genus has an important role with its distinctive biodiversity [2]. The distribution range of the *Verbascum* L. genus is narrow and the increasing risk of the destruction of its habitat creates a requirement for a strict conservation policy considering its therapeutic qualities and ecological relevance [3]. Based on the large number of species in the *Verbascum* L. genus and their varied physical traits, effective species identification using only conventional methods is difficult [4]. Therefore, the use of DNA barcoding techniques becomes essential in clearing up taxonomic uncertainty and making it easier to identify and categorize *Verbascum* L. species. Due to their physical similarity and possibility for hybridization, these endemic *Verbascum* L. species pose difficulties for correct identification and classification. In order to clarify the taxonomic relationships and boundaries and conservation *Verbascum* L. germplasm, it is crucial to use DNA barcoding approaches that cover various gene regions including the Internal Transcribed Spacer (*ITS*), maturase K (*matK*), ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*), and *trnH-psbA* intergenic spacer.

In the field of biodiversity and taxonomy, morphological traits have traditionally been used for plant species identification and classification. As the molecular methods have been developed, a ground-breaking strategy known as DNA barcoding has emerged, providing a potent tool for species identification and delineation [5]. DNA barcodes, short and

standardized genes, are used as molecular markers in DNA barcoding technique to distinguish different species [6]. In animals, the use of cytochrome c oxidase (*COXI*) as a DNA barcode could differentiate most of the species, however, due to high genomic diversity, using one universally standard barcode for plants is not possible. Several barcode regions have been proposed by Consortium for the Barcode of Life Plant Working Group (CBOL). *ITS*, *matK*, *rbcL*, and *trnH-psbA* are examples of such regions that have attracted a lot of interest [7].

In recent years, the use of barcode pairs or triples that target several barcodes, such as *ITS-matK* or *ITS-matK-rbcL*, has demonstrated tremendous potential in improving the effectiveness and accuracy of plant barcoding research. These barcode combinations offer overlapping and complementary information that strengthens the ability to distinguish between different species and improves the resolution of taxonomic problems [8]. The likelihood of receiving confusing or inconclusive results is considerably decreased by simultaneously addressing many barcodes. A more thorough evaluation of genetic variation and evolutionary links among plant species is made possible by the introduction of new barcode regions. Due to its low mutation rate and excellent species-level differentiation rate, the *ITS* region, which is located between the 18S and 26S ribosomal RNA genes, has received much attention as a potential DNA barcode [9]. The *matK* gene, which codes for a maturase K involved in the splicing of group II introns, has also shown promise as a barcode for land plants [9]. The *rbcL* gene, which codes for the large subunit of the photosynthesis-related RuBisCO enzyme, has also been useful in research on plant barcoding and lastly, because of its great variability, the *trnH-psbA* intergenic spacer, which is part of the chloroplast genome, has shown promise in resolving closely related species [9].

The objective of this study is to molecularly identify the key species in the *Verbascum* L. genus by DNA barcoding. For this purpose, the Anatolian endemic species from the *Verbascum* L. genus were selected and herbarium samples were provided by Nezahat Gökyiğit Botanik Bahçesi (NGBB). Two chloroplastic -*rbcL* and *matK*- and two nuclear -*ITS* and *trnH-psbA*- barcode regions were amplified by PCR and visualized by agarose gel electrophoresis. However, only *ITS* and *matK* barcode regions could be identified in 15 and 11 *Verbascum* L. species respectively and used for the phylogenetic analysis since herbarium samples have challenges both in the stages of DNA isolation and amplification of barcode

regions. The samples were sequenced, and the obtained data were analyzed via bioinformatic analysis. Both Next-Generation Sequencing (NGS) and Sanger Sequencing Technologies were used for the sequencing of barcode regions, and this led us to compare the pros and cons of both technologies in the issue of DNA barcoding. By verifying data from both methods, cross-validation between NGS and Sanger sequencing increases the dependability of the results. By combining the two approaches, scientists can take advantage of their individual advantages and investigate genetic variation in the barcode regions more thoroughly. The budget for the entire project may be impacted by the greater expenses associated with combining NGS and Sanger sequencing than with either technique alone. Analyzing and interpreting data can be difficult when managing and combining data from two distinct sequencing technologies.

This study will enable the addition of *Verbascum* L. to Turkey's plant inventory and make molecular-level adjustments to correct taxonomic problems if necessary. In addition, the involvement of the DNA barcodes will further contribute to the conservation of *Verbascum* L. germplasm.

2. LITERATURE REVIEW

2.1. *VERBASCUM* L. GENUS

Turkey's strategic location and a variety of ecological and climatic circumstances have led to an unusually diversified plant population [10]. The country has an impressive variety of flora due to its location at the intersection of three phytogeographical regions (Irano-Turanian, East Mediterranean, and Euro-Siberian). The number of naturally occurring plant taxa in Turkey has been estimated to be around 11700, with 30 percent of them being endemic to the area [11]. Turkey is home to 30 genera and 466 species of the Scrophulariaceae family, which has 280 genera and 3000 species globally.

Scrophulariaceae is represented by 360 species in the world, 63 species in Anatolia, and Lamiale Bromhead, considered as one of the largest families of the genus, is represented in Turkey with 4 endemic species [12]. It has one of the biggest taxon numbers in Turkey [13]. It is known that Scrophulariaceae shows a natural spread pattern in the northern hemisphere. In Scrophulariaceae, a crossbreed rate is commonly seen. Crossbreeds can be found in habitats that consist of 1 or 2 different breeds and especially in corrupted fields [2]. It is reported that most of the parental features were transferred to the newly generated plant. This feature of a natural hybridization of the plant helps the evolution and continuity as well as variation of the plant.

Verbascum L. is a Scrophulariaceae family member, commonly known as mullein. The majority of *Verbascum* L. species are herbaceous biennials or perennials that are distinguished by their tall, upright stems, woolly leaves, and beautiful inflorescences that contain numerous yellow or white color thyrse kind of flowers (Figure 2.1) [2, 14]. This clustered flower shape makes it hard to characterization of the plant. The variation can be listed such as flower number and shape, corolla size and shape, color, and flowering period. While genomic length may range between species, the common mullein, *Verbascum thapsus*, has a genome size of about 2.8 billion pair of bases (2.8 Gb) [15]. *Verbascum* L.s precise antecedents may be found in the Scrophulariaceae family, as can its evolutionary history. There is a wide variation in ploidy levels across the *Verbascum* L. species, with some being diploid (2n) or tetraploid (4n). It is significant to highlight that these genetic

traits might not apply uniformly to all *Verbascum* L. species, and further investigation of particular species is required to gain more exact genetic information [15].

Figure 2.1. *Verbascum phlomoides* (retrieved from Kew Science. *Verbascum* L.) [16].

The flowers contain essential oil and glycosides that help thoracic problems as an expectorant. Some species of *Verbascum* L. have usages as sedatives for all kinds of aches, treatment of skin diseases, dysmenorrhea, and rheumatological problems. Also, the usage of *Verbascum* L. for wound healing in animals was reported [4]. The flowers exhibit mucolytic and expectorant qualities, while the leaves have been used as a diuretic, sudorific, expectorant, and sedative [17]. Using the morphology of the seeds, Murbeck (1933) divided this genus into two sections: Bothrospermae and Aulacospermae, the latter of which had a longitudinally grooved seed (Murb.). Kamelin with alveolate seeds and diagonal grooves [18, 19].

They can colonize and establish themselves in both natural and artificial landscapes thanks to their versatility. Because they can thrive in a variety of soil types and climatic situations, *Verbascum* L. species frequently act as pioneer plants. *Verbascum* L. genus consists of over 450 different species of flowering plants. About 360 species in the genus *Verbascum* L. are found throughout temperate regions of Europe, Africa, and Asia [15]. In Turkey, most of the taxon (198) are endemic and *Verbascum* L. has a high endemism rate of 80 percent [20, 21]. *Verbascum* L. species are represented in Turkey with 255 species (413 taxa) with 107 more crossbreeds. They are considered native plants with the highest diversity rate in the

Mediterranean as well as Europe and Asia [22, 23]. 196 species of *Verbascum* L. are unique to Turkey, which is notable for its high level of endemism [21]. Since the *Verbascum* L. genus is adaptable to a variety of habitats, including sandy areas, rocky terrain, and open or semi-open natural ecosystems, *Verbascum* L. distribution can be seen in all geographical regions in Turkey (Figure 2.2, Figure 2.3, Figure 2.4, and Figure 2.5) [24].

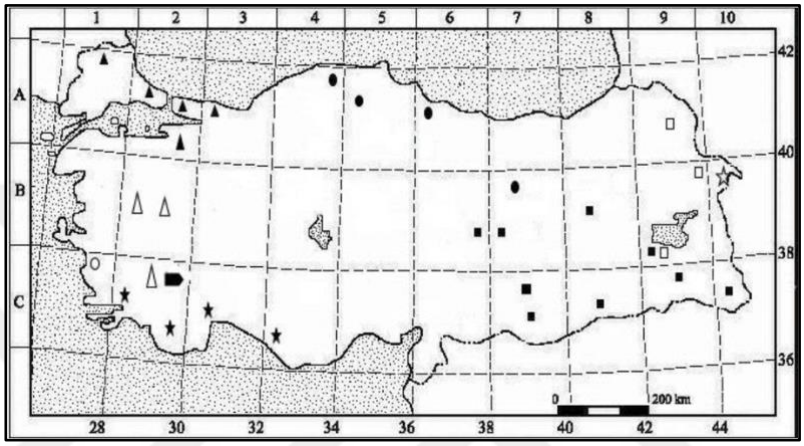


Figure 2.2. Distribution of *Verbascum* L. taxa in Turkey (■) *V. agrimoniifolium subsp. Agrimoniifolium*, (▲)*V. bugulifolium*, (●)*V. ponticum*, (★)*V. levanticum*, (☐)*V. Suworowianum var. suworowianum*, (☆)*V. suworowianum var. papillosum*, (△)*V. luciliae*, (❍)*V. rupicola* (retrieved from Karavelioğulları et al. (2016)) [25].

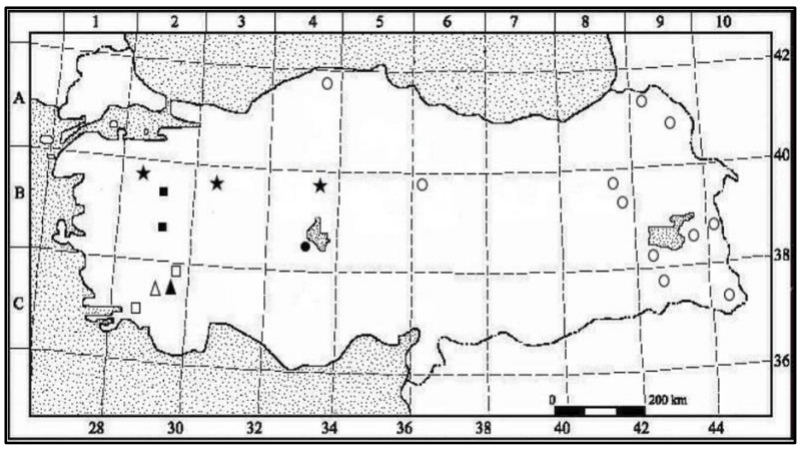


Figure 2.3. Distribution of *Verbascum* L. in Turkey (■) *V. coronopifolium*, (▲) *V. dudleyanum*, (●) *V. pyroliforme*, (★) *V. serratifolium*, (☐) *V. trapifolium*, (△) *V.*

flabellifolium, (❍) *V.oreophilum var. oreophilum* (retrieved from Karavelioğulları et al. (2016)) [25].

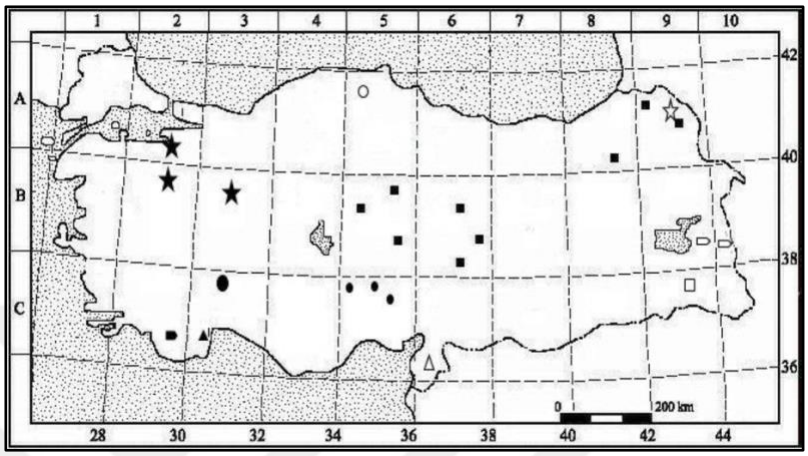


Figure 2.4. Distribution of *Verbascum* L. in Turkey (■) *V. anatolicum*, (▲) *V. spodiotrichum*, (●) *V.cilicium*, (☐) *V. bornmuellerianum*, (△) *V. gaillardotii*, (❍) *V. freynii*, (☆) *V.transcaucasicum*, (★) *V. basivelatum*, (●) *V. sorgerae*, (⚑) *V. bourgeauanum*, (⚐) *V. nudicaule* (retrieved from Karavelioğulları et al. (2016)) [25].

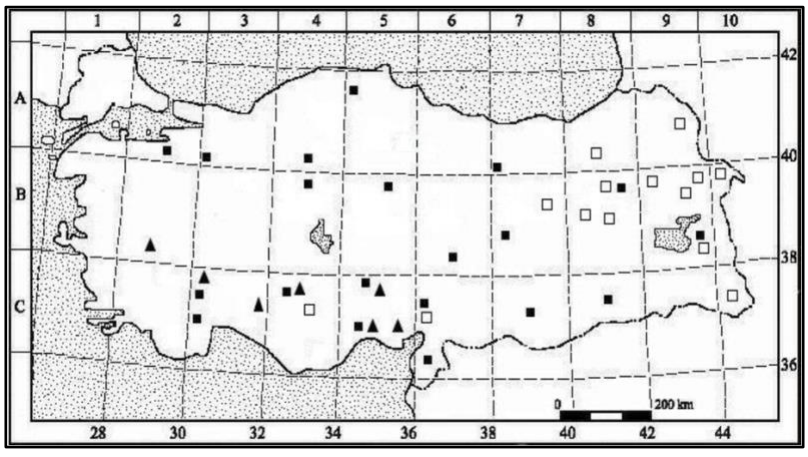


Figure 2.5. Distribution of *Verbascum* L. taxa in Turkey (■) *V. orientale*, (▲) *V. brachysepalum*, (☐) *V. oreophilum var. joannis* (retrieved from Karavelioğulları et al. (2016)) [25].

2.2 DNA BARCODING

DNA barcoding method allows identification of organisms at the molecular level by studying each species its own barcodes and supports classical taxonomy for the classification of species at the genetic level rather than examining and comparing the morphological features only. The main purpose of DNA barcoding is to recognize or distinguish species by identifying their barcode regions combinations or single barcode region [5, 6]. By providing a quick and precise approach for identifying and classifying organisms according to their genetic makeup, DNA barcoding has revolutionized the fields of taxonomy and species identification. This technique produces distinctive DNA profiles or barcodes for many species by using short, standardized gene sections as molecular markers [26]. The development of DNA barcoding has created new opportunities for a wide range of applications in numerous disciplines, including forensic science, ecological research, food verification, and biodiversity protection. Researchers may now explore the immense genetic diversity found in the natural world and decipher complex interactions between species by utilizing the power of DNA sequencing technology, ultimately leading to a deeper understanding of our planets biodiversity [27].

The choice of appropriate barcode areas is one of the most important aspects of DNA barcoding. The DNA of an organism has unique gene sequences called barcode regions that have enough genetic variation to distinguish between different species. Typically, these areas are small, conserved DNA segments that are simple to amplify and sequence [28]. The taxonomic group being studied, the accessibility of reference databases, and the desired level of resolution are only a few examples of the variables that influence the selection of barcode regions. Different taxonomic groups have proposed and frequently used several barcode areas. The animal cytochrome c oxidase subunit I (*COXI*) gene, which has demonstrated excellent discriminating power and universality, is one of the most often used [29]. Due to their various degrees of diversity and amplification success, the nuclear ribosomal internal transcribed spacer (*ITS*) region, chloroplast genes like *rbcL* and *matK*, and intergenic spacers like *trnH-psbA* have been often used as barcode areas in plants [30].

For DNA barcoding to be successfully implemented, it is essential to select the right barcode regions. These areas ought to have enough genetic diversity to distinguish between closely

related species while preserving enough biodiversity to for successful amplification across a variety of taxa [31]. In order to shed light on species evolution, adaptability, and ecological interactions, barcode areas should ideally be connected to certain biological functions or genetic features. The exploration and use of barcode areas have been further improved by the development of novel bioinformatics tools and the ongoing growth of DNA sequencing technologies, allowing researchers to push the frontiers of species identification and categorization [31].

The ability to identify species depends on taxonomists, whose work does not allow them to identify every taxon as demanded by non-specialists. The DNA Barcode of Life initiative intends to address these issues by creating a consistent, quick, and affordable species identification approach that is usable by non-specialists. For lower taxa, such as worms [26], several universal systems for molecular-based identification have been developed. By employing short, standardized gene sections as internal species markers, DNA barcoding is a revolutionary method for quickly, accurately, and automatically identifying different species. As a result, it will improve access to the Linnaean taxonomic system, benefiting ecologists, conservationists, and a variety of organizations tasked with pests, invasive species, and food safety control [27]. More generally, DNA barcoding makes it possible to imagine a time when all inquisitive minds, from scientists to students, will have simple access to the names and biological characteristics of every species on the earth. By enabling taxonomists to quickly sort specimens and by emphasizing divergent taxa that may represent new species, DNA barcoding can quicken the process of species discovery in addition to allocating specimens to known species. DNA barcoding gives taxonomists the chance to considerably increase, and eventually finish, a global inventory of the diversity of life by enhancing their capabilities in various ways [27].

The Barcode of Life project was aimed to provide a global system for an inventory of eukaryotic species based on a conventional molecular approach. The global project Consortium for the Barcode of Life (CBOL) proposed it in 2004 [32].

The emergence of DNA barcoding has improved the classification and identification of biodiversity [33]. DNA barcoding can also be used for a variety of other purposes, including confirming identity or purity to support food safety and authenticity of labeling [34, 35], to

reveal cryptic species [36], to link biological samples to crime scenes in forensics [37, 38] and in ecological and environmental genomics [39].

Because plants have far slower mutation rates than animals do, the cytochrome c oxidase 1 (*COXI*) sequence, which has been recognized as a universal barcode in animals, is unable to distinguish most plant species [40]. Our ability to precisely identify and categorize species has been transformed by DNA barcoding and the choice of appropriate barcode regions. There are a number of factors taken into account while choosing a barcode region for DNA barcoding [28]. First and foremost, universality is crucial since it ensures that the chosen barcode area may be applied widely across a variety of organisms because it must be present in most species. The second important factor is variability, as the chosen location must have enough genetic variation among species to provide precise species identification. A further important consideration is amplifiability, which states that the barcode region must be simple to amplify and sequence using common laboratory methods to guarantee accurate and timely analysis. Last but not least, functioning is crucial, with the barcode region ideally offering helpful taxonomic information to aid in both the identification and classification of species based on the acquired genetic data [28]. In DNA barcoding applications, a barcode region can efficiently fulfill its function by satisfying these requirements. With a differentiation efficiency of only 72 percent, Consortium barcode of life recently proposed the two-locus combination of *matK* + *rbcL* as the best plant barcode region combination [41]. It can be difficult to differentiate between closely related plant species or subspecies merely based on these barcode regions, though, because they occasionally have similar sequences in these regions. If this is the case, researchers may investigate additional barcode regions, such as *ITS* or other chloroplast markers, such as *trnH-psbA*, to get higher resolution and improve the ability to distinguish between closely related orchid species and subspecies. Researchers can improve the discriminatory strength of DNA barcoding in plants by using numerous barcode areas, allowing for more precise identification and categorization at both the species and subspecies levels.

With the help of this potent genetic tool, the complexities of biodiversity and its contribution to forensic analyses, conservation efforts, and quality control in sectors like agriculture and seafood can now be understood. DNA barcoding holds enormous potential for unraveling the mysteries of the natural world and expanding our understanding of the intricate web of

life thanks to ongoing technological breakthroughs and the continued research of various genetic markers [42].

Before GenBank recognizes a proposal to attach the keyword barcode to a sequence area, CBOL must approve it. Only *COXI* has such a classification so far, although it is obvious that plants require several loci and consequently a different strategy. For all land plants, the two or three-region combination options offer the best possible possibilities for the use of plastid DNA as barcodes, provided that these remaining technological concerns can be resolved.

2.2.1 DNA Barcoding in Animals

Using short, standardized gene sections to identify and categorize different animal species is known as animal DNA barcoding. The method is based on the examination of a particular gene, the cytochrome c oxidase subunit I (*COXI*) gene, which is frequently used as the animal kingdoms standardized barcode. This gene region displays enough genetic variation among many animal species to provide precise species identification. Researchers may rapidly and accurately identify the species of an animal sample by comparing the *COXI* sequences of unknown specimens to an extensive reference library [29].

The use of animal DNA barcoding has substantial effects on a number of scientific and practical areas. When dealing with specimens that are challenging to identify only based on physical traits, DNA barcoding offers an efficient and accurate method for species identification [27]. It has been shown to be especially helpful in situations when traditional taxonomy is difficult because there are cryptic species or morphologically similar taxa. Numerous animal species, including mammals, birds, fish, insects, and invertebrates, have shown success with the method. The potential for the discovery of new species and the investigation of ecological and evolutionary patterns increases as researchers continue to compile DNA barcode data and improve the reference databases [39]. Animal DNA barcoding is a useful tool for both scientific study and real-world applications, aiding in the management and conservation of animal species as well as the comprehension of global biodiversity trends.

Beyond identifying species, animal DNA barcoding has numerous uses and benefits. Its contribution to ecological and conservation studies is one important advantage. Researchers can learn more about patterns in biodiversity, population dynamics, and ecological interactions by precisely identifying and cataloging the animal species that exist within a particular habitat [43]. Additionally, DNA barcoding helps to identify invasive species, enabling early and precise management approaches to lessen their negative effects on local ecosystems [39]. Animal DNA barcoding also helps to track changes in species distributions and populations over time, which is essential for monitoring and evaluating the success of conservation activities. Additionally, DNA barcoding has been helpful in wildlife forensics, enabling the identification of endangered species in confiscated products, the detection of illegal wildlife trade, and the monitoring of animal products in the food industry, providing a method for precise species authentication. In the field of public health, DNA barcoding makes it possible to identify disease carriers, such as ticks and mosquitoes, which aids in monitoring and containing the spread of vector-borne illnesses [36]. Additionally, in the seafood business, where mislabeling of species is a regular concern, DNA barcoding has been demonstrated to be useful in assuring food safety.

2.2.2 DNA Barcoding in Plants

There is currently no standardized procedure for DNA barcoding of land plants. Due to *COXI*s slow rate of evolution and relatively low levels of variability in the mitochondrial DNA of land plants, this barcode is ineffective for identifying plants, but the *matK* and *rbcL* gene areas in the chloroplast have been accepted as the plant kingdoms respective barcodes [9, 44]. Using a combination of plastid regions is one approach to solving this issue. A project that compared the performance of a range of plastid regions to find suitable DNA barcoding regions that could be used as “the universal land plant barcoding protocol” was funded [45].

In order to keep the plant barcoding community abreast of ongoing research towards finding a standard DNA barcoding protocol an option was proposed for standardizing a plant DNA barcoding protocol; is to use combinations of plastid genes [7, 46]. It is generally known that plant mitochondrial DNA evolves slowly, thus when zoologists chose the *COXI* gene as the

standard barcode for animals, it was clear that a different approach would be needed for plants [43]. The plastid genome is the closest source of an equivalent plant barcoding region because it has many of the same favorable characteristics as animal mitochondrial DNA for barcoding, including conserved gene order and a high copy number in each cell that makes it simple to retrieve DNA for PCR and sequencing. One problem with plastid DNA, however, is its generally slow rate of evolution. An appropriate region would have enough variation within it to distinguish between different species. While it is challenging to locate sites for reasonably conserved primers in the most variable regions, markers with highly conserved primer binding sites tend to be internally more conservative [47].

Having a reading frame is another important characteristic for a prospective barcoding area since it allows researchers to assess how well sequencing reactions and editing have been done by looking for nonsense substitutions. In addition, it is not necessary for the marker to be a coding region, but it is a preferred feature because it makes the selected markers also useful for studies of phylogenetic relationships and molecular evolution. The protein-encoding plastid gene *rbcL* has been proposed as a potential plant barcode. One benefit of it is a large amount of existing information, there are more than 10,000 *rbcL* sequences already in GenBank. However, none of these have electropherogram trace files accessible, many of these are unvouchered or incorrectly recognized [47, 48].

2.2.2.1 Coding Barcode Regions

Taxonomists have hypothesized that differentiating between plant species may need the use of a multi-locus technique [49, 50, 51]. However, CBOL showed that the addition of many loci did not significantly enhance these approaches' capacity for species-level differentiation [41]. Recently, researchers suggested using the whole-plastid genome sequence to identify plants [35, 52, 53]. This idea is still not widely acknowledged. When compared to the application of single-locus barcodes, one of the key challenges is the high sequencing costs associated with collecting entire plastid genome sequences.

Gymnosperms and cryptograms had less success. While *matK* and *trnH-psbA* offer excellent resolution but need further development, *rbcL* offers high universality and good but not outstanding discriminating power. Consequently, no single locus satisfies CBOL's data standards for locus selection, necessitating the use of a synergistic combination of loci. To

enable further testing of these loci, a three-locus barcode consisting of *matK+rbcL+trnH-psbA* was preferred. However, to avoid the higher cost and prevent future delays in creating a standard barcode, the majority preferred to choose a 2-locus barcode [44]. Because *rbcL+matK* was the most popular of the two-locus barcode combinations, it was suggested as the default barcode for land plants. An intricate trade-off between universality, sequence quality, discrimination, and cost has been pragmatically resolved by this combination [45].

It is important to note that research concentrating on the plant biodiversity of a particular region or local area will have stronger discriminatory power of this standard barcode in cases involving geographically constrained sample sets. In the future, increasing the percentage of cases in which unique species identifications are archived will be a challenge for DNA barcoding in plants [54, 40].

In taxonomic groupings where direct sequencing of this locus is available, the rapidly developing internal transcribed spacers of nuclear ribosomal DNA also serve as a helpful additional barcode. Reaching consensus on a common set of loci is a crucial first step in the building of a global infrastructure for plant barcoding as well as large-scale sequencing [55, 8]. Two-locus combinations gave 59-75 percent resolution and 3-locus combinations 65-76 percent. Three locus combinations yielded 65-76 percent resolution, while two-locus combinations produced 59–75 percent. 73 percent of species could be distinguished using different loci [8]. A DNA barcode should offer the greatest amount of species differentiation, be routinely retrievable with a single primer pair, and be suitable for bidirectional sequencing with minimal manual editing of sequence traces. Four of the possible loci, while having poor discriminatory power, can be disqualified based on these criteria.

2.2.2.1.1. *matK* (maturase K)

An important molecular marker frequently used in plant DNA barcoding investigations is the *matK* barcode region. The maturase K protein, which is involved in the splicing of group II introns, is encoded by the *matK* gene, which is found within the chloroplast genome. The balance of conserved and changeable areas is one of the *matK* barcode regions many advantageous traits [9]. The conserved areas make it easier to construct primers, allowing the creation of generic primers that can amplify *matK* in a variety of plant species. While this is happening, the variable sections of *matK* offer the required sequence variation for

identifying and differentiating across species. *matK* is a useful marker for resolving phylogenetic connections and differentiating between closely related plant taxa because of its combination of conserved and variable regions [9].

matK, demonstrated significantly larger amounts of sequence variation and improved species differentiation, work is still being done to strengthen its universality by developing stronger PCR primer sets. *matK* has a low transition/transversion rate, an appropriate length, a high evolutionary rate, and clear interspecific divergence [56, 57]. Unfortunately, utilizing the primer sets that are currently on the market, *matK* is challenging to universally amplify. It was also given the options *matK* and *trnH-psbA*. The advantages of doing this include the potential for obtaining more species-level resolution. The drawback of this strategy is that because the *trnH-psbA* spacer varies and is frequently bigger in size, it poses bioinformatics concerns and issues with degraded tissue. There is not a single, globally changeable plastid DNA region that can fulfill this need, thus producing three sections will be more expensive than producing one. Additionally, CBOL has approved the multi-locus barcode concept in general [41].

There are many benefits to using the *matK* barcode region in plant DNA barcoding. First off, *matK* has gained widespread acceptance due to its ability to be successfully amplified and sequenced across a variety of plant groupings. The correct identification and classification of plant species is facilitated by the availability of trustworthy sequence data from *matK* [9]. Additionally, *matK* shows significant sequence divergence, especially in the third codon position, which improves its ability to distinguish between various taxonomic levels. The *matK* barcode region also contributes to our understanding of the evolutionary relationships across plant taxa by offering useful insights into the phylogenetics and evolution of plants. Overall, incorporating the *matK* barcode region into studies of plant DNA barcoding increases species identification, backs phylogenetic analyses, and advances the science of plant biology [9].

2.2.2.1.2. ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*)

rbcL is a good DNA barcoding area for plants at the family and genus levels because it is straightforward to amplify, sequence, and align in most land plants [58]. But *rbcL* sequences change slowly, and this locus in flowering plants has by far the lowest divergence of plastid

genes [40]. Despite these drawbacks, *rbcL* was nonetheless recommended as one of the best candidate plant barcodes based on the simplicity with which the gene sequence could be recovered, the volume of readily available data, and the good but not exceptional discriminatory power [58, 59]. The gene *rbcL* has the best characterization among plastid areas. It is now simple to retrieve across land plants thanks to improvements in primer design, and it is ideal for recovering high-quality bidirectional sequences [59].

Other popular plastid barcoding markers exist in addition to the potential barcodes mentioned above, including *rpoB, rpoC 1, atpF-atpH* and *ycf5*. These chloroplast areas are useful for phylogenetic studies and barcoding research at higher taxonomic levels, but the lack of diversity makes them unsuitable for plant DNA barcoding at lower taxonomic levels. Despite significant effort, finding a universal plant barcode similar to *COXI* in mammals has been challenging since there is insufficient variety within single loci [40, 60].

2.2.2.2 Non-coding Barcoding Regions

In addition to coding areas, noncoding barcode regions have proven to be useful tools for plant DNA barcoding. The *trnH-psbA* intergenic spacer, which is found within the chloroplast genome, is a notable non-coding region. Since there is a lot of genetic variation in this area, it can be used as a molecular marker to distinguish between different species. Due to its capacity to distinguish between closely related plant taxa and reveal information about their evolutionary relationships, the *trnH-psbA* intergenic spacer has been widely used in plant DNA barcoding investigations. Researchers may precisely identify and categorize plant species by examining the differences within this noncoding region, which advances our knowledge of plant biodiversity and evolutionary trends [9].

The Internal Transcribed Spacer (*ITS*) region is a key non-coding area used in plant DNA barcoding. The nuclear *ITS* region is made up of the *ITS1* and *ITS2* regions, which are divided by the 5.8S rRNA gene. The *ITS* region is a useful molecular marker for identifying species because of its great diversity and low mutation rates. Due to its capacity to distinguish between closely related species, the *ITS* region has been widely used in plant DNA barcoding investigations. Researchers can discover the genetic diversity and evolutionary links among plant species by examining the genetic variants within the *ITS* region [44].

Plant DNA barcoding is made more efficient and precise by including non-coding barcode sequences like *trnH-psbA* and *ITS* when combined with coding barcode regions [44]. These non-coding areas shed important light on the phylogenetic relationships, genetic diversity, and evolutionary history of plant groups. Researchers can increase the accuracy of species identification, add to our knowledge of plant evolution and biodiversity, and support conservation efforts by helping to identify and safeguard plant species by using non-coding barcode regions in plant DNA barcoding studies [30].

2.2.2.2.1. Internal Transcribed Spacer (*ITS*)

ITS has a long record of use and in most groups of flowering plants, it has performed well as a phylogenetic marker. It has almost always produced outcomes like those of plastid DNA, but it frequently has three to four times as many variable sites and evolves up to four times more quickly [40, 7]. *ITS* is also subject in most organisms to gene conversion/concerted evolution so that a single copy type is maintained. However, certain land plants preserve many copies, and even in some angiosperm families, multiple divergent yet still functioning copies of *ITS* are frequently found. As a standard barcoding area across all land plants, *ITS* is unsatisfactory due to the existence of several, divergent copies [61].

According to Alvarez and Wendel (2003), the *ITS* spacer is a potent phylogenetic marker at the species level that exhibits significant levels of interspecific divergence. In addition to being suggested as a plant barcode, the greater discriminatory power of *ITS* over plastid regions at low taxonomic levels have also been extensively studied [40, 62].

This is particularly true for parasitic plants, which provide less resolution from plastid barcodes [45]. However, *ITS* has only been viewed by CBOL as an additional locus [41]. Its shortcomings include inadequately coordinated evolution, fungus contamination, and challenges with amplification and sequencing [45]. Fungi have *ITS* sequences that can occasionally be selectively amplified and mistaken for plant sequences.

With moderate mutation rates and large sequence differences between plant species, the *ITS* region is highly changeable. The *ITS* region has distinguished itself as a valuable molecular marker for species identification and classification due to its great variability. It has been widely used in research on plant DNA barcoding, enabling scientists to precisely distinguish between closely related species that may be challenging to differentiate based solely on

physical traits [30]. Understanding the genetic diversity, evolutionary relationships, and phylogenetic trends within plant taxa is possible through the analysis of *ITS* sequences.

2.2.2.2.2. *trnH-psbA*

One of the plastid DNA regions is the non-coding *trnH-psbA* spacer [51, 63]. This region is one of the non-coding regions of the plastid genome in angiosperms with the highest level of variation. This means this intergenic spacer can offer high levels of species differentiation. The high rates of insertions and deletions at this location, however, present significant alignment challenges. It seems that there are significant length variances even among closely related taxa [51, 63]. Also, the *trnH-psbA* spacer is extremely short in some plant groups but much longer in others due to the presence of duplicates of *rpl22* and *rps19*. The *rps19* gene (pseudogene) is also found between *trnH* and *psbA* in maize, rice, and wheat [61]. Because of its long length, it is difficult to retrieve *rps19* from degraded tissue, such as highly processed material, or from forensic evidence, which makes its usage as the sole standard plant DNA barcode problematic.

Although non-coding areas like *trnH-psbA* typically evolve more quickly than genes, this is not always the case. In some populations of mosses, *matK* and *rpoCI* both have higher levels of positional variability.

Designing universal primers is possible since both sides have substantially conserved coding sequences [63]. The non-coding intergenic region contains the highest rates of insertion and deletion and the largest sequence divergence. These characteristics make *trnH-psbA* an excellent candidate for use as a plant barcode for species differentiation [63, 61]. Extensive barcoding studies have shown that the *trnH-psbA* region can identify nearly all species in some land plant groups, including *Hydrocotyle, Dendrobium*, and *Pteridophytes* [64].

2.3 DNA SEQUENCING TECHNIQUES

2.3.1 Sanger Sequencing

In DNA barcoding, Sanger sequencing is used for species identification and classification, playing a crucial role in fields such as food authentication, forensic investigations, and

biodiversity studies. Researchers can utilize Sanger sequencing to identify and classify different species, even in complicated biological materials, by focusing on specific DNA regions, such as the mitochondrial cytochrome c oxidase subunit 1 (*COXI*) gene. Fast and precise species identification is made possible by this method, which greatly advances genetic research and a number of biological and ecological applications [65].

Sanger sequencing is a DNA sequencing technique that uses chain-terminating dideoxynucleotides (ddNTPs) to replicate DNA. The procedure creates a series of labeled DNA fragments that can be separated by capillary gel electrophoresis to ascertain the original DNA sequence by introducing fluorescently labeled ddNTPs into the developing DNA strand [66].

Numerous research applications, such as DNA analysis, genetic disease diagnosis, and comprehending genetic variants, heavily rely on Sanger sequencing. It is an essential tool in the field of genomics since it has helped to sequence small to medium-sized DNA fragments. Furthermore, Sanger sequencing has been used in many large-scale genome projects, helping to sequence the whole genomes of a variety of organisms, including bacteria, plants, and humans [67].

2.3.2 Next-Generation Sequencing (NGS)

Next-Generation Sequencing (NGS) is a high-throughput DNA sequencing technology that allows for the rapid and cost-effective analysis of large amounts of DNA or RNA sequences.

It is also known as massively parallel sequencing, as it can generate millions of sequences reads simultaneously. NGS has revolutionized the field of genomics and has a wide range of applications, including genome sequencing, transcriptome analysis, epigenetics, and metagenomics [68]. The basic NGS process starts with DNA/RNA fragmentation into multiple fragments. This step is followed by the ligation of adapters that contain sequences that are required for the amplification to the ends of the fragments. The complex of fragments that are ligated with adapters are then amplified by using PCR and create large number of identical copies of each fragment. The amplified fragments are then loaded onto a sequencing instrument, where they are immobilized and amplified again to form clusters

[69]. Researchers can utilize amplicon sequencing, which entails tagging a distinct barcode onto PCR amplicons that are then combined with additional tagged amplicons and sequenced, to apply NGS for DNA barcoding. With this method, allelic variation in the barcode region can be captured, resulting in a more thorough and precise species identification [70].

2.4 AIM OF THE STUDY

Species-specific identification is an immensely important process in almost every field, but it is almost essential in biology. DNA barcoding method allows us to identify and discover plants at the molecular level by giving each species its own barcodes and supports classical taxonomy for the classification of species at the genetic level rather than examining and comparing the morphological features only.

Starting with the *Verbascum* L. barcoding research is crucial because there isn't much research in the literature that specifically address this goal. "Taxonomy, comparative genomics of Mullein (*Verbascum, Scrophulariaceae*), with implications for the evolution of *Verbascum* L. and *Lamiales*," a study by Dong et al. (2022) published in BMC Genomics, offers insights into Mullein's taxonomy and comparative genomics, illuminating *Verbascum* L.s evolutionary history and relationship to *Lamiales*. Using data from both nuclear and plastid genomic perspectives,

Ghahremaninejad et al. (2014) investigated the "Monophyly of *Verbascum* (*Scrophularieae: Scrophulariaceae*), with evidence from nuclear and plastid phylogenetic analyses" and published their findings in the Australian Journal of Botany. The authors study provides important new insights into the phylogenetic relationships within the *Verbascum* genus.

"Nomenclature and typification in *Verbascum* (*Scrophulariaceae*) from North Africa," a paper by Khamar et al. (2023) published in PhytoKeys, addresses nomenclature and typification issues related to the *Verbascum* genus, with a focus on specimens that come from North Africa.

Using molecular techniques, Srivastava and Saggoo (2014) conducted a study titled "Molecular identification of *Verbascum thapsus* L. (Ban Tambaaku) and its ITS sequence

comparison with other *Verbascum* L. species," which was published in Medicinal Plants - International Journal of Phytomedicines and Related Industries. The authors compared the Internal Transcribed Spacer (ITS) region of *Verbascum thapsus* L. strains using this information.

Olanj, Nouri, Heidari, and Shariat (2021) carried out a taxonomic study titled "Taxonomic study of seven species of *Verbascum* L. (*Scrophulariaceae*) based on molecular and morphological data in Iran," which was published in the Journal of Plant Research (Iranian Journal of Biology). The authors used both molecular and morphological data to examine and categorize seven *Verbascum* L. species from the Iranian region.

The objective of this study is the molecular identification of key species in the *Verbascum* L. genus by DNA barcoding. For this purpose, the Anatolian endemic species from the *Verbascum* L. genus were selected and identified by DNA barcoding method. This will enable the addition of *Verbascum* L. to Turkey’s plant inventory and make molecular level adjustments to correct taxonomic problems if necessary. In addition, the involvement of the DNA barcodes will further contribute to the conservation of *Verbascum* L. germplasm.

3. MATERIALS

3.1 CHEMICALS

10X PCR Buffer (Thermo Fisher Scientific, USA), Agarose (Sigma, USA), dNTP (Thermo Fisher Scientific, USA), Ethidium Bromide (Fisher Scientific, Germany), GeneRuler 100bp DNA Ladder Mix (Thermo Scientific, USA), Taq Polymerase (Thermo Fisher Scientific, USA), Loading dye (Thermo Scientific, USA), $MgCl_2$ (Thermo Scientific, USA), Ethylenediamine tetraacetic acid (Merck, Germany), Tris-HCl (Sigma, USA), Glacial Acetic Acid (Merck, Germany)

3.2 CONSUMABLES AND GLASSWARE

Eppendorf tubes 1.5 mL, 2 mL (Isolab, Germany), Erlenmeyer's flasks 250 mL (Isolab, Germany), Glass borosilicate bottles 1000 mL (Isolab, Germany), Gloves, Graduated cylinder 100 mL, Micropipette tips 10 µL, 200 µL, 1000 µL, Micropipettes 0.1-10 µL, 10-100 µL, 100-1000 µL (Thermo Fisher Scientific, USA), Mortar and pestles (Isolab, Germany),Liquid Nitrogen Tank (Isotherm, Germany), Falcon 15 ml Falcon 50 ml (Isolab, Germany).

3.3 EQUIPMENTS

Analytical scale (Shimadzu, TW423L), Autoclave (Wisd, MeXterile 60), Vortex Mixer (WiseMix, VM-10), Water bath (SUB Aqua Plus), +4°C Refrigerator (Arçelik, Turkey), -20°C Freezer (Arçelik, Turkey), Agarose Gel Electrophoresis System (BioRad, USA), Centrifuge (Eppendorf, Germany), Thermocycler (Bio-Rad, USA), UV Transilluminator (Vilber Lourmat, Germany).

3.4 PLANT DNA ISOLATION KIT

Macherey – Nagel NucleoSpin Plant II isolation kit

3.5 PLANT MATERIALS

All the *Verbascum L.* species specimen were kindly provided from the herbarium collection of Nezahat Gökyiğit Botanik Bahçesi (NGBB). Fresh tissues of few *Verbascum L.* species were also provided by NGBB.

4. METHODS

4.1 COLLECTION, DOCUMENTATION, AND IDENTIFICATION OF SPECIMEN

Both fresh and herbarium *Verbascum* L. specimen were provided by Nezahat Gökyiğit Botanik Bahçesi (NGBB) and experimental procedures were carried out at Yeditepe University Plant Tissue Culture and Genetics Laboratory. Herbarium samples were provided with their NGBB barcode numbers and the information about where, when and by whom the sample was collected and registered. The identification of the samples was done by professionals when they arrive at NGBB and even beforehand when they collected from the original location. All the *Verbascum* L. samples that were used in the current study were collected in Turkey and selected carefully to produce explanatory results between the important species of *Verbascum* L. genus by DNA barcoding. All *Verbascum* L. samples used in this study are listed in Table 4.1.

Table 4.1. List of all *Verbascum* L. samples

Herbarium Samples			
Notated number	**NGBB Barcode**	**Botanical Name**	**Location**
1	5660	*Verbascum sinatum ssp sinatum*	Istanbul, Ataşehir Nezahat Gökyiğit Botanical Garden
2	9471	*Verbascum sinatum ssp sinatum*	-
3	9817	*Verbascum speciosum Schrad*	-
4	5628	*Verbascum tuna-ekimii Karavel*	Erzincan, Kemah Refahiye
5	11352	*Verbascum vulcanicum Boiss&Hedr var vulcanicum Boiss&Hedr*	Ankara, Beypazarı, Hırkatepe

6	5281	*Verbascum glomeratum Boiss*	Gaziantep, Dülükbaba Forest
7	8856	*Verbascum glomeratum Boiss*	Muğla, Datça
8	5277	*Verbascum isauricum Boiss&Heldr*	Karaman, Ermenek Kazancı Village
9	8859	*Verbascum luridiflorum Hub-Mor*	Kayseri, Yahyalı
10	9812	*Verbascum ovalifolium ssp thracicum (Velen) Murb*	Tekirdağ, Nişantepe
11	9816	*Verbascum ovalifolium ssp thracicum Donn ex Sims*	Tekirdağ, Ahievren, Çavuşköy
12	8009	*Verbascum pestalozzoe Boiss*	Antalya, Uçarsu
13	5887	*Verbascum hasbenlii*	Çanakkale, Bodurlar Village
14	5327	*Verbascum alyssifolium*	Erzincan, Kemah
15	7794	*Verbascum blattaira*	Kyrgyzstan Bishkek Manas University Campus
16	MÖZT6011	*Verbascum cilicicum*	-
17	4801	*Verbascum faik-karaveliogullari Çıngay&Cabi*	Hakkari, Berçelan
18	9813	*Verbascum speciosum Schrad*	Tekirdağ, Yeniçiftlik
19	9723	*Verbascum xanthophoeniceum Griseb*	Tekirdağ, Ganos Dağı
20	10015	*Verbascum phoneicum*	Kayseri, Saz Deresi
21	10007	*Verbascum phoneicum*	Kayseri, Saz Deresi
22	5334	*Verbascum gypsicola Vural&Aydogdu*	Ankara, Beypazarı
23	2002	*Verbascum alyssifolium*	Erzincan, Kemah
Fresh Samples			

Notated Number	**NGBB Barcode**	**Botanical Name**	**Location**
24	F.Vul	*Verbascum vulcanicum*	Istanbul, Ataşehir Nezahat Gökyiğit Botanical Garden
25	8875	*Verbascum hasbenlii*	-

The collected herbarium samples were stored at room temperature (Figure 4.1) whereas fresh tissue samples were preserved in silica gel (Figure 4.2) which is an effective and cost-efficient way to keep samples dry and prevent molding, at room temperature until further use. The total number of collected samples was 25 with 23 herbarium and 2 fresh tissue samples (Table 4.1).

Figure 4.1. Geographical Distribution of Collected Samples in Turkey via Google Maps.

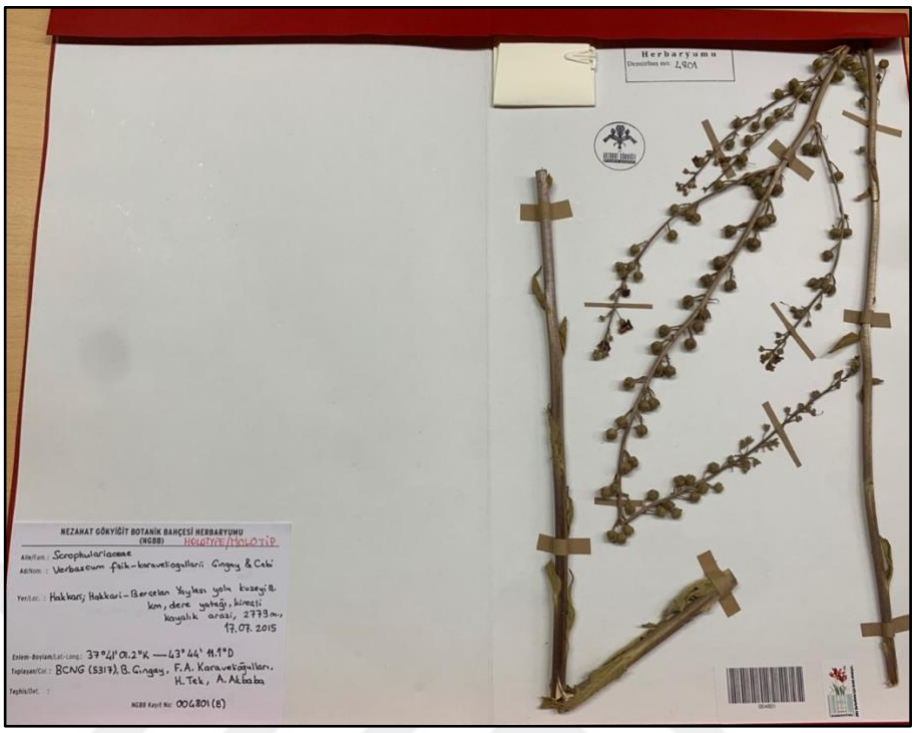


Figure 4.2. Herbarium sample preservation envelopes from Nezahat Gökyiğit Botanik Bahçesi (NGBB).

Figure 4.3. Fresh sample preservation in silica from Nezahat Gökyiğit Botanik Bahçesi (NGBB).

4.2 DNA ISOLATION

4.2.1 CTAB Isolation

The herbarium samples, weighing 250 mg each, were individually homogenized using liquid nitrogen. Subsequently, 900 µl of CTAB and 9 µl of β-Mercaptoethanol were added to each sample and vortexed. The solutions were then incubated at 65°C for 60 minutes. Following the incubation period, a chloroform:isomylalcohol (1:1) solution was added to each sample in a 24:1 ratio and vortexed. The resulting mixtures underwent centrifugation at 15000xg for 10 minutes. After centrifugation, supernatants were discarded, and a 1:1 chloroform:isomylalcohol solution was added before another round of centrifugation at 1500xg for 10 minutes. Following this, 20 µl of RNase was added, and the mixtures were incubated at 37°C for 60 minutes. After the RNase incubation, 2/3 of ice-cold isopropanol was added to each sample, followed by centrifugation under the same conditions. The supernatants were discarded, and the pellets were washed with a mixture of 600 µl 70 percent ethanol and 400 µl 0.2M sodium acetate. The resulting mixtures were then centrifuged at 1000xg for 5 minutes, after which the supernatants were discarded, and 100 µl of ddH_2O was added. After that, samples were incubated at 37°C for 60 minutes and then at 60°C for 45 minutes. Following the incubation, the sample concentration and purity were measured using a nanodrop, and the samples were stored at -20°C.

4.2.2 Commercial Plant Isolation Kit

Macherey – Nagel NucleoSpin Plant II isolation kit was used for DNA isolation. Before the kit was used, both herbarium and fresh samples were grounded using liquid nitrogen into powder. All steps were done based on the manufacturers' instructions. The powdered samples were collected in a tube and proceed with cell lysis using 400 µL Buffer PL1, followed by the addition of 10 µL RNase A, and mixing and the incubation for 10 minutes at 65°C. After the incubation, for filtration and clarification of crude lysate, NucleoSpin® Filter was placed into a new collection tube and lysate loaded onto the column. The sample was centrifuged for 2 min at 11.000 x g, and the flow-through was collected. To adjust the

DNA binding conditions, 450 µL Buffer PC was added on the flow-through and mixed thoroughly by pipetting. For DNA binding, a NucleoSpin® Plant II Column was placed into a new Collection Tube (2 mL) and 700 µL of the sample was loaded. The tube was centrifuged for 1 min at 11.000 x g and the flow-through was discarded. After that, wash step was carried out to remove impurities and dry silica membrane with the initial wash. In the first wash, 400 µL of Buffer PW1 to was added on the NucleoSpin® Plant II Column. The column was centrifuged for 1 min at 11.000 x g and the flow-through was discarded. For the second wash, 700 µL of Buffer PW2 to the NucleoSpin® Plant II Column. The column was centrifuged for 1 min at 11.000 x g and the flow-through was discarded. Last washing step was proceeded with adding another 200 µL Buffer PW2 to the NucleoSpin® Plant II column and centrifuged for 2 min at 11.000 x g to remove wash buffer and dry the silica membrane completely. Finally, to elute the genomic DNA, the NucleoSpin® Plant II Column into was placed onto a new 1.5 mL microcentrifuge tube. 50 µL of Buffer PE at 65°C was added onto the membrane and incubated for 5 min at 65°C. The sample was centrifuged for 1 min at 11.000 x g to elute the DNA. Same process was repeated, and the DNA was eluted on the same tube.

The concentration and purities of the isolated DNA samples were determined by spectrophotometric analysis (Biospec-nano Spectrophotometer, Shimadzu Biotech) and then stored at -20ºC until further use.

4.3 PCR AMPLIFICATION AND GEL ELECTROPHORESIS

ITS, *matK*, *rbcL*, and *trnH-psbA* barcode regions were amplified by polymerase chain reaction (PCR) containing 2.5 µl of 10X Taq DNA Polymerase PCR Buffer, 2.5 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix, 1 µl of 10 µM forward and reverse primer each, 0.25 µl of 5 U/µl DNA Taq Polymerase, 2 µl of DNA from *Verbascum* L. sample and adjusted to final volume 25 µl with ddH_2O.

The PCR conditions for *ITS* and *matK* gene primers are as follows: 94ºC for 1 min, 40 cycles of 94ºC for 40 secs, 55ºC for 40 secs, 72ºC for 40 secs followed by 72ºC for 5 min (Table

4.2) [71, 72]. The PCR conditions for *rbcL* gene primers are 94ºC for 3 mins, 35 cycles of 94ºC 40 secs, 58ºC for 1 min, 72ºC for 1 min followed by 72ºC for 7 min(Table 4.3) [73]. The PCR conditions for *trnH-psbA* gene primers are as follows: 94ºC for 1 min, 40 cycles of 94ºC for 30 secs, 55ºC for 40 secs, 72ºC for 40 secs followed by 72ºC for 5 min (Table 4.4) [74].

For visualization of PCR products, 1.5 percent agarose gel containing 1.5 gr of Agarose, 100 ml of 1X TAE buffer and 3 µl of EtBr, was run at 120 V until the bands of DNA Ladder Mix were separated enough. PCR products were screened by using UV Transilluminator. All primer sequences used in this study are listed in Table 4.5.

Table 4.2. PCR cycles used for amplification of *ITS* and *matK* genes.

94ºC	1 min	
94ºC	30 sec	40 cycles
55ºC	40 sec	
72ºC	40 sec	
72ºC	5 min	

Table 4.3. PCR cycles used for amplification of *rbcL* gene.

94ºC	3 min	
94ºC	40 sec	35 cycles
58ºC	1 min	
72ºC	1 min	
72ºC	7 min	

Table 4.4. PCR cycles used for amplification of *trnH-psbA* gene.

94ºC	1 min	
94ºC	30 sec	40 cycles

54.2ºC	40 sec	
72ºC	45 sec	
72ºC	5 min	

Primers were not designed and ordered instead for the following experiments of barcoding *Verbascum* L. based on their universality for plant DNA barcoding [75]. Ordered sequences and their references are shown at the Table 4.5.

Table 4.5. Gene names and corresponding primer sequences used in the study.

Gene	**Primer Sequence (5' to 3')**	**Reference**
ITS	**F:** GTCCACTGAACCTTATCATTTAGAGG **R:** GCCGTTACTAAGGGAATCCTGTTAG	Beyra Matos, A. et al. (1999) [76] Käss, E. et al. (1997) [77]
matK	**F:** ACCCAGTCCATCTGGAAATCTGGTTC **R:** CGTACAGTACTTTTGTGTTTACGAG	Chase MW et al. (1993) [78] Chase MW et al. (1993) [78]
rbcL	**F:** TGTCACCACAAACAGAAACT **R:** ACGTTTTCATCATCTTTGGTAAA	Levin, R. A. et al. (2003) [79] Kress, W. J. et al. (2007) [51]
trnH-psbA	**F:** CGCGCATGGATTCACAATCC **R:** GTTATGCATGAACGTAATGCTC	Tate, J. A. et al. (2003) [74] Sang, T. et al. (1997) [80]

4.4 DNA SEQUENCING AND BIOINFORMATIC ANALYSIS

Following PCR and gel electrophoresis, remaining samples of *Verbascum* L. which gave clear gel electrophoresis results from *matK*, *rbcL* and *ITS* primers were prepared and forwarded for the sequencing procedure without undergoing PCR clean-up. Further sequence determination process was done by outside companies. Triogen Biotechnology for Sanger sequencing by using ABI3730XL and Massive Bioinformatics for NGS by using MiniON type Nanopore sequencing.

Qiagen's CLC Genomic Workbench 23 was used for comparative sequence analysis. All the processes including creating primer alignment, assembled primer alignment, creating maximum phylogeny tree for both primers, and assembled primer, pairwise comparison as well as alignment chromatogram and FASTA files was added and done in CLC Genomics Workbench. Phred score was checked with FastQC v12.

5. RESULTS AND DISCUSSION

The increasing industrialization and environmental degradation lead to a planetary biodiversity crisis. The loss of habitat due to climate crisis and contaminants brings many animal and plant species to the edge of extinction thus the loss of genetic resources. Since almost 80 percent of the total biomass on the earth consists of plants, the loss of germplasm results in a significant diminishment in agriculture that further affects the whole ecosystem. By improving the conservation of germplasm, it is possible to protect the ecosystem as well which highly depends on the taxonomy and accurate identification of the species. From the viewpoint of genetic diversity, the location of Turkey is a significant position as a meeting point for Euro-Siberian, Mediterranean, and Irano-Turanian phytogeographical regions. The climate, topography, and geomorphology make Turkey the home of one of the most diverse plant species and endemism [1]. It is estimated that there are approximately 2700 endemic species in Turkey where the *Verbascum* L. genus has an important part of its distinctive biodiversity [2]. The distribution range of the *Verbascum* L. genus is narrow and the increasing risk of the destruction of its habitat creates a requirement for a strict conservation policy considering its therapeutic qualities and ecological relevance [3]. Based on the large number of species in the *Verbascum* L. genus and their varied physical traits, effective species identification using only conventional methods is difficult [4]. Therefore, the use of DNA barcoding techniques becomes essential in clearing up taxonomic uncertainty and making it easier to identify and categorize *Verbascum* L. species. Due to their physical similarity and possibility for hybridization, these endemic *Verbascum* L. species pose difficulties for correct identification and classification. In order to clarify the taxonomic relationships and boundaries and conserve *Verbascum* L. germplasm, it is crucial to use DNA barcoding approaches that cover various gene areas, including *ITS*, *matK*, *rbcL*, and *trnH-psbA*. DNA barcoding has become a potent technique for species identification and categorization in the field of biodiversity research, transcending the conventional dependence on morphological features. To identify and differentiate plant species, DNA barcoding uses small gene sections as molecular markers, such as *ITS*, *matK*, *rbcL*, and *trnH-psbA* [28]. Plant species have a significant genetic diversity, making it difficult for a single barcode region to accurately capture the genetic variation needed to distinguish between species [7]. Additionally, it can be challenging to differentiate between some plant species

with just one primer because of conserved or identical sequences in specific barcode areas. Researchers can gain a more thorough understanding of the genetic variety among plant species by utilizing numerous barcode regions or genomic markers, which will improve the accuracy and reliability of identifying species in DNA barcoding investigations [31]. Recent developments have demonstrated that the accuracy and efficiency of plant barcoding are improved using primer pairs or triples that target several gene areas, such as *ITS-matK* or *ITS-matK-rbcL*. These combinations offer supplementary data that helps with taxonomic determination and reduces ambiguity [7]. Among the plant species being studied, *Verbascum* L. is significant due to its medicinal properties and ecological value. Given the difficulties associated with recognizing and classifying *Verbascum* L. species using conventional techniques, DNA barcoding becomes essential for clearing up taxonomic ambiguities [15]. The objective of this study is the molecular identification of key species in the *Verbascum* L. genus by DNA barcoding. For this purpose, the Anatolian endemic species from the *Verbascum* L. genus will be selected and identified by DNA barcoding method. This will enable the addition of *Verbascum* L. to Turkey's plant inventory and make molecular level adjustments to correct taxonomic problems if necessary. In addition, the involvement of the DNA barcodes will further contribute to the conservation of *Verbascum* L. germplasm. There were 25 total samples, 23 herbarium and 2 fresh tissues, at the beginning of the experiment from 20 different species of *Verbascum* L. Yet, during isolation process, some of the samples did not give acceptable results to proceed to next steps. So, after the isolation process, there were 19 samples remained, 17 herbarium and 2 fresh tissues, which contained 17 different species of *Verbascum* L. (Table 5.1).

Botanists have selected *Verbascum* L. species with great attention in order to verify and rectify species classifications. A number of factors, including morphological traits, regional distribution, and ecological context, are taken into account throughout the meticulous selection process. The goal of specimen selection for botanists is to accurately portray the diversity within the genus *Verbascum* L. This guarantees that the chosen samples are typical of the many species and subspecies, enabling a thorough analysis of their taxonomic characteristics.

Herbarium samples are advantageous since all *Verbascum* L. were taken directly from Turkey's natural habitats, guaranteeing that the samples accurately reflect the genetic

diversity found in those particular locations. A significant benefit of herbarium specimens is their accessibility. These collections offer a wide variety of plant species from different places and eras, acting as repositories of biodiversity. Herbarium samples are freely accessible to researchers, enabling a wider range of investigations without requiring ongoing fieldwork [81]. This accessibility is especially helpful for carrying out large-scale or retrospective research, which advances our knowledge of plant genetics. Herbarium samples also present a special chance to carry out in-depth, long-term research [81]. By comparing specimens from various eras, researchers can look into how plant populations have changed throughout time. This temporal dimension sheds light on how plant species have evolved and how environmental influences have shaped their genetic composition.

Nevertheless, despite these benefits, using herbarium materials in genetic research presents several difficulties, chiefly associated with the preservation techniques used. Herbarium specimens DNA deteriorates with time as a result of aging naturally and chemical treatments. It can be more difficult to recover intact genetic material when fixatives like formaldehyde and other preservatives are used because they can create chemical alterations to the DNA [82]. One of the challenges in extracting DNA from herbarium specimens is the degradation of the material. To retrieve useful DNA fragments, researchers must use specialized methods, frequently concentrating on shorter target sequences. The extraction procedure is additionally made more difficult by the possibility of contamination and the existence of inhibitors in herbarium DNA [82].

Two distinct isolation techniques, CTAB Isolation and the Macherey-Nagel Plant DNA Isolation Kit, were used to address and solve possible degradation problems. A comparative study showed that the DNA products made with the Macherey-Nagel Kit performed better since gel electrophoresis results of Macherey-Nagel protocol isolated samples were a lot clearer compared to CTAB protocol isolated samples. Even though concentration numbers were acceptable for both isolation processes, CTAB protocol isolated samples had a lot more background noise in gel electrophoresis results [Appendix E]. This improvement might be explained by the kits ability to purify DNA more effectively than the CTAB approach, producing DNA with a higher degree of purity. Commercial kits are made to be consistent and repeatable. They adhere to established procedures, which lessens variation in the extraction procedure. On the other hand, results from manual techniques such as CTAB

extraction could be inconsistent due to human error and differences in reagent preparation [83].

Moreover, the commercial kit performs better in downstream applications due to its effectiveness in eliminating impurities such proteins, polysaccharides, and secondary metabolites. This characteristic is essential for getting trustworthy and clean genetic data, which enhances the researches overall success and validity.

Gradient PCR is a useful method used in molecular biology to determine the ideal annealing temperature for PCR amplification. Using this technique, a variety of temperatures are tested in order to evaluate how they affect the amplification process's efficiency and specificity. An important PCR parameter is the annealing temperature, which affects how well primers bind to the target DNA and, ultimately, how well the amplification goes. Scientists may optimize conditions to achieve efficient and specific amplification, guaranteeing accurate and dependable findings in molecular biology experiments, by methodically adjusting the temperature of annealing in a gradient PCR [84]. A variety of annealing temperatures are examined in a single experiment with gradient PCR. This is accomplished by progressively adjusting the temperature across the reaction tubes using a thermal cycler. Finding the optimal temperature for primer binding to the DNA template is the aim of this procedure. The observation aimed to ascertain which annealing temperature offers the maximum specificity and efficiency of DNA amplification by conducting the PCR reaction at a gradient of temperatures [85]. Through the method, the PCR settings for a specific set of primers and template DNA were optimized, which improved the amplification processes success and produced accurate results. Because there were several *Verbascum* L. samples, it was necessary to optimize the annealing temperature because one primer annealing temperature was insufficient for all of the samples [84].

As shown in Table 4.1, 4.2 and 4.3 the ideal annealing temperatures were chosen for differing primers. Gradient PCR has been configured to record temperature readings three degrees above and three degrees below the figure's stated value. However, gradient PCR results did not greatly affect the outcomes because substantial changes were not seen even at higher temperature ranges.

One common chloroplast gene used in plant DNA barcoding is *matK*. Its significant sequence variation is one of its main characteristics, which makes it useful for differentiating between closely related plant species. *MatK* showed more divergence than *rbcL*, another widely used chloroplast gene, in comparative research covering 490 vascular plant species [86]. This increased divergence was noted both between and within species (intra-specificity and inter-specificity). *matK*s increased variability improves its ability to distinguish between plant species at the individual and species levels. On the other hand, *ITS* is a nuclear gene used in plant DNA barcoding. Its considerable sequence variation and ability to distinguish closely related species make it a valuable tool in plant DNA barcoding. It is useful for differentiating plant species that have close genetic affinities because of its great diversity and quick evolution. The *ITS* regions known high sequence variation is essential for creating universal primers that allow for the amplification of this region across a broad range of plant species. *ITS* has shown itself to be effective in species distinction in a number of investigations. The research indicates that, after *matK* [87], *ITS*2 was the second-best performer in species distinction. The research also emphasizes how *ITS* can be used in concert with other genes, such *matK* and *rbcL*, to improve the accuracy of species identification [88].

5.1 DNA ISOLATION

Total DNA was isolated from samples that has initial weight within the range of 50-60 mg except *Verbascum hasbenlii* and *Verbascum faik-karaveliogullari Çıngay&Cabi*, which had initial weights of 28.5 mg and 45.6 mg respectively. Total DNA was isolated using Macherey – Nagel NucleoSpin Plant II isolation kit according to manufacturers' instructions. The concentrations and purities of the samples were analyzed with Nanodrop. The concentrations of the isolated DNA samples were ranged between 6.73 ng/µl and 37.17 ng/µl for *Verbascum alyssifolium* and *Verbascum tuna-ekimii Karavel* respectively. The nucleic acid concentrations, purity values (OD260/280, OD260/230) and initial sample weight of all isolated DNA samples were shown in Table 5.1.

Table 5.1. Concentration, purity, and initial weights of the samples.

Herbarium Samples					
Botanical Name	**NGBB Barcode**	**Nucleic Acid Concentration (ng/µl)**	**OD260/280**	**OD260/230**	**Initial Sample Weight (mg)**
Verbascum sinatum ssp sinatum	5660	31.56	1.78	0.58	56.2
Verbascum sinatum ssp sinatum	9471	12.19	1.79	0.50	58.2
Verbascum speciosum Schrad	9817	22.50	2.84	0.43	52.7
Verbascum tuna-ekimii Karavel	5628	37.17	2.00	0.37	59.2

Verbascum vulcanicum Boiss&Hedr var vulcanicum Boiss&Hedr	11352	14.41	2.02	0.83	54.6
Verbascum glomeratum Boiss	5281	11.80	2.27	0.39	58.4
Verbascum glomeratum Boiss	8856	14.68	2.40	0.47	51.8
Verbascum isauricum Boiss&Heldr	5277	22.05	1.81	0.58	54.7
Verbascum luridiflorum Hub-Mor	8859	12.69	2.20	0.53	55.5
Verbascum ovalifolium ssp thracicum (Velen) Murb	9812	15.39	3.07	0.92	50
Verbascum ovalifolium ssp thracicum Donn ex Sims	9816	14.73	2.83	0.53	50
Verbascum pestalozzoe Boiss	8009	36.37	1.71	0.59	56.3
Verbascum hasbenlii	5887	6.83	1.64	0.56	28.5

Verbascum alyssifolium	5327	6.73	2.76	0.52	56.7
Verbascum blattaira	7794	10.69	1.85	0.30	54.6
Verbascum cilicicum	MÖZT6011	22.59	2.12	0.31	55.4
Verbascum faik-karaveliogullari Çıngay&Cabi	4801	31.57	2.29	0.59	45.6
Fresh Samples					
Botanical Name	**NGBB Barcode**	**Nucleic Acid Concentration (ng/µl)**	**OD260/280**	**OD260/230**	**Initial Sample Weight (mg)**
Verbascum vulcanicum	F.Vul	25.27	1.75	0.61	53.6
Verbascum hasbenlii	8875	11.49	2.22	0.52	54.5

5.2 PCR AMPLIFICATION OF *ITS* BARCODE REGION

ITS, *matK*, and *rbcL* barcode regions were proposed by Consortium for the Barcode of Life Plant Working Group (CBOL) to distinguish species [89]. Before sequencing, the target barcode regions were amplified by using gene-specific primers in PCR.

According to Hilu and Barthet (2008), the size of the *matK* gene is approximately 1500 base pairs [90]. In the case of *ITS* gene, the size is ranged from 260 base pairs to 1794 (*ITS1* and *ITS2*) base pairs [91] whereas for *rbcL*, the approximate size is around 1500 base pairs [92].

PCR amplification was performed with the isolated DNA from samples 9812, 9816, 5887 and 4801 in replicates to amplify *ITS* region. After that, the amplified samples were visualized by 1.5 percent agarose gel. Only the replicates of 5887 samples did not produce

a band in agarose gel. The produced bands were observed within the range 900-1000 base pairs in size for *ITS* (Figure 5.1). Each DNA sample were added to the gel with two replicates, and they represented as a and b in the gel results.

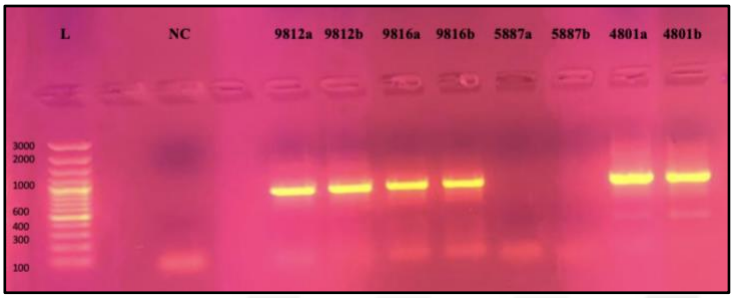


Figure 5.1. 1.5 percent agarose gel result of PCR amplification with *ITS* gene-specific primers of the *Verbascum* L. samples 9812, 9816, 5887 and 4801. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed on the isolated DNA from samples 5628, 8009,5887 and 7794 in replicates to amplify *ITS* region. The amplified samples were visualized by agarose gel electrophoresis. Only samples the replicates of 5628 and 7794 yielded results, showing distinct bands on the gel. The samples produced bands within the range 900-1000 base pairs. The intensity of the bands was higher in replicates of sample 7794 than 5628 (Figure 5.2).

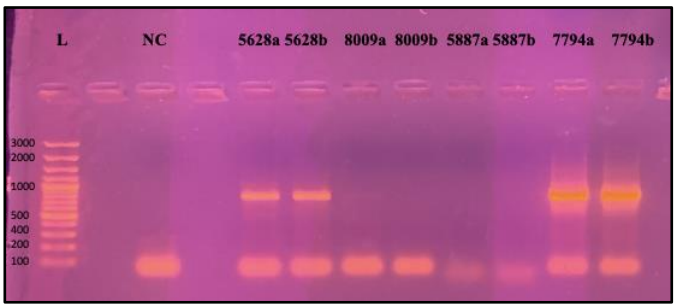


Figure 5.2. 1.5 percent agarose gel result of PCR amplification with *ITS* gene-specific primers of the *Verbascum* L. samples 5628 and 7794. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed by using *ITS* gene specific primers for samples 9471, 8859, 7794 and 5281 in replicates to amplify *ITS* barcode region. Subsequently, the

amplified samples were visualized by agarose gel electrophoresis. For the *ITS* gene, all samples except one replicate of 7794a yielded results. The remaining samples exhibited bands of approximately 1000 base pairs. However, even though sample 9471 produced a band at the expected bp, there were non-specific bindings thus more optimization was required in PCR conditions especially for annealing temperature. The lowest band intensity was observed for sample 5281. The faint bands observed at the end of the gel do not represent any significant findings as they are caused by primer dimers as well as the band observed in negative control. The intensity difference within the same replicate can be caused by technical errors. These deviations may arise from issues such as uneven mixing of the solution or uneven loading of samples due to pipetting errors. Unexpected bands, such as those observed in samples 8859, 9471, can arise due to nonspecific binding of DNA or RNA to the gel matrix. Sample preparation and PCR processes can potentially cause DNA can degradation and fragmentation, leading in the appearance of unexpected bands. Another factor contributing to these bands could be cross-contamination during sample preparation and loading procedures (Figure 5.3).

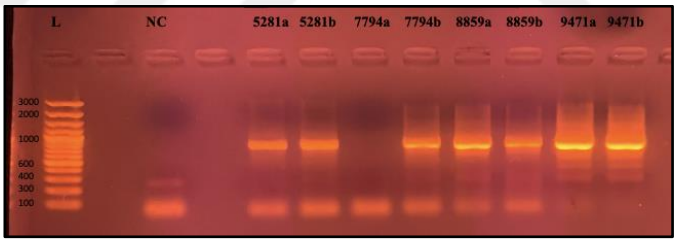


Figure 5.3. 1.5 percent agarose gel result of PCR amplification with *ITS* gene-specific primers of the *Verbascum* L. samples 5281, 7794, 8859, 9471. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed on the isolated DNA from samples 9817 and M6011 in replicates to amplify *ITS* region and the amplified samples were visualized by agarose gel electrophoresis. Only the replicates of 9817 samples yielded results within the range 900-1000 base pairs in size (Figure 5.4).

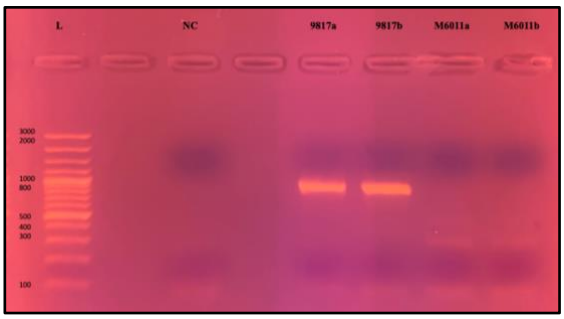


Figure 5.4. 1.5 percent agarose gel result of PCR amplification with *ITS* gene-specific primers of the *Verbascum* L. sample 9817. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed on the isolated DNA from samples 5628, 8009, 5887 and 7794 in replicates to amplify *ITS* region. The amplified samples were visualized by agarose gel electrophoresis. Only samples the replicates of 5628 and 7794 yielded results, showing distinct bands on the gel. The samples produced bands within the range 900-1000 base pairs. Higher band intensity was observed in the replicates of sample 7794 compared to sample 5628, indicating a possible influence of technical errors (Figure 5.5).

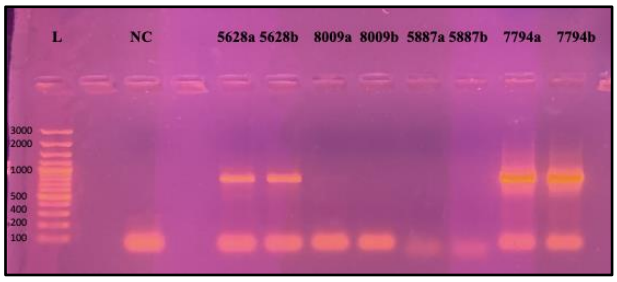


Figure 5.5. 1.5 percent agarose gel result of PCR amplification with *ITS* gene-specific primers of the *Verbascum* L. samples 5628 and 7794. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed on the isolated DNA using *ITS* gene-specific primers. Different samples were utilized in replicates for the *ITS* primers. The amplified samples were visualized by agarose gel electrophoresis. However, none of the samples showed results (Figure 5.6).

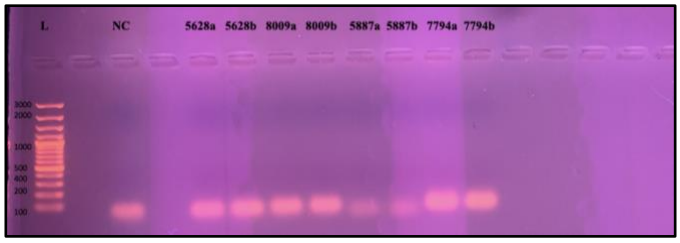


Figure 5.6. 1.5 percent agarose gel result of PCR amplification with *rbcL-TmF* gene-specific primers of the *Verbascum* L. samples 8859 and 4801. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

5.3 PCR AMPLIFICATION OF *matK* BARCODE REGION

PCR amplification was performed on the isolated DNA from samples number 9812, 9816, 5887 and 4801 in replicates to amplify the *matK* regions and visualized by agarose gel electrophoresis. Only samples the replicates of 4801 yielded results, showing distinct bands on the gel. The bands observed were corresponded to the expected size of approximately 1000 base pairs in size (Figure 5.7). Each DNA sample were added to the gel with two replicates, and they represented as a and b in the gel results.

Figure 5.7. 1.5 percent agarose gel result of PCR amplification with *matK* gene-specific primers of the *Verbascum* L. sample 4801. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed by using *matK* gene-specific primers separately on the isolated DNA from samples 9471, 8859, 7794 and 5281 in replicates to amplify *matK* regions. Subsequently, the amplified samples were visualized by agarose gel electrophoresis. In the case of the *matK* gene, all samples yielded results except for both replicates of 7794.

The observed bands in these samples fell within a similar size range of approximately 900-1000 base pairs. Notably, the bands of sample 9471 displayed highest intensity compared to the other samples, while sample 5281 exhibited lower band intensity. There are many reasons for the intensity variation within the same replicates of samples. Uneven distribution or mixing of the samples components, which might cause changes in the concentration of the target molecules, is one potential cause. Inadequate mixing may leave some areas with concentrations that are greater or lower than others, which will change how intense the bands appear on the gel electrophoresis. Uneven amounts of the sample being put onto the gel as a result of loading mistakes is another possibility. Because the amount of target molecules in each lane varies, this can lead to different band intensities. Intensity discrepancies can also be caused by technical problems like incorrect pipetting or uneven voltage application during electrophoresis. Additionally, several factors can affect the intensity difference in gel electrophoresis between different samples. Variations in the concentration or quantity of the target molecules present in each sample are one of the main causes. In comparison to samples with lower concentrations, samples with higher concentrations of the target molecules are likely to show stronger band intensities. Differences in sample quality or purity can also be a factor since contaminants or impurities might impact how the target molecules move and appear on the gel. Furthermore, because the samples are herbarium specimens, their preservation techniques can differ, particularly in terms of the duration of time that preservation is required. This variance in preservation time may have an effect on the samples' quality. In order to reduce these discrepancies and ensure precise and reliable outcomes, it is crucial to thoroughly control and standardize the experimental techniques (Figure 5.8).

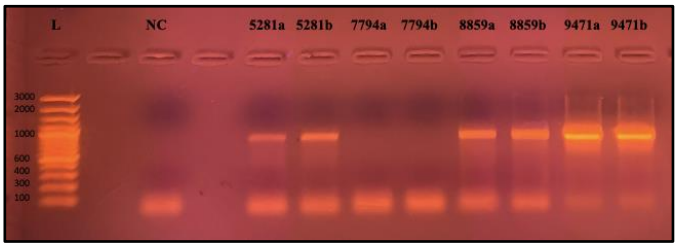


Figure 5.8. 1.5 percent agarose gel result of PCR amplification with *matK* gene-specific primers of the *Verbascum* L. sample 5281, 8859 and 9471. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed on the isolated DNA from samples 5628, 11352, 9812, 9816, M6011 and 7794 in replicates to amplify *matK* region followed by the visualization by agarose gel electrophoresis. Only samples the replicates of 5628, 11352 yielded results, showing distinct bands on the gel. The samples produced bands within the range 900-1000 base pairs (Figure 5.9).

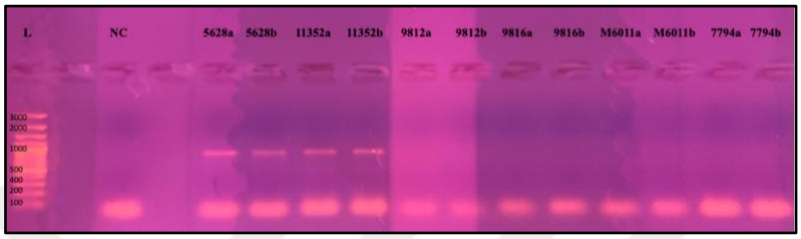


Figure 5.9. 1.5 percent agarose gel result of PCR amplification with *matK* gene-specific primers of the *Verbascum* L. samples 5628 and 11352. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

5.4 PCR AMPLIFICATION OF *rbcL* BARCODE REGION

PCR amplification was performed on the isolated DNA using separate sets of primers for *rbcL* gene. Different samples were utilized in replicates for both *rbcL* primers. The amplified samples were visualized by agarose gel electrophoresis. Each DNA sample were added to the gel with two replicates, and they represented as a and b in the gel results.

For *rbcL*, both 8859 and 4801 samples produced bands within the range of 700-900 base pairs. Notably, sample 4801 exhibited higher band intensity compared to sample 8859. As it was mentioned, differences in gel electrophoresis intensity between samples are influenced by a variety of factors. Band strength can vary depending on the concentration of the target chemical, with larger concentrations producing stronger intensities. The quality and purity of the sample can also be important since impurities alter how molecules migrate across the gel. The preservation procedure is crucial for both the 8859 and 4801 samples since they are herbarium specimens. There are differences in the preservation methods used for herbarium specimens, which could have an impact on sample quality due to different preservation times. Neither sample 8859 nor 4801 produced bands on agarose gel with *rbcL* gene-specific primers. The failed amplification might be a result of non-compatibility of primers with plant

species. However, since the used primers are universal the highest possibility of the failure is non-optimized PCR conditions (Figure 5.10).

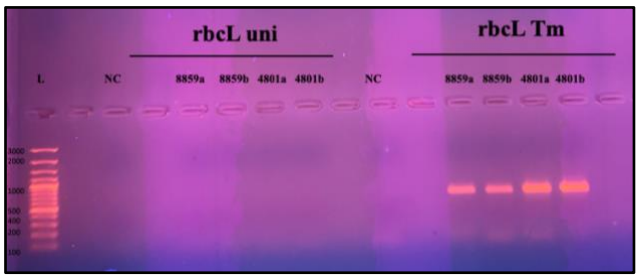


Figure 5.10. 1.5 percent agarose gel result of PCR amplification with *rbcL-Tm* gene-specific primers of the *Verbascum* L. samples 8859 and 4801. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

PCR amplification was performed by using two different *rbcL* primers separately on the isolated DNA from samples 8875 and 5277 in replicates to amplify *rbcL* region. The amplified samples were visualized by agarose gel electrophoresis. Only the replicates of 8875 samples from *rbcL Tm* primer yielded results within the range 900-1000 base pairs in size. *rbcL* provides distinction within the same species. That may explain the contrasting results obtained from the *rbcL* analysis of 8857 and 5277. This suggests that genetic variations within the species are responsible for this discrepancy. Neither sample 8875 nor 5277 produced bands on agarose gel with *rbcL uni* gene-specific primers. The failed amplification might be a result of non-compatibility of primers with plant species. However, since the used primers are universal the highest possibility of the failure is non-optimized PCR conditions (Figure 5.11).

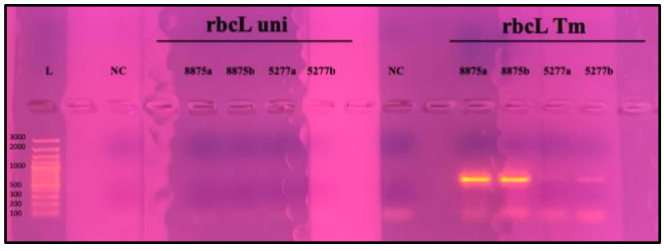


Figure 5.11. 1.5 percent agarose gel result of PCR amplification with *rbcL-uni* and *rbcL-Tm* gene-specific primers of the *Verbascum* L. sample 8875. L: Ladder (GeneRuler DNA Ladder Mix), NC: Negative Control.

Table 5.2. Final PCR amplification results.

Herbarium Samples			
Botanical Name	**NGBB Barcode**	***ITS***	***matK***
Verbascum sinatum ssp sinatum	5660		
Verbascum sinatum ssp sinatum	9471	x	x
Verbascum speciosum Schrad	9817	x	x
Verbascum tuna-ekimii Karavel	5628	x	x
Verbascum vulcanicum Boiss&Hedr var vulcanicum Boiss&Hedr	11352	x	x
Verbascum glomeratum Boiss	5281	x	x
Verbascum glomeratum Boiss	8856	x	x
Verbascum isauricum Boiss&Heldr	5277	x	x
Verbascum luridiflorum Hub-Mor	8859	x	x
Verbascum ovalifolium ssp thracicum (Velen) Murb	9812	x	
Verbascum ovalifolium ssp thracicum Donn ex Sims	9816	x	
Verbascum pestalozzoe Boiss	8009		
Verbascum hasbenlii	5887		
Verbascum alyssifolium	5327	x	x
Verbascum blattaira	7794	x	
Verbascum cilicicum	MÖZT6011		
Verbascum faik-karaveliogullari Çıngay&Cabi	4801	x	
Fresh Samples			
Botanical Name	**NGBB Barcode**		
Verbascum vulcanicum	F.Vul	x	x
Verbascum hasbenlii	8875	x	x

This studies main goal is to use DNA barcoding for the molecular identification of important *Verbascum* L. genus species, with a particular focus on Turkey. The *Verbascum* L. genus indigenous species from Anatolia were chosen for analysis. To do this, PCR was used to amplify two putative barcode regions, *ITS* and *matK*, from 15 and 11 different *Verbascum* L. species, respectively (Table 5.2.).

This work also successfully amplified from several *Verbascum* L. species an additional DNA barcode area, *rbcL*, in addition to the *ITS* and *matK* regions (Figure 5.10. and Figure 5.11.). Combining *rbcL* with the *matK* and *ITS* data allowed for effective amplification and analysis.

It is important for a number of reasons that *matK* and *ITS*, as well as *rbcL*, produced successful outcomes. First off, the genetic fingerprint of the investigated *Verbascum* L. species is provided by these three barcode areas. Each area gives distinctive genetic data, enabling a more thorough and precise identification of species. Additionally, the effective amplifying and analyzing of numerous barcode regions raises the accuracy and trust of the molecular identification procedure. It enables cross-referencing and results validation, lowering the likelihood of misidentification or misunderstanding. Researchers can better comprehend the evolutionary history and phylogenetic relationships of *Verbascum* L. species in Turkey by merging the data from these three barcode regions. A more thorough evaluation of the genetic diversity and differentiation within the species is made possible by this integrated approach, revealing light on their evolutionary processes, adaptability, and potential ecological importance. Finally, the positive outcomes from *rbcL*, *matK*, and *ITS*, support the legitimacy and efficacy of employing DNA barcoding for species identification, taxonomic revisions, and conservation efforts within the *Verbascum* L. genus.

5.5 DNA SEQUENCE ANALYSIS

Table 5.3. Sequenced samples for NGS and Sanger.

Herbarium Samples		NGS		Sanger	
Botanical Name	**NGBB Barcode**	ITS	matK	ITS	matK
Verbascum sinatum ssp sinatum	005660	-	-	-	-
Verbascum sinatum ssp sinatum	009471	x	x	-	-
Verbascum speciosum Schrad	009817	x	x	-	-
Verbascum tuna-ekimii Karavel	005628	x	x	-	-
Verbascum vulcanicum Boiss&Hedr var vulcanicum Boiss&Hedr	11352	-	-	x	x
Verbascum glomeratum Boiss	005281	x	x	-	-
Verbascum glomeratum Boiss	008856			x	x
Verbascum isauricum Boiss&Heldr	005277	x	x	-	-
Verbascum luridiflorum Hub-Mor	008859	-	-	x	x
Verbascum ovalifolium ssp thracicum (Velen) Murb	009812	-	-	x	-
Verbascum ovalifolium ssp thracicum Donn ex Sims	009816	-	-	x	-
Verbascum pestalozzoe Boiss	008009	-	-	-	-
Verbascum hasbenlii	005887	-	-	-	-
Verbascum alyssifolium	005327	x	x	-	-
Verbascum blattaira	007794	-	-	x	-
Verbascum cilicicum	MÖZT6011	-	-	-	-
Verbascum faik-karaveliogullari Çıngay&Cabi	004801	x	x	-	-
Verbascum speciosum Schrad	009813	-	-	-	-
Verbascum xanthophoeniceum Griseb	009723	-	-	-	-
Verbascum phoneicum	10015	-	-	-	-
Verbascum phoneicum	10007	-	-	-	-
Verbascum gypsicola Vural&Aydogdu	005334	-	-	-	-
Verbascum alyssifolium	002002	-	-	-	-
Fresh Samples					
Botanical Name	**NGBB Barcode**				
Verbascum vulcanicum	F.Vul	-	-	x	x
Verbascum hasbenlii	008875	-	-	x	x

The Table 5.3 gives a thorough overview of the testing status for two sequencing approaches: Next-Generation Sequencing (NGS) and Sanger Sequencing on *matK* and Internal Transcribed Spacer (*ITS*) barcode areas. "x"-designated entries show that the appropriate category has been examined in the designated barcode region with the shown sequencing technique. On the other hand, entries that have a "-" next to them indicate that no testing has been done in that category using the matching sequencing method in the designated barcode region.

The phylogenetic trees illustrate the evolutionary relationships based on the analysis of *ITS* and *matK* gene sequences, both by Sanger and NGS, within the *Verbascum* L. genus and its relationships with selected outgroups, including *Veronicastrum sibiricum, Antirrhinum majus, Coffea canephora*, and *Coffea arabica* outgroups. The branches of the phylogenetic tree in this representation reflect the evolutionary distances that accumulate over time in addition to the diverse paths that various *Verbascum* L. species have taken [93].

Outgroups are essential to building the evolutionary structure among the taxa under study in phylogenetic tree research. The main goal of adding an outgroup is to give scientists a point of reference to establish the trees root and help them distinguish between derived and ancestral lineages [93,94]. The absence of an outgroup makes it difficult to determine the direction of evolution within the tree, which leaves the study lacking a clear point of reference for comprehending relationships between various species. In tree rooting, the usage of outgroups helps to prevent circular thinking. In the absence of an external standard, there is a chance that the tree will be rooted only on the basis of characters seen in the ingroup, which could lead to bias or illogical reasoning. By offering a separate point of reference for determining the root, the addition of an outgroup reduces this risk and improves the impartiality and accuracy of the phylogenetic study [93]. After reviewing the literature, outgroups for *Verbascum* L. barcoding were selected as *Veronicastrum sibiricum*, *Antirrhinum maju, Coffea canephora* and *Coffea arabica* [15, 58].

Two essential elements in the field of phylogenetic trees improve our comprehension of the trees dependability: bootstrap values in a tree that has undergone bootstrap analysis, and the reference bar in a maximum likelihood (ML) tree. Based on observed sequence data, Maximum Likelihood algorithms calculate the ideal branch lengths and tree structure [95]. The evolutionary distance is shown by the reference bar, which is positioned beneath the tree branches. Longer branches indicate higher genetic divergence, whereas shorter branches indicate closer genetic ties. It is possible to understand the historical paths that have shaped the *Verbascum* L. genus more thoroughly by this complex representation [93].

A resampling method called bootstrap analysis assesses how resilient the branching pattern of a phylogenetic tree is. Bootstrap values connected with tree branches indicate the percentage of times a particular branch or clade is supported throughout the resampled datasets. Several datasets are constructed by sampling from the original dataset. Strong support for matching branches is indicated by high bootstrap values (usually above 70–80 percent), whereas lower values imply less confidence. Values above 70–80 percent are generally regarded as reasonably strong support, while there is no hard barrier; values below 50 percent should be regarded cautiously. Relative evolutionary distances can be better understood with the help of the reference bar, and the bootstrap values provide a level of statistical support for the inferred branching patterns. When combined, these elements give

researchers useful instruments for evaluating the validity and dependability of the phylogenetic relationships that are shown in a tree [96].

5.5.1 Sanger Sequencing Analysis

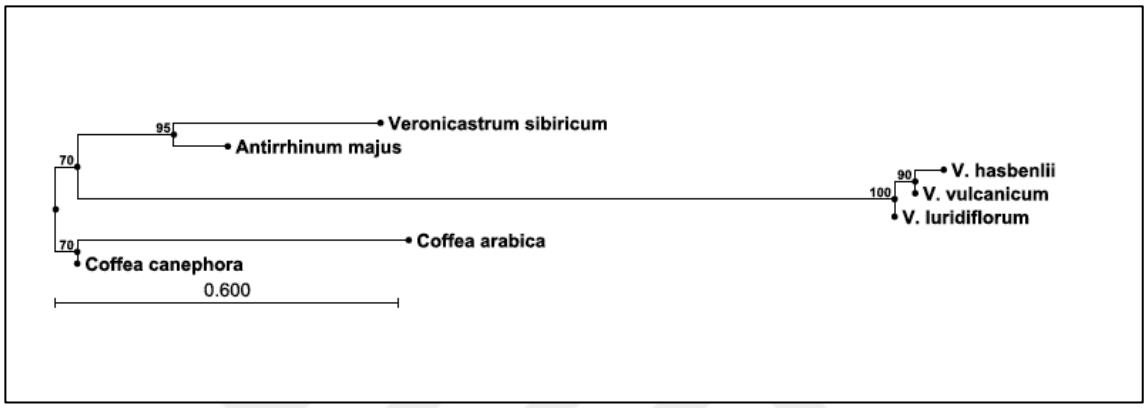


Figure 5.12. Phylogenetic tree of *Verbascum* L. species based on *ITS* gene sequences and Sanger sequencing.

In figure 5.12, the phylogenetic tree was constructed using *ITS* gene sequences obtained through Sanger sequence analysis. Notably, *Coffea canephora* and *Coffea arabica* branched early, indicating an early divergence within the *Coffea* genus as it was expected since they are the members of outgroup. Following this, the tree bifurcates into two main clades: the *Verbascum* L. clade and the clade comprising *Veronicastrum sibiricum* and *Antirrhinum majus*. The *Verbascum* L. clade, consisting of *V. hasbenli, V. vulcanicum,* and *V. luridiflorum* exhibits high bootstrap values of 100 and 90, indicating strong support for their close relationship and their affiliation with the *Verbascum* L. genus.

Within the *Verbascum* L. genus, two major clades with varying branch support are discernible. Specifically, *V. luridiflorum* diverges earliest in the phylogenetic tree, followed by *V. vulcanicum*. Interestingly, *V. hasbenlii* is positioned as the most recently evolved species within this context. These insights provide a comprehensive understanding of the evolutionary relationships among the studied species, shedding light on the temporal aspects of their divergence within the genus *Verbascum* L.

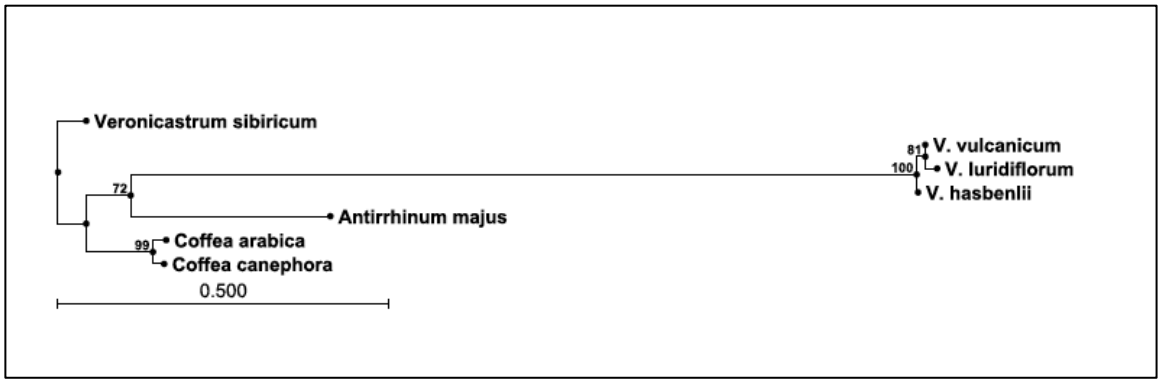


Figure 5.13. Phylogenetic tree of *Verbascum* L. species based on *matK* gene sequences and Sanger sequencing.

In figure 5.13, the construction of the phylogenetic tree relied on *matK* gene sequences. Notably, *Veronicastrum sibiricum* takes an early branching position, followed by *Coffea arabica* and *Coffea canephora*. The subsequent branch comprises *Antirrhinum majus* and the *Verbascum* L. clade. Within this clade, *V. hasbenlii* emerges as the earliest diverging species in the tree, followed by *V. vulcanicum* and, most recently, *V. luridiflorum*. This sequencing implies that *V. luridiflorum* is the most recently evolved species within the context of this phylogenetic analysis. The insights derived from the *matK* gene sequences provide a parallel perspective on the evolutionary relationships, contributing to a more comprehensive understanding of the genetic divergence and historical connections among the studied plant species.

When comparing two genes in the framework of a phylogenetic tree, it is necessary to carefully consider the branching and distance discrepancies. Variations in the evolutionary events that each gene records can be emphasized by branching differences, which indicate the organization and connection of taxa on the trees. Divergent branching patterns in the trees indicate that the genes can represent different theories about the evolutionary relationships between the taxa under study. Gene duplications, horizontal gene transfers, and other distinct evolutionary events specific to each gene could be the cause of these discrepancies [93].

In addition, the distance discrepancies between the trees built from the two distinct genes reveal information on the evolutionary dynamics and genetic divergence that are specific to

each gene. Based on the unique genetic information that each gene encodes, evolutionary divergence between taxa is indicated by genetic distances. Variations in these distances could be caused by variations in replacement models, rates of mutation, or other evolutionary restrictions on each gene. In order to interpret the biological information offered by different genes in the evolutionary context more nuancedly, researchers attempting to uncover the complexities of evolutionary relationships must possess an extensive knowledge of both branching and distance differences [93].

Table 5.4. Sanger Sequencing-Based Pairwise Comparison of *Verbascum* L. Species

		1	2	3	4	5	6	7
Coffea canephora	1		466	385	165	476	460	472
Coffea arabica	2	472		697	557	708	724	716
Antirrhinum majus	3	653	967		420	529	515	521
Veronicastrum sibiricum	4	481	873	587		517	503	509
V. vulcanicum	5	1245	1525	1239	1283		18	14
V. luridiflorum	6	1230	1533	1229	1272	42		12
V. hasbenlii	7	1245	1532	1233	1278	53	52	

Pairwise comparison is the process of matching up DNA, RNA, or protein sequences between two species in order to investigate the evolutionary relationships between species of the same plant (Table 5.4.). In molecular biology and bioinformatics, this phase is essential. Sequences are compared pair by pair after alignment to determine how similar or dissimilar they are at each point. There is a scoring system in place that takes gaps, mismatches, and matches into account. The result is a matrix that facilitates the analysis of evolutionary relationships by displaying the degree of resemblance or dissimilarity. With the aid of sequencing data, phylogenetic trees, which show the evolutionary relationships between species, can be constructed. Greater evolutionary divergence is suggested by more distant branches, whilst closer branches represent closely related species [97]. The genetic or molecular sequence data is the variable utilized for pairwise comparison in phylogenetic tree analysis. These sequences can come from protein, RNA, or DNA, depending on the kind of organisms being studied and the analysis's focus. Certain genomic regions—such as genes

or non-coding regions—that are chosen for their ability to shed light on evolutionary relationships may be included in the particular genetic material that is utilized as a variable [97].

In Table 5.4, the pairwise comparisons between *Antirrhinum majus, Veronicastrum sibiricum, Coffea canephora, Coffea arabica, V. vulcanicum, V. luridiflorum*, and *V. hasbenlii* are represented. In order to shed light on the evolutionary links among these different entities, the main goal of this investigation is to evaluate species correlation and identify relationships. Each row and column in the matrix correspond to a specific species, with the diagonal elements providing a measure of self-comparison. Off-diagonal entries represent pairwise comparisons between different species, and the values within these cells indicate the degree of similarity or dissimilarity in terms of evolutionary distance. Higher numbers indicate greater evolutionary divergence or dissimilarity, whereas lower values often indicate a greater evolutionary relationship or higher similarity. When comparing the *Verbascum* L. clade, which includes *V. vulcanicum, V. luridiflorum, and V. hasbenlii*, a clear pattern becomes apparent. These comparisons are highlighted in blue, indicating lower scores that categorically indicate higher similarity between these specific species. This color-coded distinction emphasizes the close evolutionary relationships within the *Verbascum* L. clade. Conversely, when considering the comparisons between the *Verbascum* L. clade and the outgroup species, the tree reveals higher values, indicative of greater dissimilarity. This distinction is crucial in understanding the broader context of the evolutionary relationships between the *Verbascum* L. clade and the outgroup species.

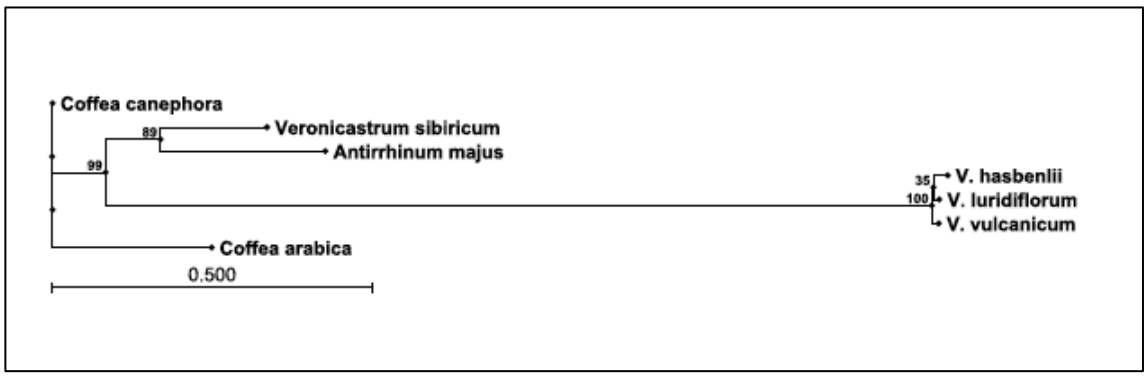


Figure 5.14. Assembled phylogenetic tree of *Verbascum* L. species based on Sanger sequencing.

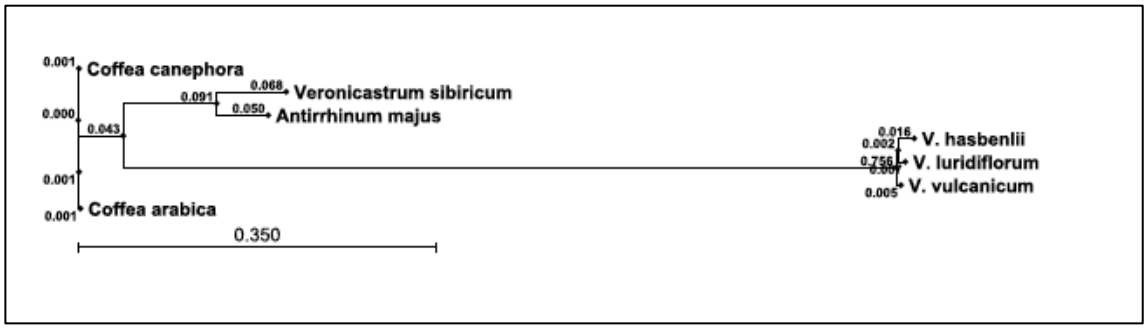


Figure 5.15. Maximum Likelihood phylogenetic tree of *Verbascum* L. species based on Sanger sequencing.

Branch lengths are represented by the combined values of the scale bar value 0.350 (Figure 5.15.) and the numerical value 0.756 for *Verbascum* L. clade distance to outgroup (Figure 5.15.) in a phylogenetic tree. These lengths, which are measured in units similar to substitutions per site for obtained molecular data, indicate the degree of evolutionary change along each branch [93]. For example, in the framework of molecular phylogenetics, such as plant DNA sequences, a branch length of 0.756 may represent an average of 0.756 substitutions per site along that specific branch which indicates significant difference in evolutionary tree (Figure 5.15.). When understanding these branch lengths in connection to the actual magnitude of evolutionary change becomes extremely important. A scale bar value of 0.350 in the example given indicates that there are 0.350 units of evolutionary change in one unit of length of the scale bar. Branch length estimate in relation to the standard reference is made easier by this scale.

When a branch with a length of 0.043 between first and second outgroup is evaluated alongside a scale bar of 0.350 (Figure 5.15.), it suggests that the branch in question reflects a comparatively little degree of evolutionary change in relation to the scale bar between outgroups. Put more simply, if the branch is shorter than the scale bar, it indicates that the genetic or morphological variations along that branch are not as great as the scale bars reference value. Understanding evolutionary connections within the phylogenetic tree depends in large part on the length of the branches; longer branches typically imply greater divergence or more cumulative changes over time, while shorter branches typically indicate less divergence or fewer evolutionary changes [93]. As a result, the scale bar offers a consistent visual reference for analyzing branch lengths in regard to evolutionary

relationships. In conclusion to that, with scale number 0.756, our *Verbascum* L. clade diverged enough from outgroup for classification purposes with the help of bootstrap values range from 89 to 100 (Figure 5.14.) for this group.

5.5.2 Next-Generation (NGS) Sequencing Analysis

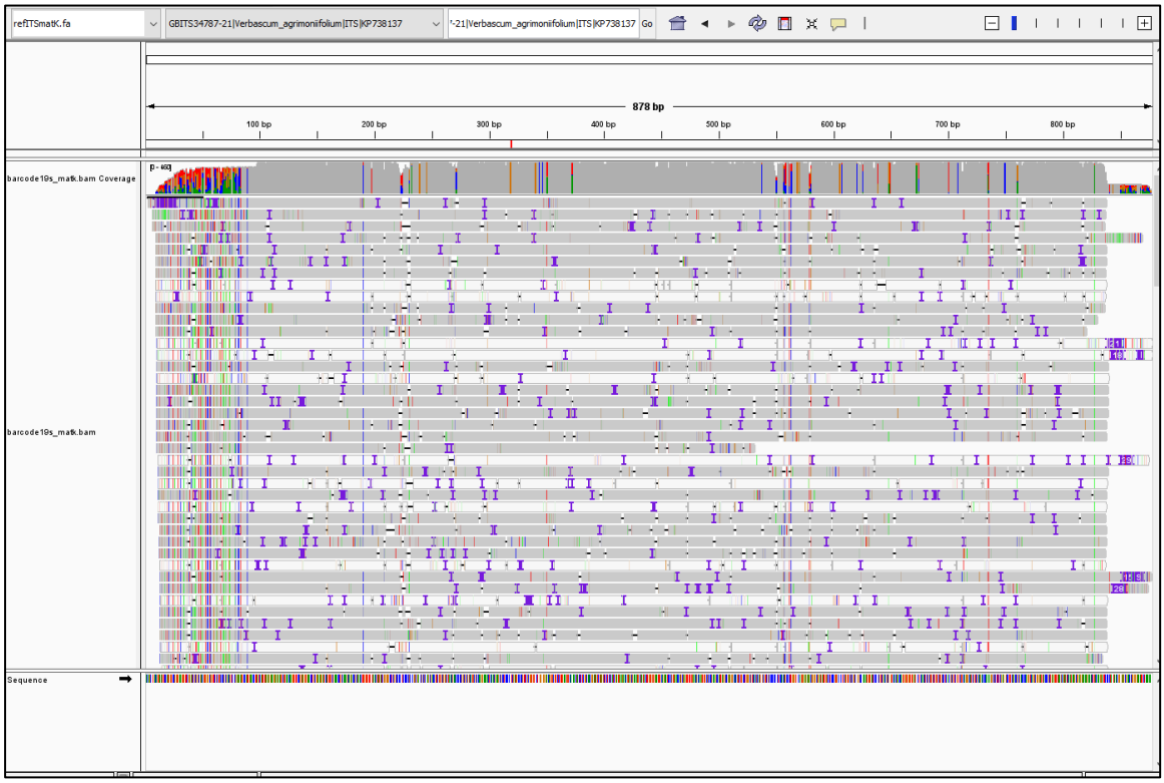


Figure 5.16. NGS data example of *Verbascum agrimonifolium* based on *ITS* gene sequences.

With its high-throughput capabilities, next-generation sequencing (NGS) is leading the way in genomic discovery and transforming the study of DNA. Due to its high match efficiency, millions of DNA fragments may be sequenced at once, yielding an abundance of data in remarkably quick periods of time. NGS's real strength is in how accurately it can identify alterations in DNA sequences [98]. NGS makes it easier to identify genetic changes such as single nucleotide polymorphisms (SNPs), deletions, insertions, and structural abnormalities because of its deep coverage and precision. Understanding population genetics, determining the evolutionary processes within and between species, and learning the genetic basis of

diseases are all made possible by this mutation-detecting skill. This ability, which is marked by accuracy and deep coverage, enables the thorough identification of genetic variations, from structural changes [99].

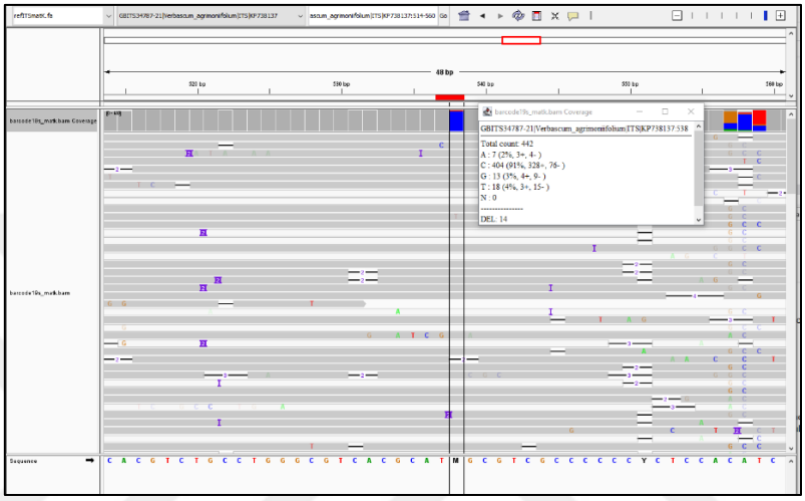


Figure 5.17. NGS data example of nucleotide count within *Verbascum agrimonifolium* based on *ITS* gene sequences.

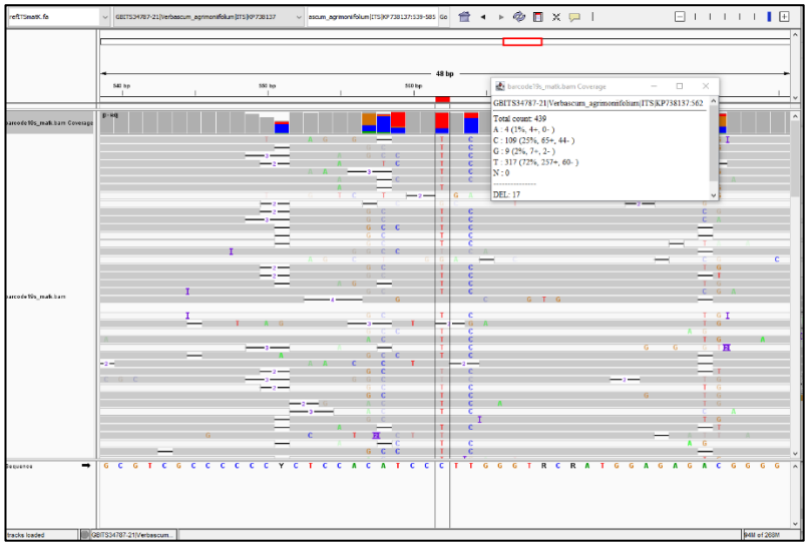


Figure 5.18. NGS data example of nucleotide counts within *Verbascum agrimonifolium* based on *ITS* gene sequences.

When this genomic power is used to plant DNA sequences of *Verbascum* L., NGS becomes an invaluable resource for revealing the genetic diversity present in many *Verbascum* L. species. NGS exposes mutations and genetic variations through pairwise comparisons from different reads, providing insight into the complex genomic topography of the *Verbascum* L. genus. This study argues that using the mutation-detecting powers of NGS not only improves our understanding of the evolutionary history of the plant but also offers important insights into the molecular factors influencing the various characteristics and adaptations seen in different *Verbascum* L. species.

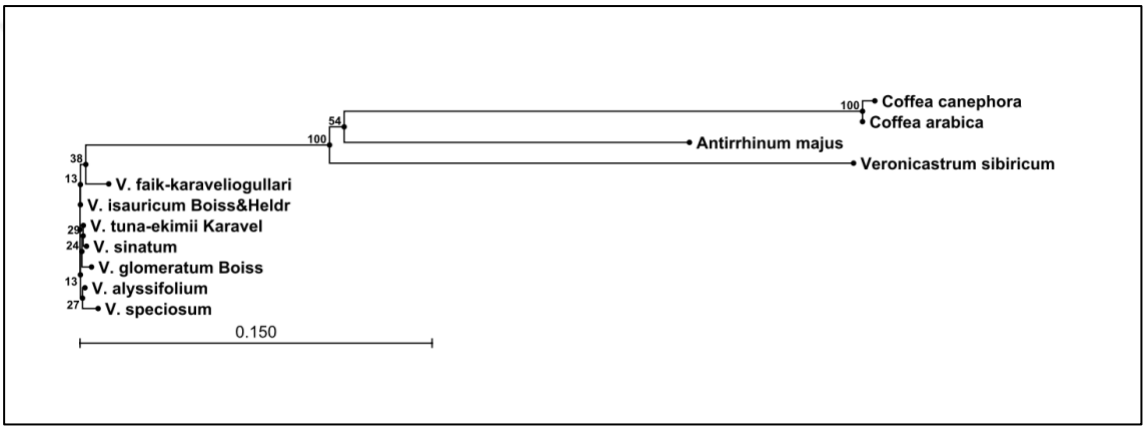


Figure 5.19. Phylogenetic tree of *Verbascum* L. species based on *ITS* gene sequences and NGS.

In Figure 5.19, the phylogenetic tree was constructed using *ITS* gene sequences obtained through NGS analysis. Notably, the initial branching occurred within the *Verbascum* L. clade, with outgroups (*Coffea canephora, Coffea arabica, Veronicastrum sibirium* and *Antirrhinum majus*) appearing later in the branching pattern. This divergence highlights the separation between the *Verbascum* L. clade and the outgroups. Upon closer inspection of the *Verbascum* L. clade, the species exhibit a close relationship. However, when considering the bootstrap values, their low magnitudes suggest conflicting or insufficient signal in the dataset. This could be attributed to a scarcity of informative sites or issues with alignment quality [96].

The diminished bootstrap values imply that different columns of the alignments yield varied tree topologies. One potential source of error contributing to these low values could be the

presence of rogue taxa, significantly impacting bootstrap support [96]. To enhance these values, diverse phylogenetic reconstruction methods, such as neighbor joining (NJ), minimum evolution (ME), and maximum parsimony (MP), can be employed [100]. Additionally, the combination of different markers, such as nuclear or mitochondrial genes, can augment the robustness of phylogenetic inferences for further analysis.

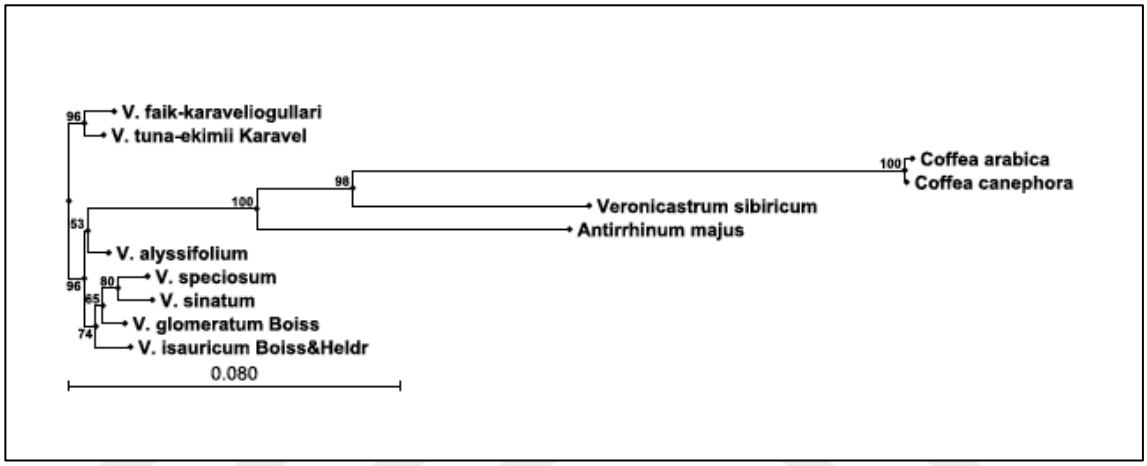


Figure 5.20. Phylogenetic tree of *Verbascum* L. species based on *matK* gene sequences and NGS.

In Figure 5.20, the phylogenetic tree was constructed utilizing *matK* gene sequences acquired through NGS analysis. Similar to the phylogenetic tree based on *ITS* gene, the initial branch in this tree corresponds to the *Verbascum* L. clade, distinct and separate from the outgroups. Notably, this tree exhibits greater clarity, characterized by higher bootstrap values. The primary branches comprise *V. faik-karaeliogullari* together with *V. tuna-ekimli Karavel*, followed by the remaining *Verbascum* L. clade and outgroups. Outgroups show significant difference in both Figure 5.19 and 5.20.

Table 5.5. NGS-based Pairwise Comparison of *Verbascum* L. species

		1	2	3	4	5	6	7	8	9	10	11
V. sinatum	1		0	1	1	0	1	2	87	87	88	77
V. speciosum	2	19		1	1	0	1	2	87	87	88	77
V. alyssifolium	3	19	24		2	1	0	3	86	86	89	78
V. glomeratum Boiss	4	20	21	17		1	2	3	88	88	89	78
V. tuna-ekimii Karavel	5	26	28	16	18		1	2	87	87	88	77
V. faik-karaveliogullari	6	36	40	25	32	18		3	86	86	89	78
V. isauricum Boiss&Heldr	7	20	21	16	15	20	30		85	85	86	79
Coffea canephora	8	350	350	340	347	351	349	343		0	99	108
Coffea arabica	9	348	348	338	345	349	347	341	5		99	108
Veronicastrum sibiricum	10	293	292	282	286	290	286	282	382	380		151
Antirrhinum majus	11	252	253	248	253	253	256	246	380	381	313	

In Table 5.5, when examining the comparisons within the *Verbascum* L. clade, a resemblance to the outcomes of the phylogenetic tree derived from Sanger sequencing is observed. These specific comparisons, similar to the Sanger sequencing-based results, are accentuated in blue, denoting lower scores that unequivocally signify heightened similarity among these particular species. Conversely, the comparisons between the *Verbascum* L. clade and the outgroups reveal higher scores, indicating diminished similarity between these species, aligning with the anticipated expectations. The *Verbascum* L. clade exhibits its lowest scores in pairwise comparisons, specifically between *V. sinatum* and *V. speciosum*, *V. sinatum* and *V. tuna-ekimii Karavel*, *V. speciosum* and *V. tuna-ekimii Karavel*, as well as *V. alyssiofolium* and *V. faik-karaveliogullari*. Singularly, the lowest score within the *Verbascum* L. clade is observed in the comparison between *V. speciosum* and *V. faik-karaveliogullari*.

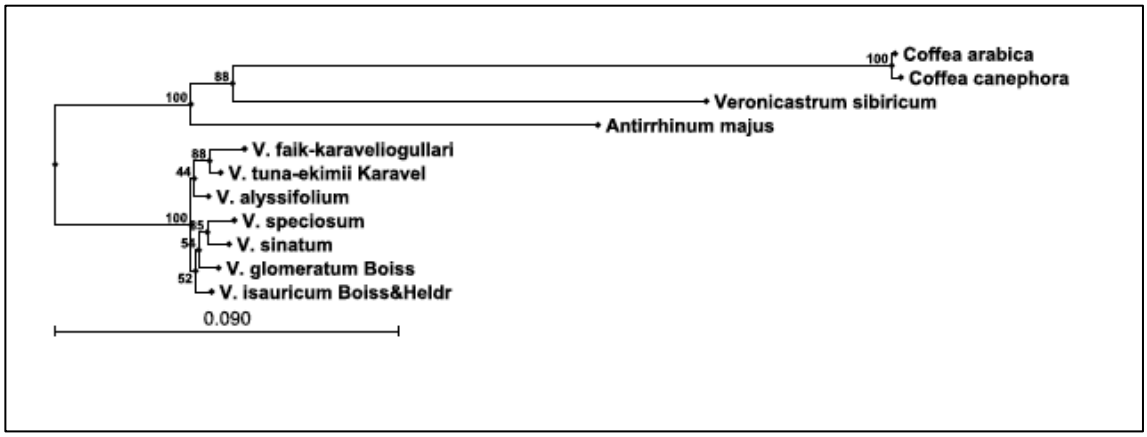


Figure 5.21. Assembled phylogenetic tree of *Verbascum* L. species based on NGS.

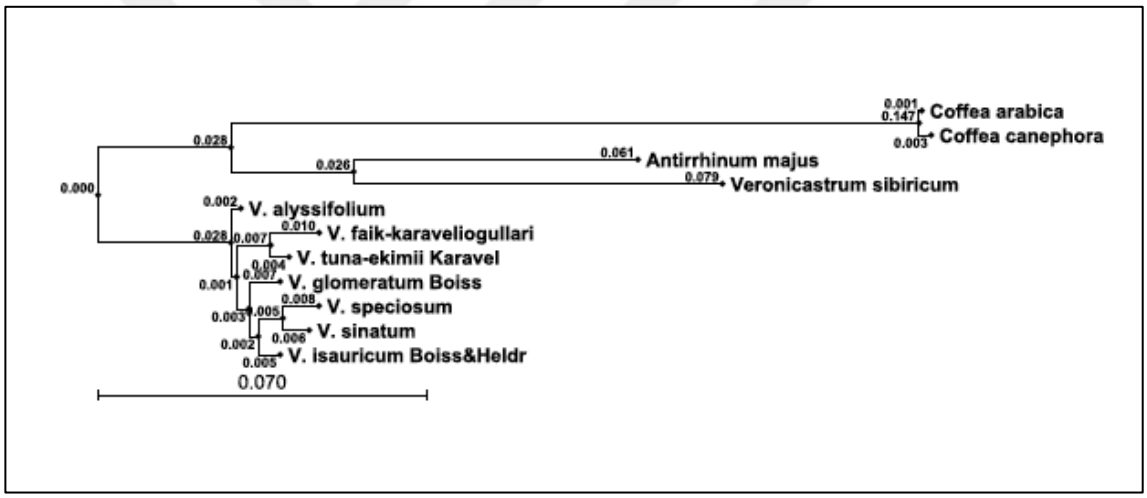


Figure 5.22. Maximum Likelihood phylogenetic tree of *Verbascum* L. species based on NGS.

In Figure 5.22, the phylogenetic tree was constructed using assembled *matK* and *ITS* gene sequences obtained through NGS analysis. Notably, the initial branching occurred between the *Verbascum* L. clade and all outgroups (*Coffea canephora*, *Coffea arabica*, *Veronicastrum sibirium* and *Antirrhinum major*). This divergence highlights the separation between the *Verbascum* L. clade and the outgroups. Upon closer inspection of the *Verbascum* L. clade, the species exhibit a close relationship supporting our values from separate constructed trees (Figure 5.19, Figure 5.20.). When considering the bootstrap values, their magnitudes suggest our *Verbascum* L. clade diverged enough from outgroup

for classification purposes for this group (Figure 5.21). The lower bootstrap values imply that different columns of the alignments show different tree topologies. One potential source of error contributing to these low values could be again the presence of rogue taxa, significantly impacting bootstrap support [96] also reason could be attributed to a issues with alignment quality.

Branch length of 0.028 between all outgroups from *Verbascum* L. clade shows clear evolutionary divergence in phylogenetic tree (Figure 5.22). Outgroup clade then diverges into branches itself with branch lengths with 0.026 and 0.147. *Verbascum* L. clade which consists of 7 *Verbascum* L. species, shows branch length values ranges from 0.002 to 0.010 which shows a comparatively little degree of evolutionary change in relation between *Verbascum* L. species. Within this clade, *V. alyssifolium* emerges as the earliest diverging species in the tree, followed by *V. glomeratum Boiss* and, most recently, *V. speciosum*. This sequencing implies that *V. speciosum* is the most recently evolved species within the context of this phylogenetic analysis based on NGS data. Longer branches imply greater evolutionary divergence or more cumulative changes over time which shows in Figure 5.19 between ourgroups and *Verbascum* L. clade.

6. CONCLUSION AND FUTURE PROSPECTS

The objective of this study is the molecular identification of key species in the *Verbascum* L. genus by DNA barcoding and to advance our knowledge of *Verbascum* L. species, their genetic diversity, and their ecological value within the Turkish phytogeographical environment by integrating a thorough molecular methodology. For this purpose, the Anatolian endemic species from the *Verbascum* L. genus will be selected and identified by DNA barcoding method. The target sequences were successfully confirmed to be present in the samples by PCR amplification, which was made possible using *ITS* and *matK* gene-specific primers. The PCR results were visualized by agarose gel electrophoresis, which revealed distinct bands that corresponded to the anticipated fragment sizes, demonstrating accurate amplification. The proposed barcode regions, *ITS* and *matK*, were amplified in *Verbascum* L. species respectively by PCR (Table 5.2). Different samples had different band intensities, indicating changes in DNA content or amplification effectiveness. Genetic variations are mostly determined by a species' evolutionary history. Significant genetic divergence from other species in the same genus can occur when a species experiences major evolutionary changes or adaptations. The genetic composition of species is shaped by evolutionary paths, which accounts for the differences seen in genetic comparisons.

Different ecological stresses and environments have a significant role in genetic differentiation. Differentially adapted species may be subject to distinct selective pressures, which can result in adaptations that modify their genetic makeup. The genetic difference between species inhabiting distinct ecological niches is influenced by the ecological setting, which encompasses elements including climate, resources, and competition. Genetic comparisons can be greatly impacted by variations in selective stresses, such as exposures to different diseases, climatic circumstances, or ecological niches. Genetic divergence may be caused by a species' exposure to various selective forces throughout time. Comprehending the selective pressures exerted on every species offers valuable perspectives on the genetic variations detected in comparisons between pairs. Also, sequence recovery for *matK* and *rbcL* was significantly decreased for herbarium specimens aged 10–30 years compared to those collected during the last ten years, as revealed by beta regression analysis. The success rate was even lower for older specimens. In contrast, only samples older than 50 years old

showed a significant loss in sequence recoveries for *ITS2*, even though this marker's average success rate at this age was similar to that of the plastid markers [101].

The genetic composition of different species varies due to variances in mutation rates. Over time, genetic divergence may occur more quickly in cases where mutation rates are higher. The degree of genetic differences seen in pairwise comparisons is influenced by the rate that genetic mutations accumulate, adding a dynamic element to the evolutionary process.

Pairwise comparisons of species reveal that structural alterations in the genome, including deletions, insertions, or chromosomal rearrangements, also have a role in the genetic differences between individuals. Gene order and function changes resulting from genomic rearrangements can add to the overall divergence across species. Comprehending these structural modifications offers significant understanding of the genetic dynamics influencing the variations in pairwise comparisons.

Overall, the combined outcomes of PCR and gel electrophoresis offered solid proof of successful DNA amplification verifying the effectiveness of the experimental technique and setting the stage for additional studies and research. The merging of Sanger and NGS analyses was both used to the exploration of *Verbascum* L. genetic database. Sanger sequencing, a conventional but trustworthy technique, was used to clarify the *ITS* and *matK* of the *Verbascum* L. species that were being studied. A fundamental knowledge of the genetic terrain was made possible by the thorough analysis of individual DNA sequences made possible by the accuracy and precision of Sanger sequencing. Numerous DNA sequences could be explored simultaneously thanks to NGS's enormous sequencing capability. A thorough review of the genomic region was made possible by this parallelized methodology, which made it possible to collect a variety of data that would have been difficult to obtain using traditional sequencing techniques.

The use of DNA barcodes has great potential for increasing the conservation of *Verbascum* L. germplasm, helping to protect and preserve this plant genus. DNA barcoding has revolutionized identifying species and genetic assessment, which greatly benefits *Verbascum* L. conservation. DNA barcoding tackles the problem of differentiating *Verbascum* L. species that share similar physical traits by providing accurate species identification. For creating efficient conservation plans that are unique to each species, this

accuracy is essential. Furthermore, *Verbascum* L. populations genetic diversity can be assessed using DNA barcoding, which offers important information regarding the adaptability and condition of the germplasm. Regular DNA barcode analyses help follow population dynamics, detect genetic diversity declines, and execute appropriate conservation measures. Additionally, DNA barcoding verifies the legitimacy of plant samples essential for study, reintroduction, and restoration projects, protecting the integrity of germplasm collections. Finally, DNA barcoding provides a uniform approach to species identification, which makes research results reliable and comparable. This supports larger research endeavors related to *Verbascum* L.s taxonomy, ecology, and evolution.

REFERENCES

1. Karagöz, A. Plant genetic resources conservation in Turkey. In: *International Symposium on Sustainable Use of Plant Biodiversity to Promote New Opportunities for Horticultural Production* 598. 2001:17-25.
2. Karavelioğulları, F. A., Duran, A., Hamzaoğlu, E. Verbascum tuna-ekimii (Scrophulariaceae), a new species from Turkey. *Ann Bot Fennici*. 2004;41(3):227-231.
3. Petrova, G., Petrov, S., & Bancheva, S. Molecular characterization of Verbascum anisophyllum (Scrophulariaceae) genetic resources through inter-simple sequence repeat (ISSR) markers. *Nat Prod Commun*. 2017;12(2):1934578X1701200212
4. Abougazar, H., Bedir, E., Khan, I. A., Çalış, İ. Wiedemanniosides A-E: New phenylethanoid glycosides from the roots of Verbascum wiedemannianum. *Planta Med*. 2003;69(9):814-819.
5. Packer, L., Gibbs, J., Sheffield, C., Hanner, R. DNA barcoding and the mediocrity of morphology. *Mol Ecol Resour*. 2009; 9:42-50.
6. Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., Ma, X., ... Chen, S. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *J Ethnopharmacol*. 2010;130(1):116-121.
7. Chase, M. W., Cowan, R. S., Hollingsworth, P. M., Van Den Berg, C., Madriñán, S., Petersen, G., ... Wilkinson, M. A proposal for a standardised protocol to barcode all land plants. *Taxon*. 2007;56(2):295-299.
8. Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C., ... Husband, B. C. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers. *Mol Ecol Resour*. 2009; 9:130-139.
9. Nock, C. J., Waters, D. L., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., & Henry, R. J. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J.* 2011;9(3):328-333.
10. Christenhusz, M. J., Byng, J. W. The number of known plant species in the world and its annual increase. *Phytotaxa*. 2016;261(3):201-217.
11. Şekercioğlu ÇH, Anderson S, Akçay E, Bilgin R, Can ÖE, Semiz G, Tavşanoğlu Ç, Yokeş MB, Soyumert A, Ipekdal K, Sağlam İK. Turkey's globally important biodiversity in crisis. *Biological Conservation*. 2011 Dec 1;144(12):2752-69.

12. Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J., Fay, M. F., Byng, J. W., Judd, W. S., ... Stevens, P. F. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc*. 2016;181(1):1-20.
13. Demir, O., ÇINGAY, B., Evren, C. A. B. İ. Verbascum L.(Sığırkuyruğu) Cinsinde Eksik Olan Bilimsel Türkçe Adlar İçin Ad Önerisi. *Bağbahçe Bilim Dergisi*. 2021;8(2):38-42.
14. Yılmaz, Ö. Lectotypification of the Names Verbascum bombyciferum and V. vacillans (Scrophulariaceae). *Novon*. 2018;26(4):323-326.
15. Dong, X., Mkala, E. M., Mutinda, E. S., Yang, J. X., Wanga, V. O., Oulo, M. A., ... Wang, Q. F. Taxonomy, comparative genomics of Mullein (Verbascum, Scrophulariaceae), with implications for the evolution of Verbascum and Lamiales. *BMC Genomics*. 2022;23(1):1-22.
16. Kew Science. Verbascum L. [Internet]. [cited 2023 Jul 17]. Available from: https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:30049308-2/images
17. Kaur, V., Upadhyaya, K. Antibacterial activity of Verbascum chinense (Scrophulariaceae) extracts. *Int J Curr Microbiol App Sci*. 2016;5(4):578-584.
18. Ghahremaninejad, F., Riahi, M., Babaei, M., Attar, F., Behçet, L., Sonboli, A. Monophyly of Verbascum (Scrophularieae: Scrophulariaceae): evidence from nuclear and plastid phylogenetic analyses. *Aust J Bot*. 2014;62(8):638-646.
19. Riahi, M., Babaei, M., Ghahremaninejad, F. Genomic DNA isolation from Scrophularieae dried leaves using a simple, high-throughput protocol. *Bangladesh J Bot*. 2019;48(4):1231-1235.
20. Aytaç, Z., Duman, H. Verbascum hasbenlii (Scrophulariaceae), a new species from Turkey. *Turk J Bot*. 2012;36(4):322-327.
21. ÇEÇEN, Ö., KARAVELİOĞULLARI, F. A., ÜNAL, A. Verbascum misirdalianum (Scrophulariaceae), a new species from central Anatolia, Turkey. *Phytotaxa*. 2015;217(1):96-99.
22. Sotoodeh, A., Attar, F., Andalo, C., Mirtadzadini, M., Civeyrel, L. Focusing on three Verbascum L. taxa (Scrophulariaceae) of the Flora of Iran. *Adansonia*. 2018;40(2):171-181.

23. Sotoodeh, A., Attar, F., Civeyrel, L. A new species of Verbascum L. (Scrophulariaceae) from the Gilan province (Iran), based on morphological and molecular evidences. *Adansonia*. 2016;38(1):127-132.
24. Catara S, Cristaudo A, Gualtieri A, Galesi R, Impelluso C, Onofri A. Threshold temperatures for seed germination in nine species of Verbascum (Scrophulariaceae). *Seed Science Research.* 2016 Mar;26(1):30-46.
25. Karavelioğulları FA, Çelik S, Işık G. The Altitudinal, Climatical and Phenological Classification of Verbascum L. Species From Different Phytogeographical Regions of Turkey.
26. Frézal, L., Leblois, R. Four years of DNA barcoding: current advances and prospects. *Infect Genet Evol.* 2008;8(5):727-736.
27. Ali, M. A., Gyulai, G., Hidvegi, N., Kerti, B., Al Hemaid, F. M., Pandey, A. K., Lee, J. The changing epitome of species identification–DNA barcoding. *Saudi J Biol Sci*. 2014;21(3):204-231.
28. Badotti, F., de Oliveira, F. S., Garcia, C. F., Vaz, A. B. M., Fonseca, P. L. C., Nahum, L. A., ... Góes-Neto, A. Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiol.* 2017;17(1):1-12.
29. Irinyi, L., Serena, C., Garcia-Hermoso, D., Arabatzis, M., Desnos-Ollivier, M., Vu, D., ... Meyer, W. International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database—the quality controlled standard tool for routine identification of human and animal pathogenic fungi. *Med Mycol*. 2015;53(4):313-337.
30. Cowan, R. S., & Fay, M. F. Challenges in the DNA barcoding of plant material. In: *Plant DNA Fingerprinting and Barcoding*: *Methods and Protocols*. 2012:23-33.
31. KUMAR, S., YADAV, S., KADYAN, S., SINGH, K., SINGH, M., & Yadav, J. P. DNA Barcoding: A Novel Technique for Plants Taxonomists. *Annals of Agri-Bio Research*. 2012;17(2):139-143.
32. Kress WJ. Plant DNA barcodes: Applications today and in the future. *Journal of systematics and evolution*. 2017 Jul;55(4):291-307.
33. Fontaneto, D., Flot, J. F., Tang, C. Q. Guidelines for DNA taxonomy, with a focus on the meiofauna. *Mar Biodivers*. 2015; 45:433-451.

34. Galimberti, A., Casiraghi, M., Bruni, I., Guzzetti, L., Cortis, P., Berterame, N. M., Labra, M. From DNA barcoding to personalized nutrition: the evolution of food traceability. *Curr Opin Food Sci*. 2019; 28:41-48.
35. Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., ... Labra, M. DNA barcoding as a new tool for food traceability. *Food Res Int*. 2013;50(1):55-63.
36. Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., Francis, C. M. Identification of birds through DNA barcodes. *PLoS Biol.* 2004;2(10): e312.
37. Coyle, H. M., Lee, C., Lin, W., Lee, H. C., Palmbach, T. M. Forensic botany: using plant evidence to aid in forensic death investigation. *Croat Med J.* 2005;46(4):606.
38. Mildenhall, D. C., Wiltshire, P. E., Bryant, V. M. Forensic palynology: why do it and how it works. *Forensic Sci Int*. 2006;163(3):163-172.
39. Valentini, A., Pompanon, F., Taberlet, P. DNA barcoding for ecologists. *Trends Ecol Evol*. 2009;24(2):110-117.
40. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., Janzen, D. H. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci U S A*. 2005;102(23):8369-8374.
41. CBOL Plant Working Group. A DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2009;106(31):12794-12797.
42. Hebert, P. D., & Gregory, T. R. The promise of DNA barcoding for taxonomy. *Syst Biol*. 2005;54(5):852-859.
43. Hebert, P. D. N., Ratnasingham, S., De Waard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B Biol Sci*. 2003;270(suppl_1):S96-S99.
44. Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., ... Chen, S. Utility of the trnH–psbA intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. *PLoS One*. 2012;7(11):e48833.
45. Devey, D. S., Chase, M. W., Clarkson, J. J. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon*. 2009;58(1):7-15.
46. Cowan, R. S., Chase, M. W., Kress, W. J., Savolainen, V. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon*. 2006;55(3):611-616.

47. Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N., ... Savolainen, V. Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc B Biol Sci*. 2005;360(1462):1889-1895.
48. Newmaster, S. G., Fazekas, A. J., Ragupathy, S. D. N. A. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Botany*. 2006;84(3):335-341.
49. Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., Chen, S. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90(1):157-166.
50. Kress, W. J., Erickson, D. L. DNA barcodes: genes, genomics, and bioinformatics. *Proc Natl Acad Sci U S A*. 2008;105(8):2761-2762.
51. Kress, W. J., Erickson, D. L. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding trnH-psbA spacer region. *PLoS One*. 2007;2(6):e508.
52. Sucher, N. J., Hennell, J. R., Carles, M. C. DNA fingerprinting, DNA barcoding, and next-generation sequencing technology in plants. In: *Plant DNA Fingerprinting and Barcoding: Methods and Protocols*. 2012:13-22.
53. Sucher, N. J., Carles, M. C. Genome-based approaches to the authentication of medicinal plants. *Planta Med*. 2008;74(6):603-623.
54. Coissac, E., Hollingsworth, P. M., Lavergne, S., Taberlet, P. From barcodes to genomes: extending the concept of DNA barcoding.
55. Hollingsworth, P. M. Refining the DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2011;108(49):19451-19452.
56. Yu, J., XUE, J. H., ZHOU, S. L. New universal matK primers for DNA barcoding angiosperms. *J Syst Evol*. 2011;49(3):176-181.
57. Selvaraj, D., Sarma, R. K., Sathishkumar, R. Phylogenetic analysis of chloroplast matK gene from Zingiberaceae for plant DNA barcoding. *Bioinformation*. 2008;3(1):24.
58. Hollingsworth, P. M., Graham, S. W., Little, D. P. Choosing and using a plant DNA barcode. *PLoS One*. 2011;6(5):e19254.
59. Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., ... Barrett, S. C. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One.* 2008;3(7):e2802.

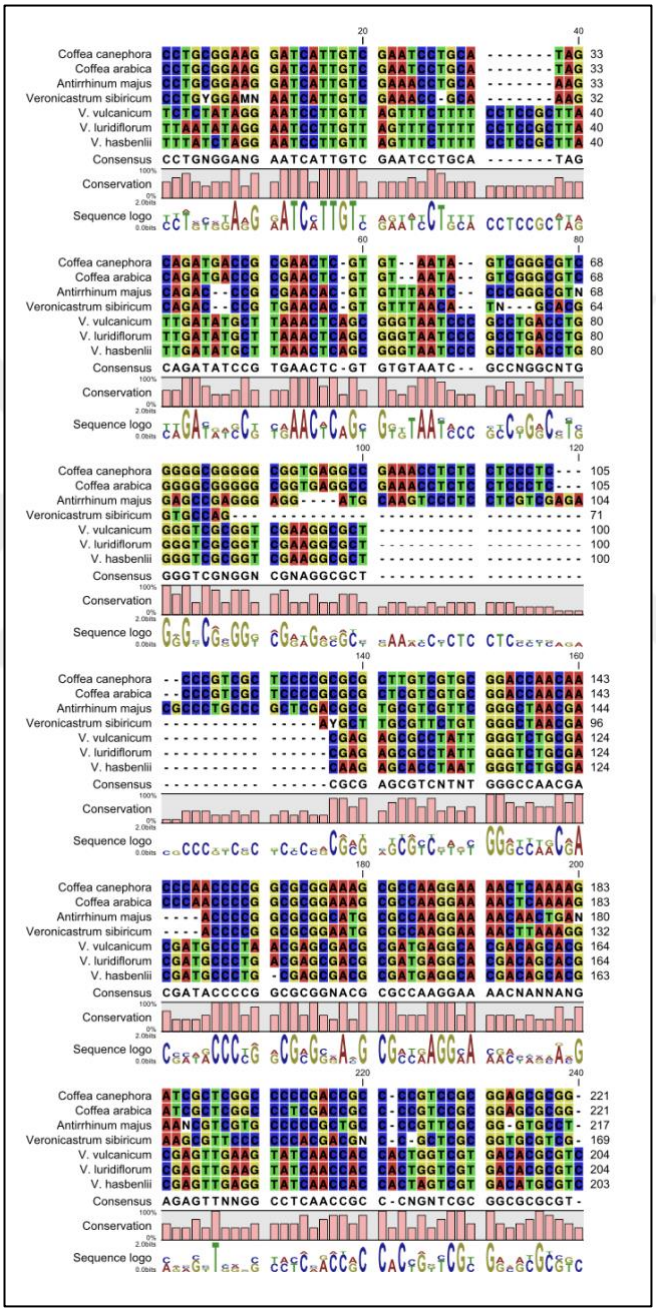
60. Sass, C., Little, D. P., Stevenson, D. W., Specht, C. D. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS One*. 2007;2(11):e1154.
61. Tripathi, A. M., Tyagi, A., Kumar, A., Singh, A., Singh, S., Chaudhary, L. B., Roy, S. The internal transcribed spacer (ITS) region and trnhH-psbA are suitable candidate loci for DNA barcoding of tropical tree species of India. *PLoS One*. 2013;8(2):e57934.
62. Stoeckle, M. Taxonomy, DNA, and the barcode of life. *Bioscience*. 2003;53(9):796-797.
63. Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., Newmaster, S. G., ... Barrett, S. C. Discriminating plant species in a local temperate flora using the rbcL+ matK DNA barcode. *Methods Ecol Evol*. 2011;2(4):333-340.
64. Van De Wiel, C. C. M., Van Der Schoot, J., Van Valkenburg, J. L. C. H., Duistermaat, H., Smulders, M. J. M. DNA barcoding discriminates the noxious invasive plant species, floating pennywort (Hydrocotyle ranunculoides Lf), from non-invasive relatives. *Mol Ecol Resour*. 2009;9(4):1086-1091.
65. Littlefair JE, Clare EL. Barcoding the food chain: from Sanger to high-throughput sequencing. *Genome*. 2016;59(11):946-958.
66. Men AE, Wilson P, Siemering K, Forrest S. Sanger DNA sequencing. *Next Generation Genome Sequencing: Towards Personalized Medicine*. 2008;1-11.
67. Valencia CA, Pervaiz MA, Husami A, Qian Y, Zhang K, Valencia CA, et al. Sanger sequencing principles, history, and landmarks. *Next Generation Sequencing Technologies in Medical Genetics*. 2013;3-11.
68. Reis-Filho JS. Next-generation sequencing. *Breast cancer research*. 2009;11(3):1-7.
69. Slatko BE, Gardner AF, Ausubel FM. *Overview of next-generation sequencing technologies*. Current protocols in molecular biology. 2018;122(1):e59.
70. Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular ecology resources*. 2014;14(5):892-901.
71. Möller M, Cronk QCB. Evolution of morphological novelty: a phylogenetic analysis of growth patterns in Streptocarpus (Gesneriaceae). *Evolution*. 2001;55:918–929.

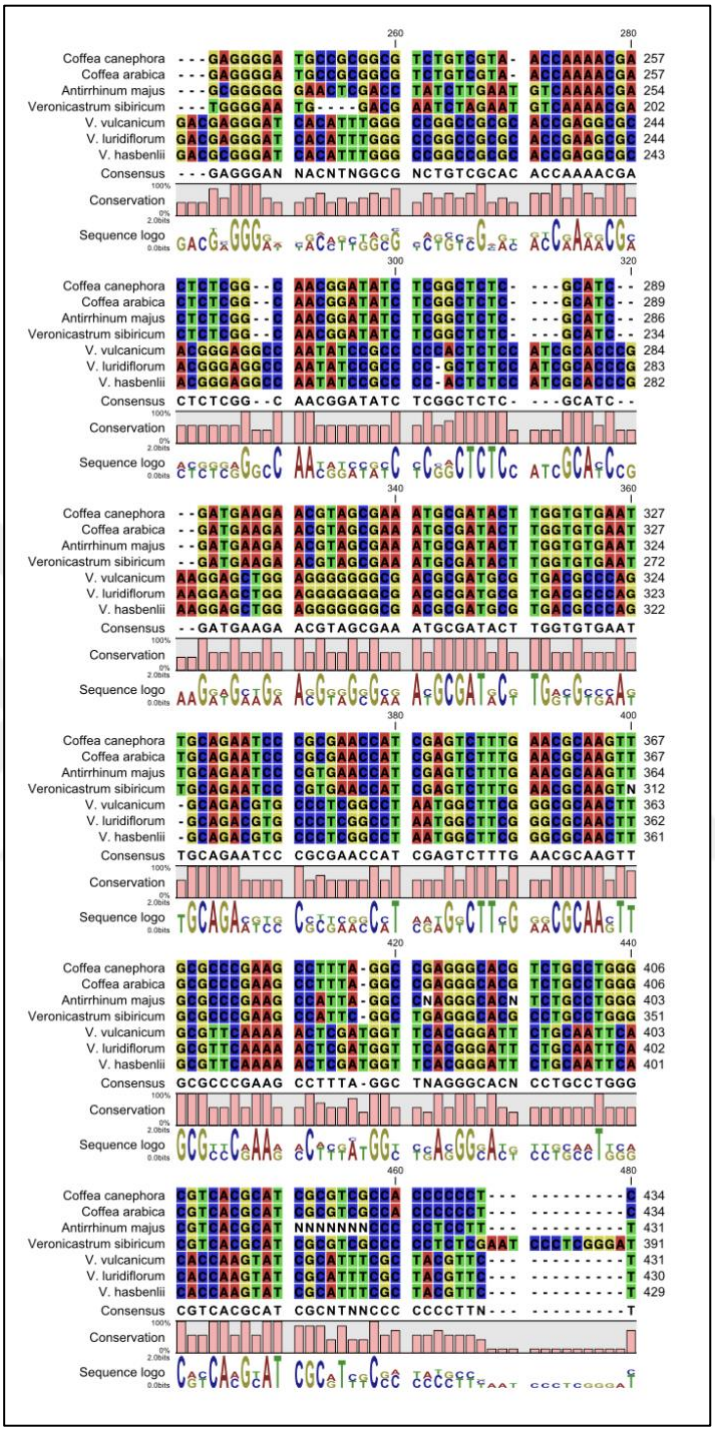
72. Kim YK, Jo SJ, Kim KJ. Phylogenetic position of Neofinetia and Sedirea (Orchidaceae) and their species identification using the chloroplast matK and the nuclear ITS sequences. *Korean Journal of Plant Taxonomy*. 2014;44(1):39-50.
73. Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. Plant DNA barcodes and a community phylogeny of a tropical forest dynamic plot in Panama. *Proceedings of the National Academy of Science USA*. 2009;106:18621–18626.
74. Tate JA, Simpson BB. Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany*. 2003;28:723–737.
75. Yang L, Abduraimov O, Tojibaev K, Shomurodov K, Zhang YM, Li WJ. Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of Ferula L. *BMC genomics*. 2022 Dec;23(1):1-4.
76. Beyra Matos A, Lavin M. Monograph of Pictetia (Leguminosae-Papilionoideae) and review of the Aeschynomeneae. (No Title). 1999.
77. Käss, E., & Wink, M. (1997). Phylogenetic relationships in the Papilionoideae (family Leguminosae) based on nucleotide sequences of cpDNA (rbcL) and ncDNA (ITS 1 and 2). *Molecular Phylogenetics and evolution*, 8(1), 65-88.
78. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden*. 1993 Jan 1:528-80.
79. Levin, R. A. et al. (2003). Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. *American Journal of Botany*, 90(1), 107-115.
80. Sang, T. et al. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *American Journal of Botany*, 84(8), 1120-1136.
81. Willis CG, Ellwood ER, Primack RB, Davis CC, Pearson KD, Gallinat AS, et al. Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology & Evolution.* 2017;32(7):531-546.
82. Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT. DNA damage in plant herbarium tissue. *PLoS One*. 2011;6(12):e28448.
83. Attitalla IH. Modified CTAB method for high quality genomic DNA extraction from medicinal plants. *Pak J Biol Sci*. 2011;14(21):998-999.

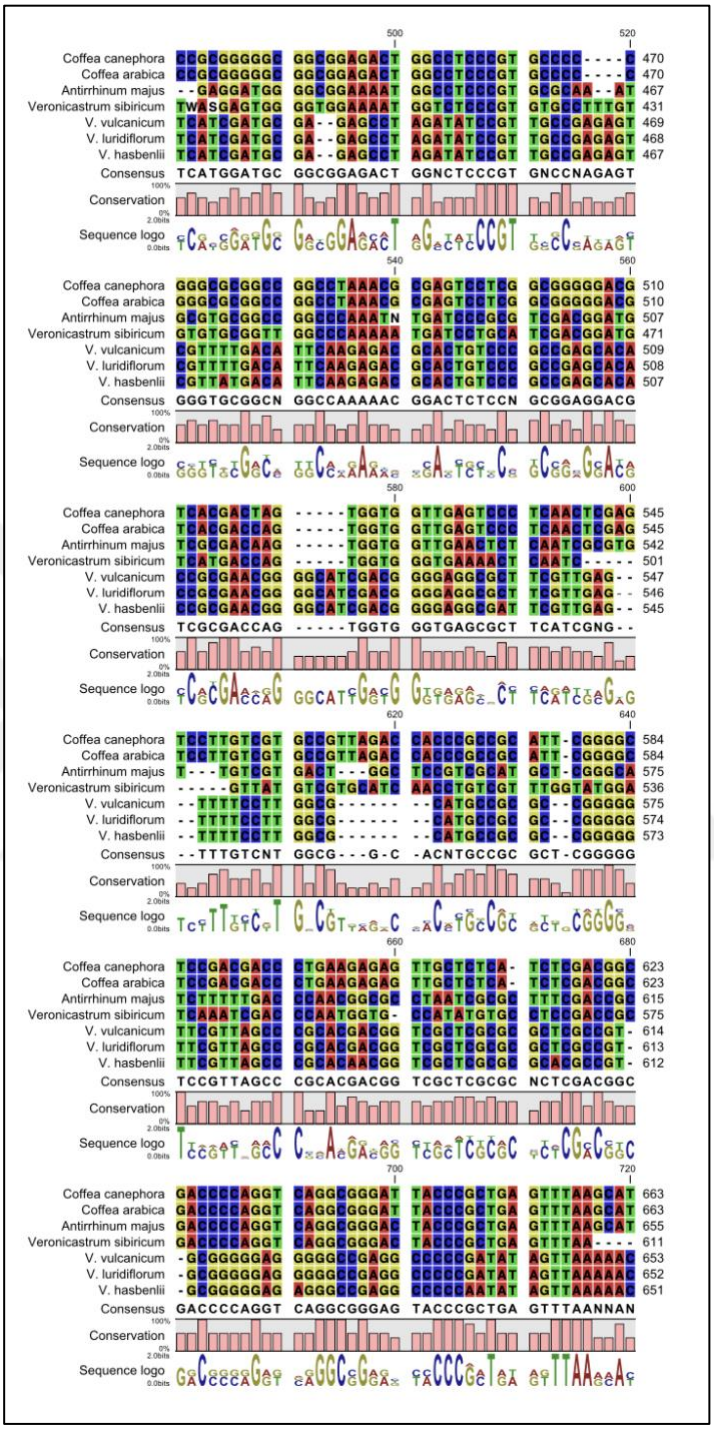
84. Zhang C, Xing D. Parallel DNA amplification by convective polymerase chain reaction with various annealing temperatures on a thermal gradient device. *Anal Biochem*. 2009;387(1):102-112.
85. Yu SL, Tang YQ, Li Y, Zhang H, Wu XL. Gradient decrement of annealing time can improve PCR with fluorescent-labeled primers. *J Biosci Bioeng*. 2010;110(4):500-504.
86. Kurt Y, Parmaksiz A, Kaya ÖF, Ulusal D, Cecen C. A simple and cost-effective protocol for DNA isolation from herbarium specimens. *InAnnales Botanici Fennici* 2022 Sep (Vol. 59, No. 1, pp. 251-259). Finnish Zoological and Botanical Publishing Board.
87. Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S. Use of ITS2 region as the universal DNA barcode for plants and animals. *PloS one*. 2010 Oct 1;5(10):e13102.
88. Enan MR, Palakkott AR, Ksiksi TS. DNA barcoding of selected UAE medicinal plant species: a comparative assessment of herbarium and fresh samples. *Physiology and molecular biology of plants.* 2017 Jan;23:221-7.
89. Li, F. W., Kuo, L. Y., Rothfels, C. J., Ebihara, A., Chiou, W. L., Windham, M. D., Pryer, K. M. rbcL and matK earn two thumbs up as the core DNA barcode for ferns. *PLoS One*. 2011;6(10):e26597.
90. Hilu, K. W., Barthet, M. M. Mode and tempo of matK: gene evolution and phylogenetic implications. In: *Evolutionary Biology from Concept to Application*. 2008:165-179.
91. Yang, R. H., Su, J. H., Shang, J. J., Wu, Y. Y., Li, Y., Bao, D. P., ... Yao, Y. J. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS One*. 2018;13(10):e0206428.
92. Chen, Y., Wang, B., Chen, J., Wang, X., Wang, R., Peng, S., ... Luo, J. Identification of Rubisco rbcL and rbcS in Camellia oleifera and their potential as molecular markers for selection of high tea oil cultivars. *Front Plant Sci*. 2015;6:189.
93. McLennan DA. How to read a phylogenetic tree. Evolution: *Education and Outreach*. 2010 Dec;3(4):506-19.
94. Charr JC, Garavito A, Guyeux C, Crouzillat D, Descombes P, Fournier C, Ly SN, Raharimalala EN, Rakotomalala JJ, Stoffelen P, Janssens S. Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on Coffea canephora (Robusta coffee). *Molecular phylogenetics and evolution*. 2020 Oct 1;151:106906.

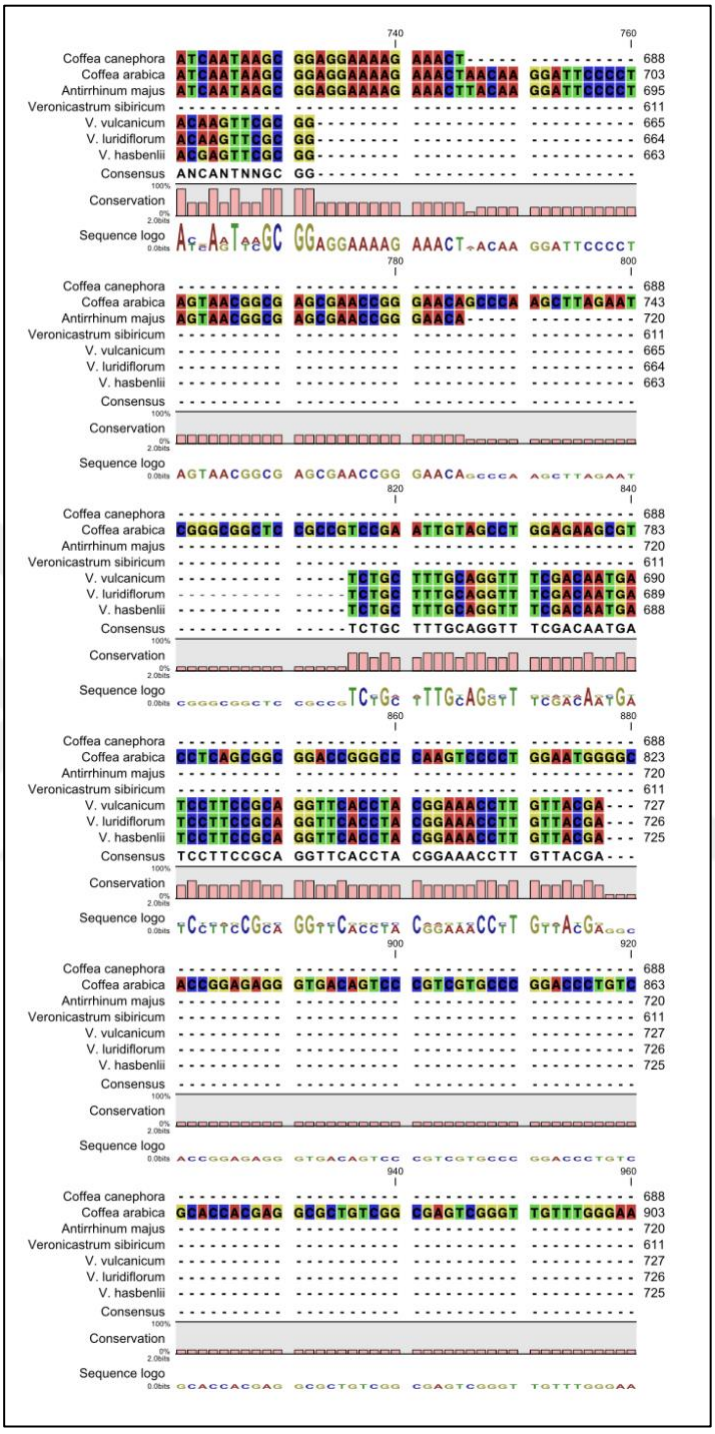
95. Lin Y, Hu F, Tang J, Moret BM. Maximum likelihood phylogenetic reconstruction from high-resolution whole-genome data and a tree of 68 eukaryotes. *InBiocomputing* 2013 2013 (pp. 285-296).
96. Russo CA, Selvatti AP. Bootstrap and rogue identification tests for phylogenetic analyses. *Molecular Biology and Evolution*. 2018 Sep 1;35(9):2327-33.
97. Maddison WP. Testing character correlation using pairwise comparisons on a phylogeny. *Journal of Theoretical Biology*. 2000 Feb 7;202(3):195-204.
98. Jain M. Next-generation sequencing technologies for gene expression profiling in plants. *Brief Funct Genomics.* 2012 Jan;11(1):63-70.
99. Barba M, Czosnek H, Hadidi A. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*. 2014 Jan;6(1):106-136.
100. Rosenberg MS, Kumar S. Traditional phylogenetic reconstruction methods reconstruct shallow and deep evolutionary relationships equally well. *Molecular Biology and Evolution*. 2001 Sep 1;18(9):1823-7.
101. Kuzmina ML, Braukmann TW, Fazekas AJ, Graham SW, Dewaard SL, Rodrigues A, Bennett BA, Dickinson TA, Saarela JM, Catling PM, Newmaster SG. Using herbarium-derived DNAs to assemble a large-scale DNA barcode library for the vascular plants of Canada. *Applications in Plant Sciences*. 2017 Dec;5(12):1700079.

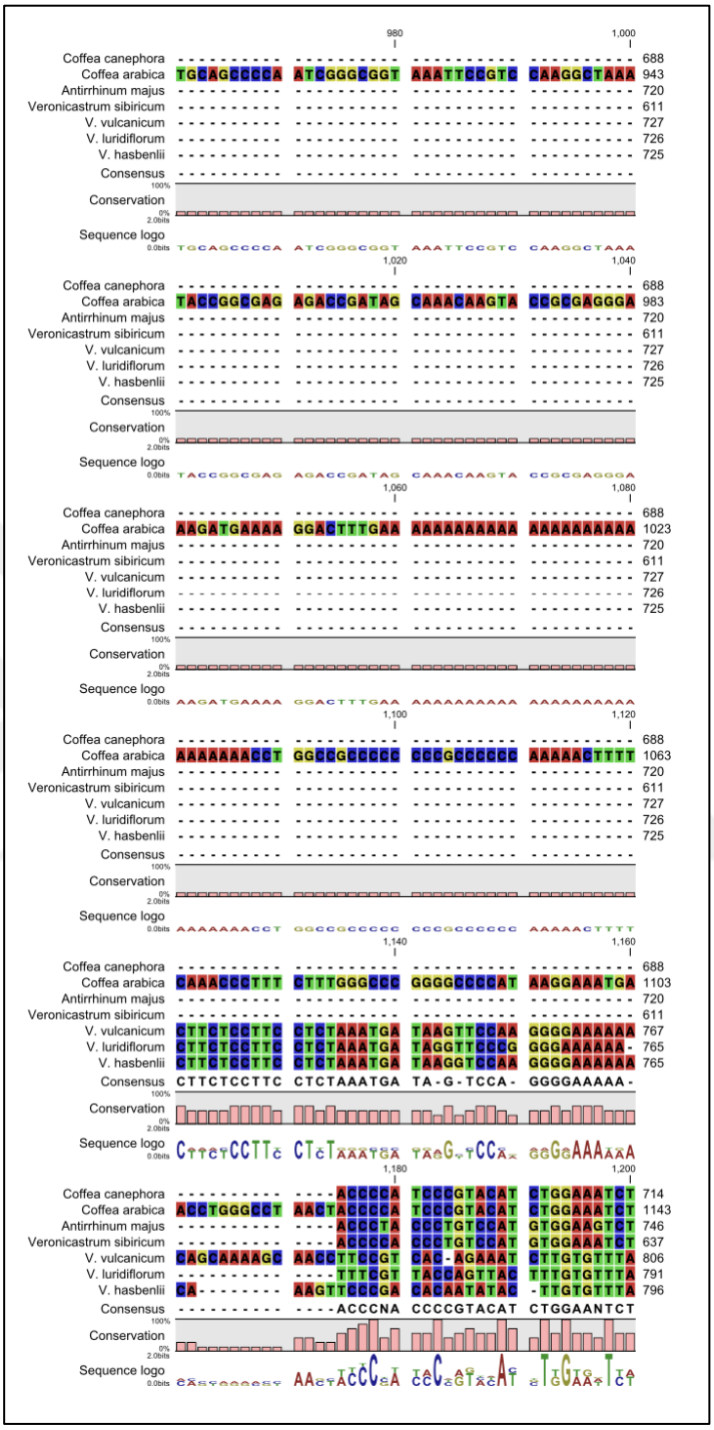
APPENDIX A: *ITS – matk* ASSEMBLED CHROMATOGRAM RESULTS OF SANGER SEQUENCED *VERBASCUM* L. SAMPLES

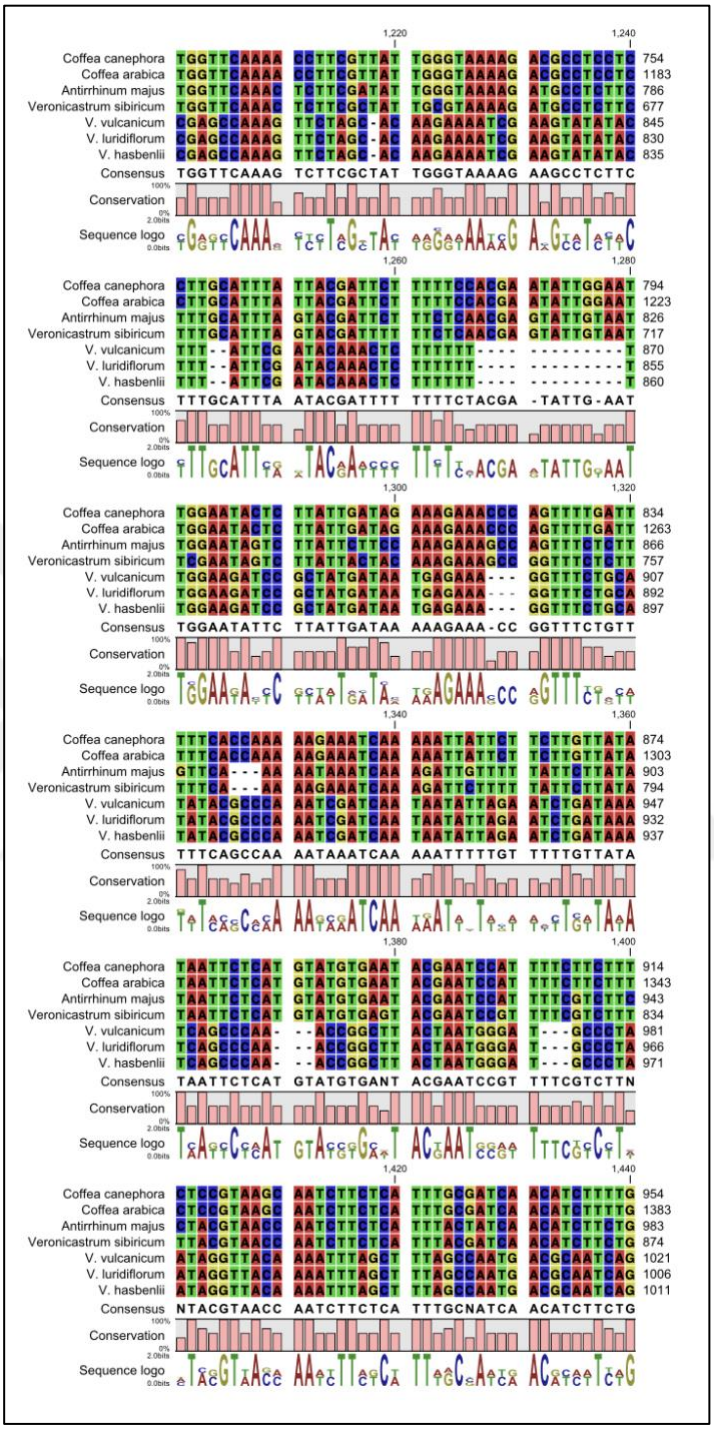


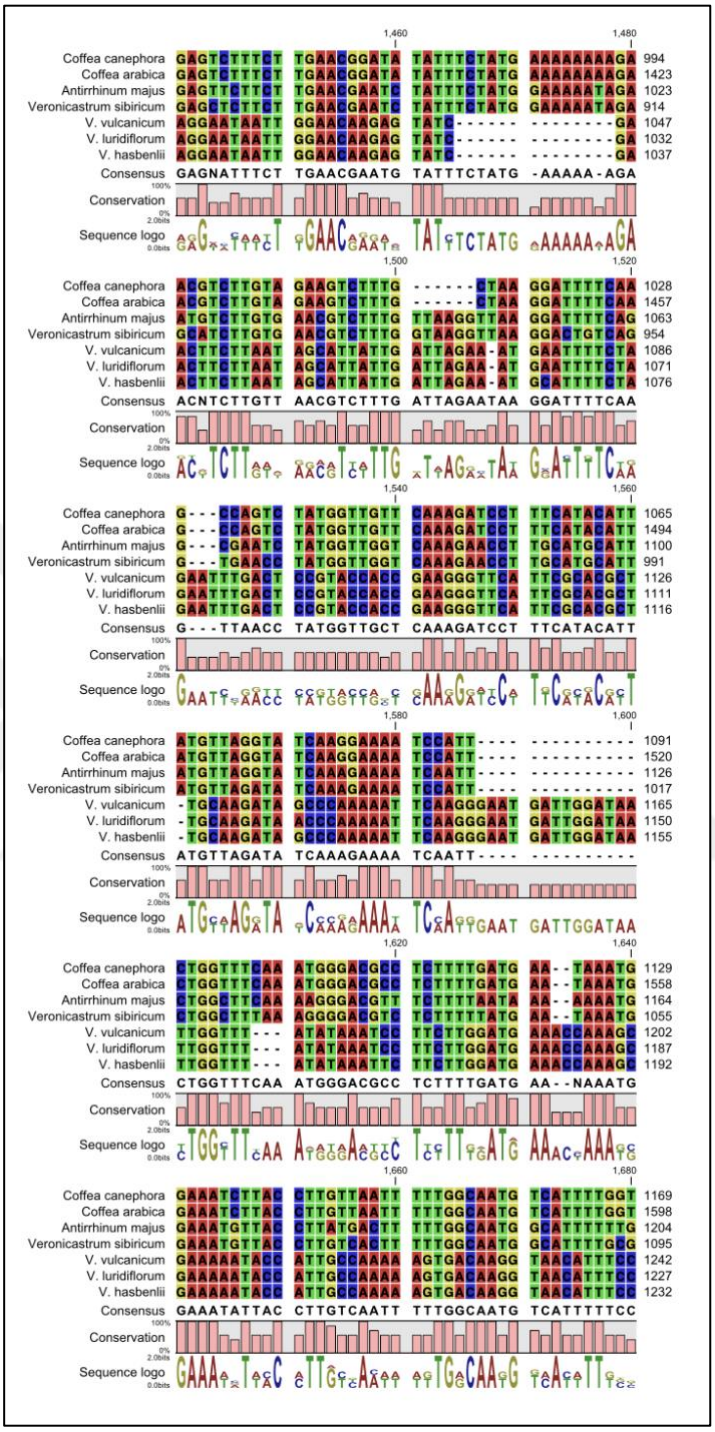

Coffea canephora
Coffea arabica
Antirrhinum majus
Veronicastrum sibiricum
V. vulcanicum
V. luridiflorum
V. hasbenlii
Consensus
Conservation
Sequence logo

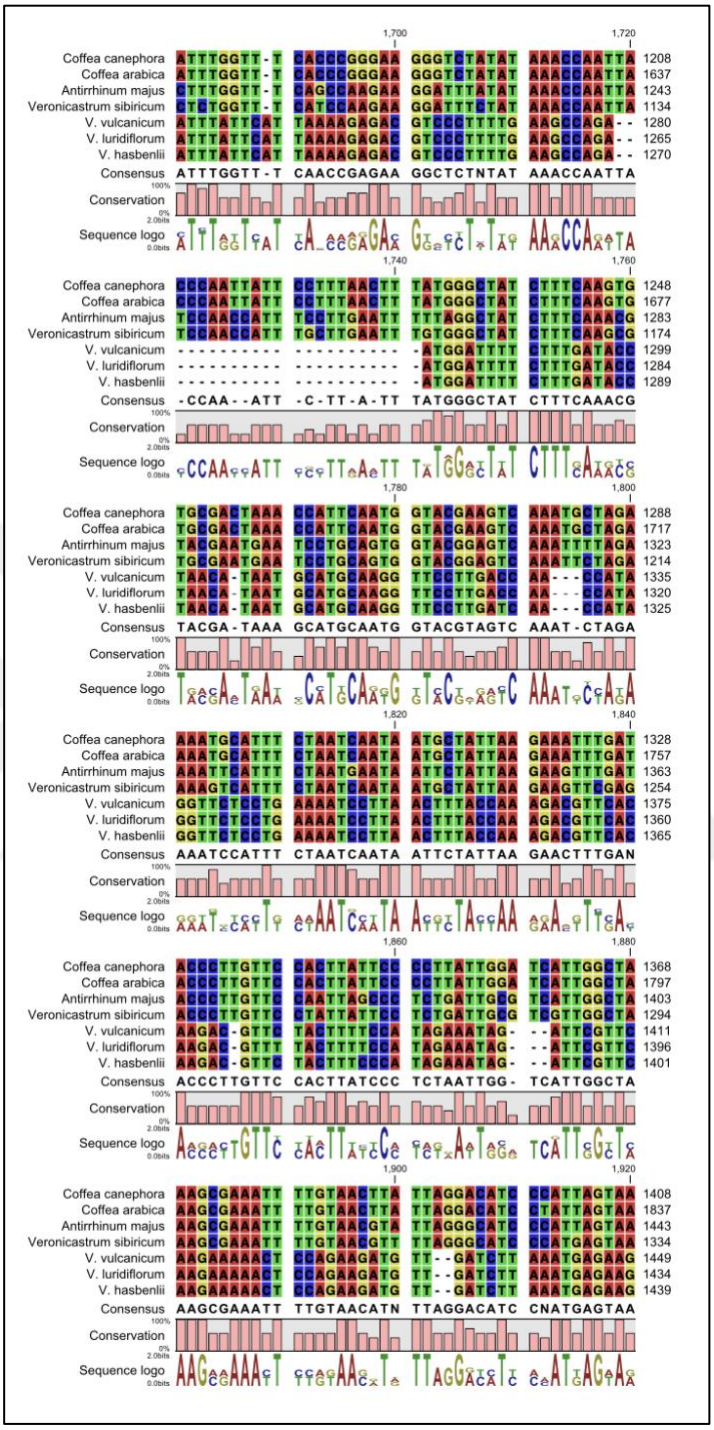

Coffea canephora
Coffea arabica
Antirrhinum majus
Veronicastrum sibiricum
V. vulcanicum
V. luridiflorum
V. hasbenlii
Consensus
Conservation
Sequence logo

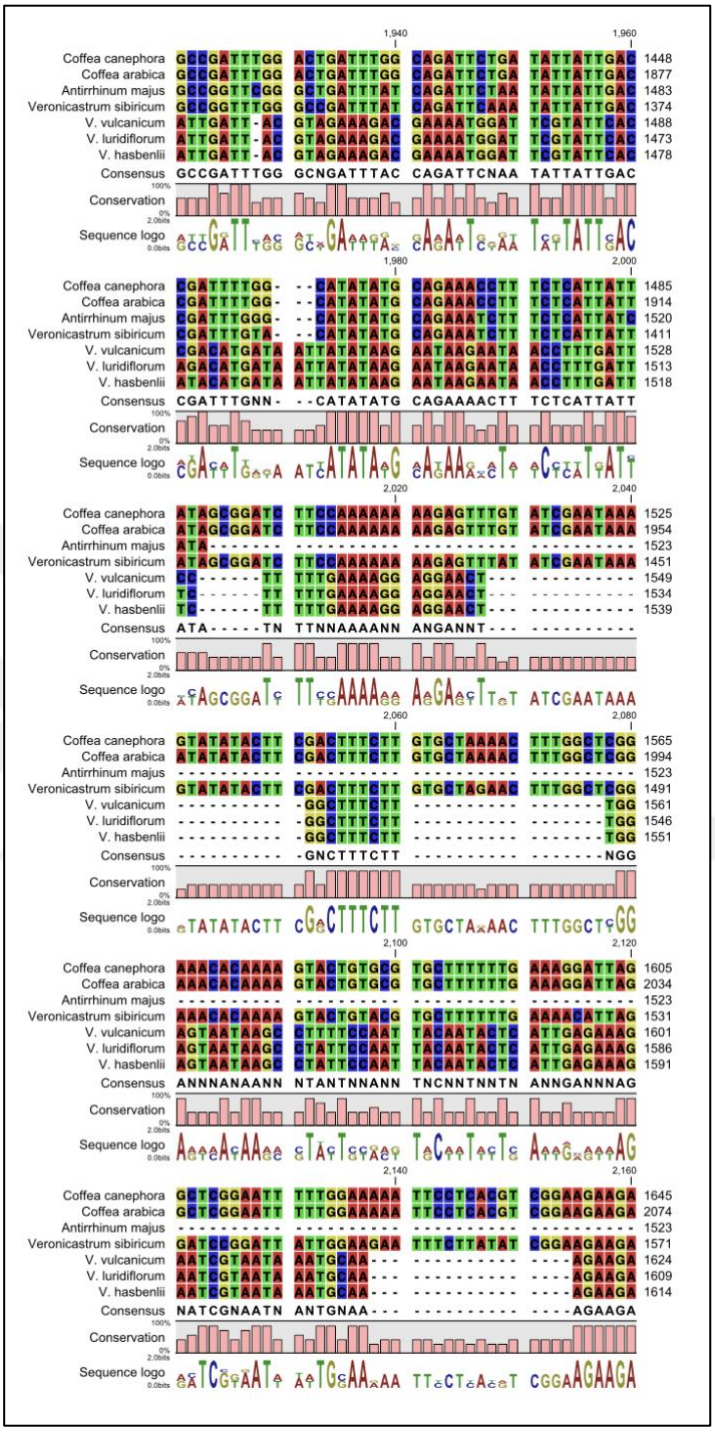

740
760
Coffea canephora ATCAATAAGC GGAGGAAAAG AAACT----- ---------- 688
Coffea arabica ATCAATAAGC GGAGGAAAAG AAACTAACAA GGATTCCCCT 703
Antirrhinum majus ATCAATAAGC GGAGGAAAAG AAACTTACAA GGATTCCCCT 695
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum ACAAGTTCGC GG-------- ---------- ---------- 665
V. luridiflorum ACAAGTTCGC GG-------- ---------- ---------- 664
V. hasbenlii ACGAGTTCGC GG-------- ---------- ---------- 663
Consensus ANCANTNNGC GG-------- ---------- ----------
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
780
800
Coffea canephora ---------- ---------- ---------- ---------- 688
Coffea arabica AGTAACGGCG AGCGAACCGG GAACAGCCCA AGCTTAGAAT 743
Antirrhinum majus AGTAACGGCG AGCGAACCGG GAACA----- ---------- 720
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum ---------- ---------- ---------- ---------- 665
V. luridiflorum ---------- ---------- ---------- ---------- 664
V. hasbenlii ---------- ---------- ---------- ---------- 663
Consensus ---------- ---------- ---------- ----------
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
820
840
Coffea canephora ---------- ---------- ---------- ---------- 688
Coffea arabica CGGGCGGCTC CGCCGTCCGA ATTGTAGCCT GGAGAAGCGT 783
Antirrhinum majus ---------- ---------- ---------- ---------- 720
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum ---------- -----TCTGC TTTGCAGGTT TCGACAATGA 690
V. luridiflorum ---------- -----TCTGC TTTGCAGGTT TCGACAATGA 689
V. hasbenlii ---------- -----TCTGC TTTGCAGGTT TCGACAATGA 688
Consensus ---------- -----TCTGC TTTGCAGGTT TCGACAATGA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
860
880
Coffea canephora ---------- ---------- ---------- ---------- 688
Coffea arabica CCTCAGCGGC GGACCGGGCC CAAGTCCCCT GGAATGGGGC 823
Antirrhinum majus ---------- ---------- ---------- ---------- 720
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum TCCTTCCGCA GGTTCACCTA CGGAAACCTT GTTACGA--- 727
V. luridiflorum TCCTTCCGCA GGTTCACCTA CGGAAACCTT GTTACGA--- 726
V. hasbenlii TCCTTCCGCA GGTTCACCTA CGGAAACCTT GTTACGA--- 725
Consensus TCCTTCCGCA GGTTCACCTA CGGAAACCTT GTTACGA---
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
900
920
Coffea canephora ---------- ---------- ---------- ---------- 688
Coffea arabica ACCGGAGAGG GTGACAGTCC CGTCGTGCCC GGACCCTGTC 863
Antirrhinum majus ---------- ---------- ---------- ---------- 720
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum ---------- ---------- ---------- ---------- 727
V. luridiflorum ---------- ---------- ---------- ---------- 726
V. hasbenlii ---------- ---------- ---------- ---------- 725
Consensus ---------- ---------- ---------- ----------
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
940
960
Coffea canephora ---------- ---------- ---------- ---------- 688
Coffea arabica GCACCACGAG GCGCTGTCGG CGAGTCGGGT TGTTTGGGAA 903
Antirrhinum majus ---------- ---------- ---------- ---------- 720
Veronicastrum sibiricum ---------- ---------- ---------- ---------- 611
V. vulcanicum ---------- ---------- ---------- ---------- 727
V. luridiflorum ---------- ---------- ---------- ---------- 726
V. hasbenlii ---------- ---------- ---------- ---------- 725
Consensus ---------- ---------- ---------- ----------
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

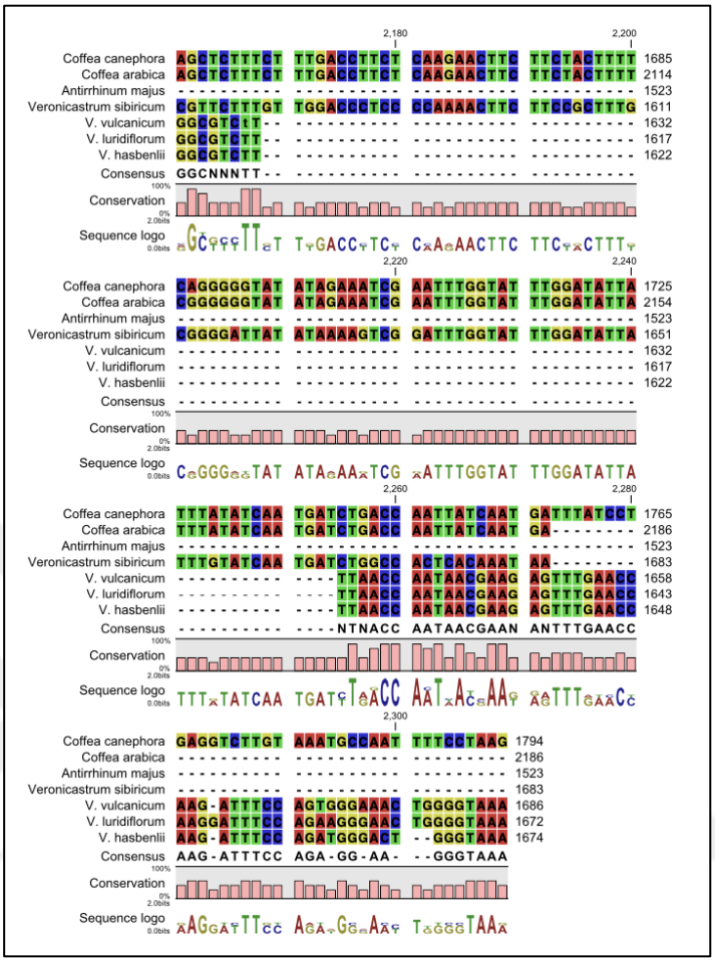

980
1,000
Coffea canephora
Coffea arabica
Antirrhinum majus
Veronicastrum sibiricum
V. vulcanicum
V. luridiflorum
V. hasbenlii
Consensus
Conservation
Sequence logo
1,020
1,040
1,060
1,080
1,100
1,120
1,140
1,160
1,180
1,200


1,220
1,240
Coffea canephora TGGTTCAAAA CCTTCGTTAT TGGGTAAAAG AAGCCTCTTC 754
Coffea arabica TGGTTCAAAA GCTTCGTTAT TGGGTAAAAG AAGCCTCTTC 1183
Antirrhinum majus TGGTTCAAAC TCTTCGATAT TGGGTAAAAG ATGCCTCTTC 786
Veronicastrum sibiricum TGGTTCAAAC TCTTCGCTAT TGCGTAAAAG ATGCCTCTTC 677
V. vulcanicum CGAGCCAAAG TTCTAGC-AC AAGAAAATCG AAGTATATAC 845
V. luridiflorum CGAGCCAAAG TTCTAGC-AC AAGAAAATCG AAGTATATAC 830
V. hasbenlii CGAGCCAAAG TTCTAGC-AC AAGAAAATCG AAGTATATAC 835
Consensus TGGTTCAAAG TCTTCGCTAT TGGGTAAAAG AAGCCTCTTC
Conservation
Sequence logo
1,260
1,280
Coffea canephora CTTGCATTTA TTACGATTCT TTTTCCACGA ATATTGGAAT 794
Coffea arabica CTTGCATTTA TTACGATTCT TTTTCCACGA ATATTGGAAT 1223
Antirrhinum majus TTTGCATTTA GTACGATTCT TTGTCAACGA GTATTGTAAT 826
Veronicastrum sibiricum TTTGCATTTA GTACGATTTT TTCTCAACGA GTATTGTAAT 717
V. vulcanicum TTT--ATTCG ATACAAACTC TTTTTT---- ---------T 870
V. luridiflorum TTT--ATTCG ATACAAACTC TTTTTT---- ---------T 855
V. hasbenlii TTT--ATTCG ATACAAACTC TTTTTT---- ---------T 860
Consensus TTTGCATTTA ATACGATTTT TTTTCTACGA -TATTG-AAT
Conservation
Sequence logo
1,300
1,320
Coffea canephora TGGAATACTC TTATTGATAG AAAGAAACCC AGTTTTGATT 834
Coffea arabica TGGAATACTC TTATTGATAG AAAGAAACCC AGTTTTGATT 1263
Antirrhinum majus TGGAATAGTC TTATTCTTCC AAAGAAAGCC AGTTTCTCTT 866
Veronicastrum sibiricum TGGAATAGTC TTATTACTAC AAAGAAAGCC GGTTTCTCTT 757
V. vulcanicum TGGAAGATCC GCTATGATAA TGAGAAA--- GGTTTCTGCA 907
V. luridiflorum TGGAAGATCC GCTATGATAA TGAGAAA--- GGTTTCTGCA 892
V. hasbenlii TGGAAGATCC GCTATGATAA TGAGAAA--- GGTTTCTGCA 897
Consensus TGGAATATTC TTATTGATAA AAAGAAA-CC GGTTTCTGTT
Conservation
Sequence logo
1,340
1,360
Coffea canephora TTTCAGCCAA AAGAAATCAA AAATTATTCT TCTTGTTATA 874
Coffea arabica TTTCAGCCAA AAGAAATCAA AAATTATTCT TCTTGTTATA 1303
Antirrhinum majus GTTCA---AA AATAAATCAA AGATTGTTTT TATTCTTATA 903
Veronicastrum sibiricum TTTCA---AA AAGAAATCAA AGATTCTTTT TATTCTTATA 794
V. vulcanicum TATACGCCCA AATCGATCAA TAATATTAGA ATCTGATAAA 947
V. luridiflorum TATACGCCCA AATCGATCAA TAATATTAGA ATCTGATAAA 932
V. hasbenlii TATACGCCCA AATCGATCAA TAATATTAGA ATCTGATAAA 937
Consensus TTTCAGCCAA AATAAATCAA AAATTTTTGT TTTTGTTATA
Conservation
Sequence logo
1,380
1,400
Coffea canephora TAATTCTCAT GTATGTGAAT ACGAATCCAT TTTCTTCTTT 914
Coffea arabica TAATTCTCAT GTATGTGAAT ACGAATCCAT TTTCTTCTTT 1343
Antirrhinum majus TAATTCTCAT GTATGTGAAT ACGAATCCAT TTTCGTCTTC 943
Veronicastrum sibiricum TAATTCTCAT GTATGTGAGT ACGAATCCGT TTTCGTCTTT 834
V. vulcanicum TCAGCCCAA- --ACCGGCTT ACTAATGGGA T---GCCCTA 981
V. luridiflorum TCAGCCCAA- --ACCGGCTT ACTAATGGGA T---GCCCTA 966
V. hasbenlii TCAGCCCAA- --ACCGGCTT ACTAATGGGA T---GCCCTA 971
Consensus TAATTCTCAT GTATGTGANT ACGAATCCGT TTTCGTCTTN
Conservation
Sequence logo
1,420
1,440
Coffea canephora CTCCGTAAGC AATCTTCTCA TTTGCGATCA ACATCTTTTG 954
Coffea arabica CTCGGTAAGC AATCTTCTCA TTTGCGATCA ACATCTTTTG 1383
Antirrhinum majus CTACGTAACC AATCTTCTCA TTTACTATCA ACATCTTCTG 983
Veronicastrum sibiricum TTAGGTAACC AATCTTCTCA TTTACGATCA ACATCTTCTG 874
V. vulcanicum ATAGGTTACA AAATTTAGCT TTAGCCAATG AGCAATCAG 1021
V. luridiflorum ATAGGTTACA AAATTTAGCT TTAGCCAATG AGCAATCAG 1006
V. hasbenlii ATAGGTTACA AAATTTAGCT TTAGCCAATG AGCAATCAG 1011
Consensus NTACGTAACC AATCTTCTCA TTTGCNATCA ACATCTTCTG
Conservation
Sequence logo

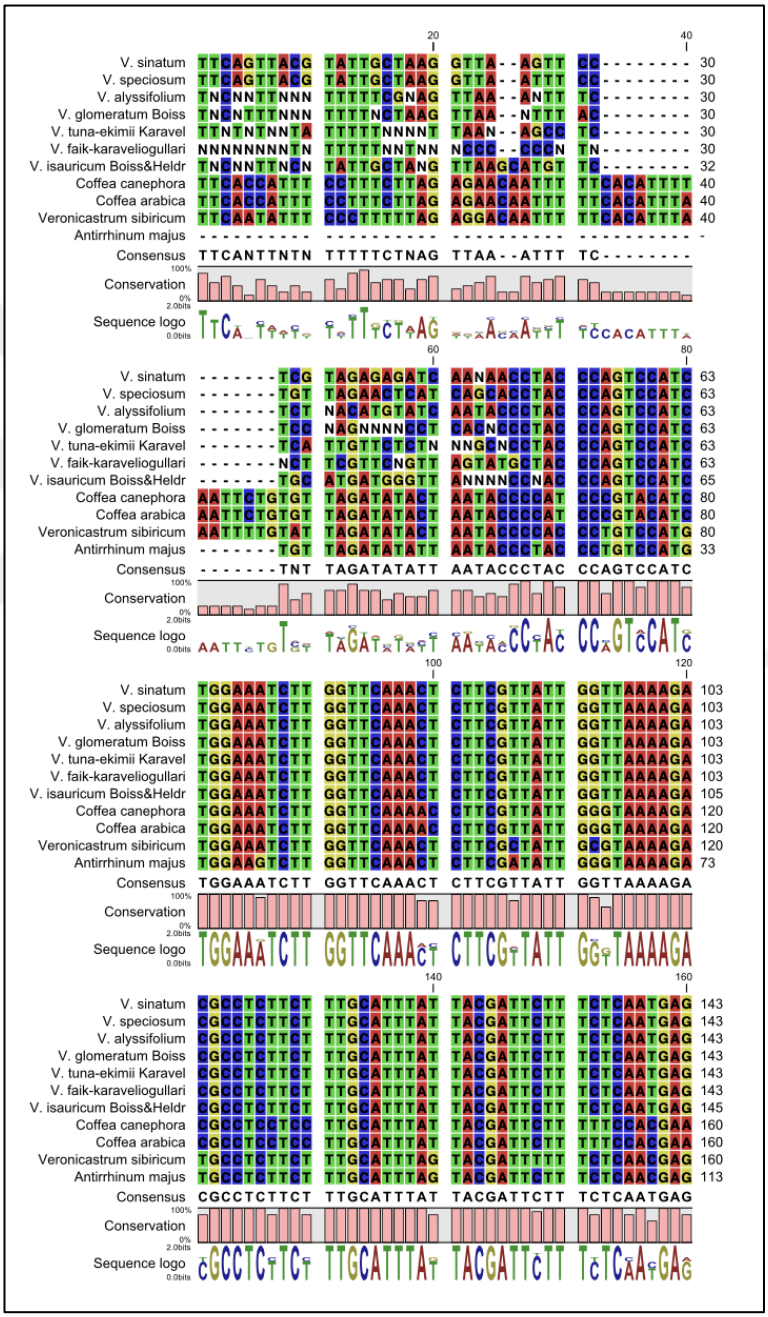

1,460
1,480
Coffea canephora GAGTCTTTCT TGAACGGATA TATTTCTATG AAAAAAAAGA 994
Coffea arabica GAGTCTTTCT TGAACGGATA TATTTCTATG AAAAAAAAGA 1423
Antirrhinum majus GAGTTCTTCT TGAACGAATC TATTTCTATG GAAAAATAGA 1023
Veronicastrum sibiricum GAGCTCTTCT TGAACGAATC TATTTCTATG GAAAAATAGA 914
V. vulcanicum AGGAATAATT GGAACAAGAG TATC------ --------GA 1047
V. luridiflorum AGGAATAATT GGAACAAGAG TATC------ --------GA 1032
V. hasbenlii AGGAATAATT GGAACAAGAG TATC------ --------GA 1037
Consensus GAGNATTTCT TGAACGAATG TATTTCTATG -AAAAA-AGA
Conservation
Sequence logo
1,500
1,520
Coffea canephora ACGTCTTGTA GAAGTCTTTG ------CTAA GGATTTTCAA 1028
Coffea arabica ACGTCTTGTA GAAGTCTTTG ------CTAA GGATTTTCAA 1457
Antirrhinum majus ATGTCTTGTG AACGTCTTTG TTAAGGTTAA GGATTTTCAG 1063
Veronicastrum sibiricum GCATCTTGTG AACGTCTTTG GTAAGGTTAA GGACTGTCAG 954
V. vulcanicum ACTTCTTAAT AGCATTATTG ATTAGAA-AT GAATTTTCTA 1086
V. luridiflorum ACTTCTTAAT AGCATTATTG ATTAGAA-AT GAATTTTCTA 1071
V. hasbenlii ACTTCTTAAT AGCATTATTG ATTAGAA-AT GCATTTTCTA 1076
Consensus ACNTCTTGTT AACGTCTTTG ATTAGAATAA GGATTTTCAA
Conservation
Sequence logo
1,540
1,560
Coffea canephora G---CCAGTC TATGGTTGTT CAAAGATCCT TTCATACATT 1065
Coffea arabica G---CCAGTC TATGGTTGTT CAAAGATCCT TTCATACATT 1494
Antirrhinum majus G---CGAATC TATGGTTGGT CAAAGAACCT TGCATGCATT 1100
Veronicastrum sibiricum G---TGAACC TATGGTTGGT CAAAGAACCT TGCATGCATT 991
V. vulcanicum GAATTTGACT CCGTACCACC GAAGGGTTCA TTCGCACGCT 1126
V. luridiflorum GAATTTGACT CCGTACCACC GAAGGGTTCA TTCGCACGCT 1111
V. hasbenlii GAATTTGACT CCGTACCACC GAAGGGTTCA TTCGCACGCT 1116
Consensus G---TTAACC TATGGTTGCT CAAAGATCCT TTCATACATT
Conservation
Sequence logo
1,580
1,600
Coffea canephora ATGTTAGGTA TCAAGGAAAA TCCATT---- ---------- 1091
Coffea arabica ATGTTAGGTA TCAAGGAAAA TCCATT---- ---------- 1520
Antirrhinum majus ATGTTAGGTA TCAAAGAAAA TCAATT---- ---------- 1126
Veronicastrum sibiricum ATGTTAGATA TCAAAGAAAA TCCATT---- ---------- 1017
V. vulcanicum -TGCAAGATA GCCCAAAAAT TCAAGGGAAT GATTGGATAA 1165
V. luridiflorum -TGCAAGATA ACCCAAAAAT TCAAGGGAAT GATTGGATAA 1150
V. hasbenlii -TGCAAGATA GCCCAAAAAT TCAAGGGAAT GATTGGATAA 1155
Consensus ATGTTAGATA TCAAAGAAAA TCAATT---- ----------
Conservation
Sequence logo
1,620
1,640
Coffea canephora CTGGTTTCAA ATGGGACGCC TCTTTTGATG AA--TAAATG 1129
Coffea arabica CTGGTTTCAA ATGGGACGCC TCTTTTGATG AA--TAAATG 1558
Antirrhinum majus CTGGCTTCAA AAGGGACGTT TCTTTTAATA AA--AAAATG 1164
Veronicastrum sibiricum CTGGCTTTAA AGGGGACGTC TCTTTTTATG AA--TAAATG 1055
V. vulcanicum TTGGTTT--- ATATAAATCC TTCTTGGATG AAACCAAAGC 1202
V. luridiflorum TTGGTTT--- ATATAAATCC TTCTTGGATG AAACCAAAGC 1187
V. hasbenlii TTGGTTT--- ATATAAATTC TTCTTGGATG AAACCAAAGC 1192
Consensus CTGGTTTCAA ATGGGACGCC TCTTTTGATG AA--NAAATG
Conservation
Sequence logo
1,660
1,680
Coffea canephora GAAATCTTAC CTTGTTAATT TTTGGCAATG TCATTTTGGT 1169
Coffea arabica GAAATCTTAC GTTGTTAATT TTTGGCAATG TCATTTTGGT 1598
Antirrhinum majus GAAATGTTAC GTTATGAATT TTTGGCAATG GCATTTTTTG 1204
Veronicastrum sibiricum GAAATGTTAC CTTGTCAATT TTTGGCAATG GCATTTTGCG 1095
V. vulcanicum GAAAAATACC ATTGCCAAAA AGTGACAAGG TAACATTTCC 1242
V. luridiflorum GAAAAATACC ATTGCCAAAA AGTGACAAGG TAACATTTCC 1227
V. hasbenlii GAAAAATACC ATTGCCAAAA AGTGACAAGG TAACATTTCC 1232
Consensus GAAATATTAC CTTGTCAATT TTTGGCAATG TCATTTTTCC
Conservation
Sequence logo

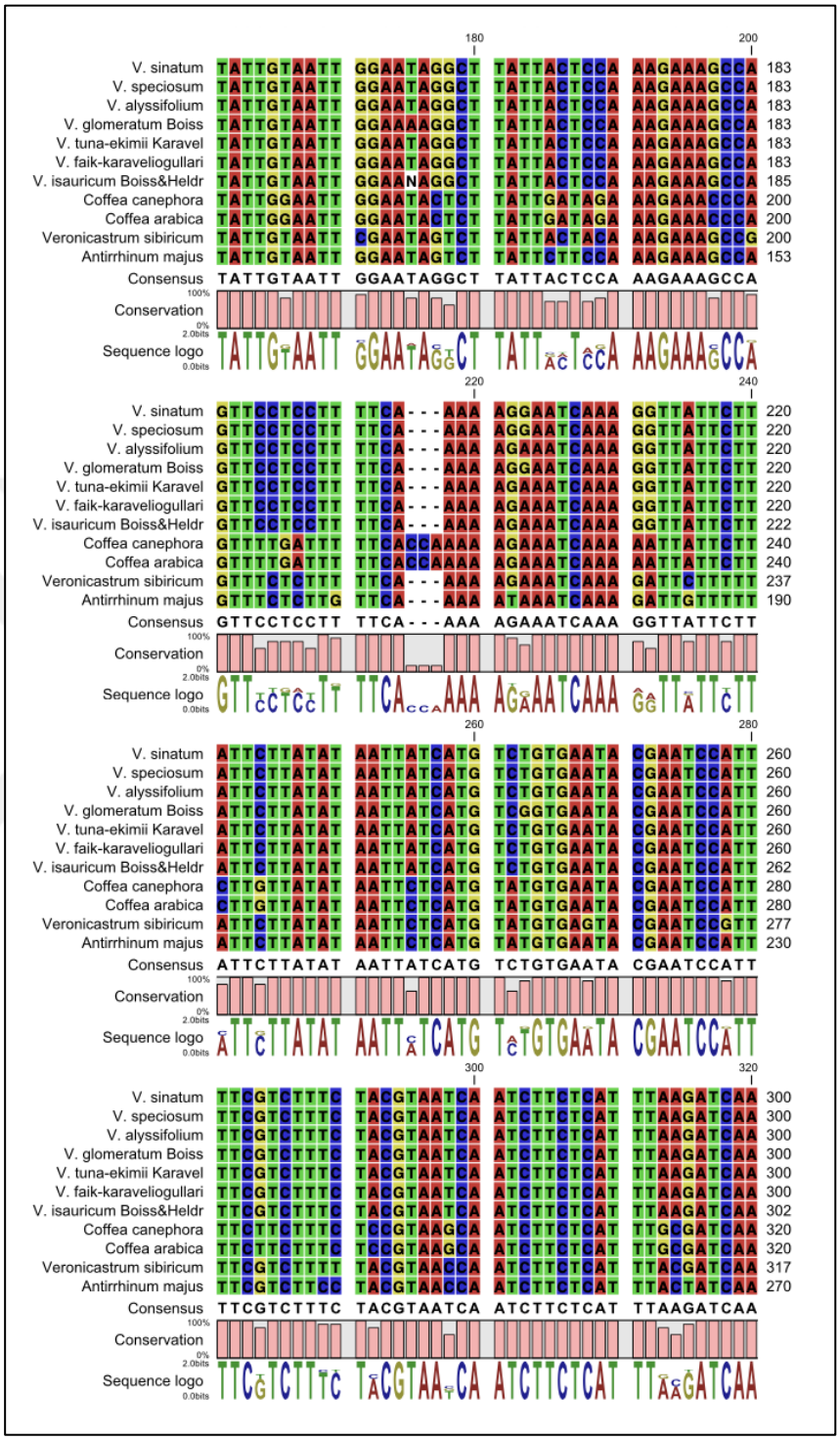

1,700
1,720
Coffea canephora ATTTGGTT-T CACCCGGGAA GGGTCTATAT AAACCAATTA 1208
Coffea arabica ATTTGGTT-T CACCCGGGAA GGGTCTATAT AAACCAATTA 1637
Antirrhinum majus CTTTGGTT-T CAGCCAAGAA GGATTTATAT AAACCAATTA 1243
Veronicastrum sibiricum CTCTGGTT-T CATCCAAGAA GGATTTCTAT AAACCAATTA 1134
V. vulcanicum ATTTATTCAT TAAAAGAGAC GTCCCTTTTG AAGCCAGA-- 1280
V. luridiflorum ATTTATTCAT TAAAAGAGAC GTCCCTTTTG AAGCCAGA-- 1265
V. hasbenlii ATTTATTCAT TAAAAGAGAC GTCCCTTTTG AAGCCAGA-- 1270
Consensus ATTTGGTT-T CAACCGAGAA GGCTCTNTAT AAACCAATTA
Conservation
Sequence logo
1,740
1,760
Coffea canephora CCCAATTATT CCTTTAACTT TATGGGCTAT CTTTCAAGTG 1248
Coffea arabica CCCAATTATT CCTTTAACTT TATGGGCTAT CTTTCAAGTG 1677
Antirrhinum majus TCCAACCATT TCCTTGAATT TTTAGGCTAT CTTTCAAACG 1283
Veronicastrum sibiricum TCCAACCATT TGCTTGAATT TGTGGGCTAT CTTTCAAGCG 1174
V. vulcanicum ---------- ---------- -ATGGATTTT CTTTGATACC 1299
V. luridiflorum ---------- ---------- -ATGGATTTT CTTTGATACC 1284
V. hasbenlii ---------- ---------- -ATGGATTTT CTTTGATACC 1289
Consensus -CCAA--ATT -C-TT-A-TT TATGGGCTAT CTTTCAAACG
Conservation
Sequence logo
1,780
1,800
Coffea canephora TGCGACTAAA CCATTCAATG GTACGAAGTC AAATGCTAGA 1288
Coffea arabica TGCGACTAAA CCATTCAATG GTACGAAGTC AAATGCTAGA 1717
Antirrhinum majus TACGAATGAA TCCTGCAGTG GTACGGAGTC AAATTTTAGA 1323
Veronicastrum sibiricum TGCGAATGAA TCCTGCAGTG GTACGGAGTC AAATTCTAGA 1214
V. vulcanicum TAACA-TAAT GCATGCAAGG TTCCTTGACC AA----CCATA 1335
V. luridiflorum TAACA-TAAT GCATGCAAGG TTCCTTGACC AA----CCATA 1320
V. hasbenlii TAACA-TAAT GCATGCAAGG TTCCTTGATC AA----CCATA 1325
Consensus TACGA-TAAA GCATGCAATG GTACGTAGTC AAAT-CTAGA
Conservation
Sequence logo
1,820
1,840
Coffea canephora AAATGCATTT CTAATCAATA ATGCTATTAA GAAATTTGAT 1328
Coffea arabica AAATGCATTT CTAATCAATA ATGCTATTAA GAAATTTGAT 1757
Antirrhinum majus AAATTCATTT CTAATGAATA ATTCTATTAA GAAGTTTGAT 1363
Veronicastrum sibiricum AAAGTCATTT CTAATCAATA ATGCTATTAA GAAGTTCGAG 1254
V. vulcanicum GGTTCTCGTG AAAATCCTTA ACTTTACCAA AGACGTTCAC 1375
V. luridiflorum GGTTCTCGTG AAAATCCTTA ACTTTACCAA AGACGTTCAC 1360
V. hasbenlii GGTTCTCGTG AAAATCCTTA ACTTTACCAA AGACGTTCAC 1365
Consensus AAATCCATTT CTAATCAATA ATTCTATTAA GAACTTTGAN
Conservation
Sequence logo
1,860
1,880
Coffea canephora ACCCTTGTTC CACTTATCCC TCTTATTGGA TCATTGGCTA 1368
Coffea arabica ACCCTTGTTC CACTTATTCC GCTTATTGGA TCATTGGCTA 1797
Antirrhinum majus ACCCTTGTTC CAATTAGCCC TCTGATTGCG TCATTGGCTA 1403
Veronicastrum sibiricum ACCCTTGTTC CTATTATTCC TCTGATTGCG TCGTTGGCTA 1294
V. vulcanicum AAGAC-GTTC TACTTTTCCA TAGAAATAG- --ATTCGTTC 1411
V. luridiflorum AAGAC-GTTT TACTTTTCCA TAGAAATAG- --ATTCGTTC 1396
V. hasbenlii AAGAC-GTTC TACTTTCCCA TAGAAATAG- --ATTCGTTC 1401
Consensus ACCCTTGTTC CACTTATCCC TCTAATTGG- TCATTGGCTA
Conservation
Sequence logo
1,900
1,920
Coffea canephora AAGCGAAATT TTGTAACTTA TTAGGACATC CCATTAGTAA 1408
Coffea arabica AAGCGAAATT TTGTAACTTA TTAGGACATC CTATTAGTAA 1837
Antirrhinum majus AAGCGAAATT TTGTAACGTA TTAGGGCATC CCATTAGTAA 1443
Veronicastrum sibiricum AAGCGAAATT TTGTAACGTT TTAGGGCATC CCATGAGTAA 1334
V. vulcanicum AAGAAAAACT CCAGAAGATG TT--GATCTT AAATGAGAAG 1449
V. luridiflorum AAGAAAAACT CCAGAAGATG TT--GATCTT AAATGAGAAG 1434
V. hasbenlii AAGAAAAACT CCAGAAGATG TT--GATCTT AAATGAGAAG 1439
Consensus AAGCGAAATT TTGTAACATN TTAGGACATC CNATGAGTAA
Conservation
Sequence logo

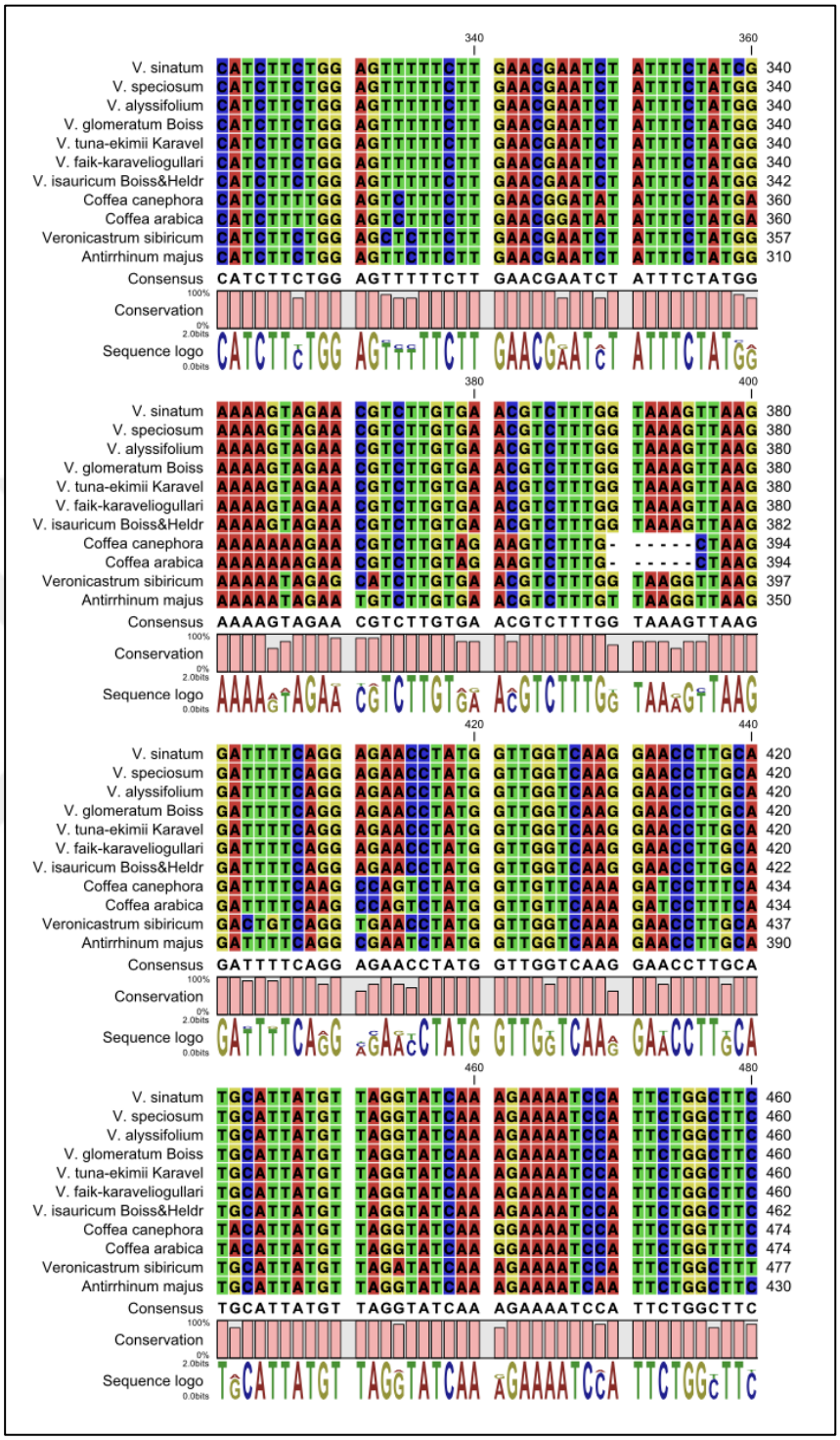

1,940
1,960
Coffea canephora GCCGATTTGG ACTGATTTGG CAGATTCTGA TATTATTGAC 1448
Coffea arabica GCCGATTTGG ACTGATTTGG CAGATTCTGA TATTATTGAC 1877
Antirrhinum majus GCCGGTTCGG GCTGATTTAT CAGATTCTAA TATTATTGAC 1483
Veronicastrum sibiricum GCCGGTTTGG GCCGATTTAT CAGATTCAAA TATTATTGAC 1374
V. vulcanicum ATTGATT-AC GTAGAAAGAC GAAAATGGAT TCGTATTCAC 1488
V. luridiflorum ATTGATT-AC GTAGAAAGAC GAAAATGGAT TCGTATTCAC 1473
V. hasbenlii ATTGATT-AC GTAGAAAGAC GAAAATGGAT TCGTATTCAC 1478
Consensus GCCGATTTGG GCNGATTTAC CAGATTCNAA TATTATTGAC
Conservation
Sequence logo
1,980
2,000
Coffea canephora CGATTTTGG- --CATATATG CAGAAACCTT TCTCATTATT 1485
Coffea arabica CGATTTTGG- --CATATATG CAGAAACCTT TCTCATTATT 1914
Antirrhinum majus CGATTTGGG- --CATATATG CAGAAATCTT TCTCATTATC 1520
Veronicastrum sibiricum CGATTTGTA- --CATATATG CAGAAATCTT TCTCATTATT 1411
V. vulcanicum CGACATGATA ATTATATAAG AATAAGAATA ACCTTTGATT 1528
V. luridiflorum AGACATGATA ATTATATAAG AATAAGAATA ACCTTTGATT 1513
V. hasbenlii ATACATGATA ATTATATAAG AATAAGAATA ACCTTTGATT 1518
Consensus CGATTTGNN- --CATATATG CAGAAAACTT TCTCATTATT
Conservation
Sequence logo
2,020
2,040
Coffea canephora ATAGCGGATC TTCCAAAAAA AAGAGTTTGT ATCGAATAAA 1525
Coffea arabica ATAGCGGATC TTCCAAAAAA AAGAGTTTGT ATCGAATAAA 1954
Antirrhinum majus ATA------- ---------- ---------- ---------- 1523
Veronicastrum sibiricum ATAGCGGATC TTCCAAAAAA AAGAGTTTAT ATCGAATAAA 1451
V. vulcanicum CC------TT TTTGAAAAGG AGGAACT--- ---------- 1549
V. luridiflorum TC------TT TTTGAAAAGG AGGAACT--- ---------- 1534
V. hasbenlii TC------TT TTTGAAAAGG AGGAACT--- ---------- 1539
Consensus ATA-----TN TTNNAAAANN ANGANNT--- ----------
Conservation
Sequence logo
2,060
2,080
Coffea canephora GTATATACTT CGACTTTCTT GTGCTAAAAC TTTGGCTCGG 1565
Coffea arabica ATATATACTT CGACTTTCTT GTGCTAAAAC TTTGGCTCGG 1994
Antirrhinum majus ---------- ---------- ---------- ---------- 1523
Veronicastrum sibiricum GTATATACTT CGACTTTCTT GTGCTAGAAC TTTGGCTCGG 1491
V. vulcanicum ---------- -GGCTTTCTT ---------- -------TGG 1561
V. luridiflorum ---------- -GGCTTTCTT ---------- -------TGG 1546
V. hasbenlii ---------- -GGCTTTCTT ---------- -------TGG 1551
Consensus ---------- -GNCTTTCTT ---------- -------NGG
Conservation
Sequence logo
2,100
2,120
Coffea canephora AAACACAAAA GTACTGTGCG TGCTTTTTTG AAAGGATTAG 1605
Coffea arabica AAACACAAAA GTACTGTGCG TGCTTTTTTG AAAGGATTAG 2034
Antirrhinum majus ---------- ---------- ---------- ---------- 1523
Veronicastrum sibiricum AAACACAAAA GTACTGTACG TGCTTTTTTG AAAACATTAG 1531
V. vulcanicum AGTAATAAGC CTTTTCCAAT TACAATACTC ATTGAGAAAG 1601
V. luridiflorum AGTAATAAGC CTATTCCAAT TACAATACTC ATTGAGAAAG 1586
V. hasbenlii AGTAATAAGC CTATTCCAAT TACAATACTC ATTGAGAAAG 1591
Consensus ANNNANAANN NTANTNNANN TNCNNTNNTN ANNGANNNAG
Conservation
Sequence logo
2,140
2,160
Coffea canephora GGTCGGAATT TTTGGAAAAA TTCCTCACGT CGGAAGAAGA 1645
Coffea arabica GGTCGGAATT TTTGGAAAAA TTCCTCACGT CGGAAGAAGA 2074
Antirrhinum majus ---------- ---------- ---------- ---------- 1523
Veronicastrum sibiricum GATCCGGATT ATTGGAAGAA TTTCTTTATAT CGGAAGAAGA 1571
V. vulcanicum AATCGTAATA AATGCAA--- ---------- ----AGAAGA 1624
V. luridiflorum AATCGTAATA AATGCAA--- ---------- ----AGAAGA 1609
V. hasbenlii AATCGTAATA AATGCAA--- ---------- ----AGAAGA 1614
Consensus NATCGNAATN ANTGNAA--- ---------- ----AGAAGA
Conservation
Sequence logo

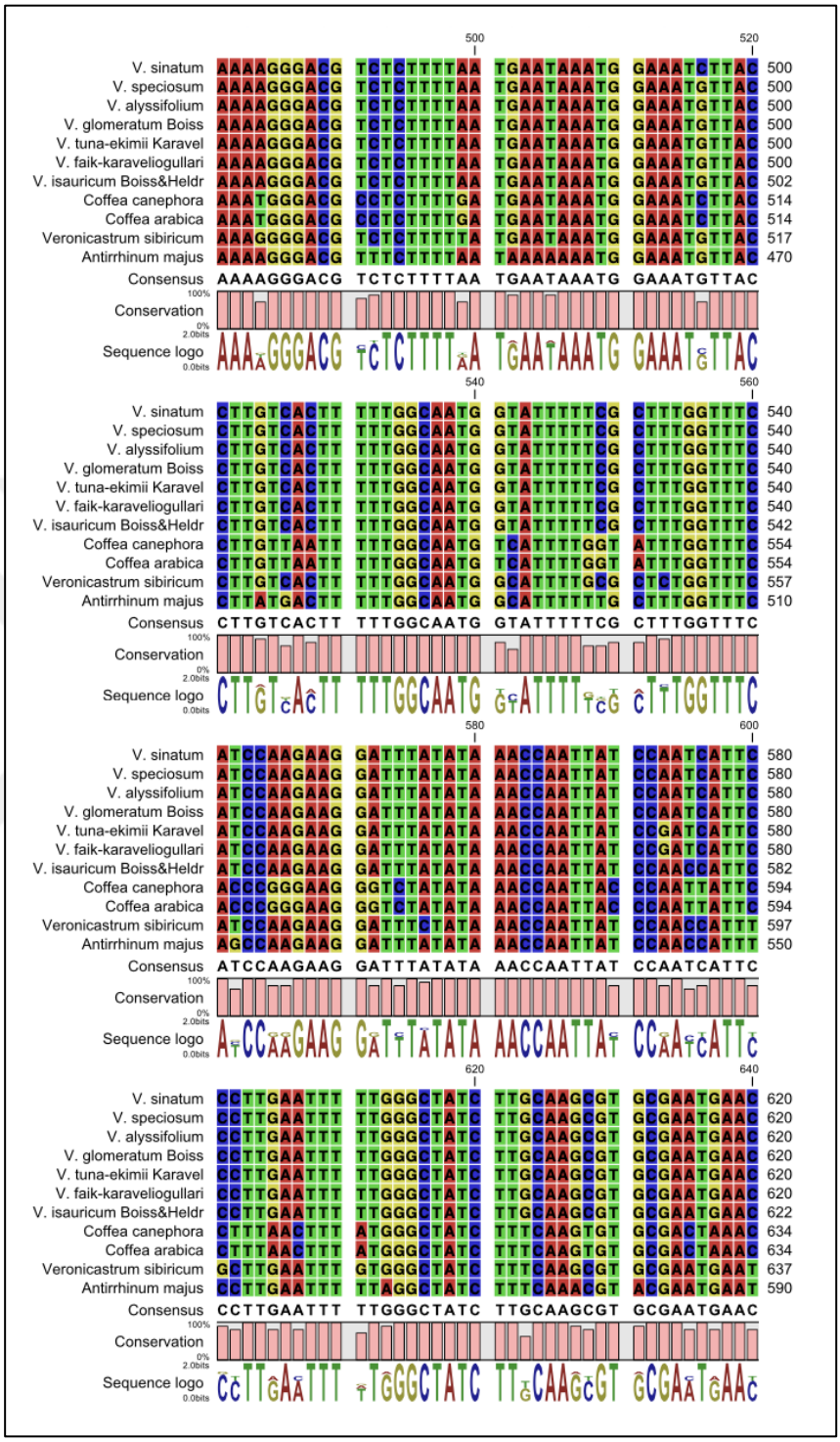

2,180
2,200
Coffea canephora 1685
Coffea arabica 2114
Antirrhinum majus 1523
Veronicastrum sibiricum 1611
V. vulcanicum 1632
V. luridiflorum 1617
V. hasbenlii 1622
Consensus
Conservation
Sequence logo
2,220
2,240
Coffea canephora 1725
Coffea arabica 2154
Antirrhinum majus 1523
Veronicastrum sibiricum 1651
V. vulcanicum 1632
V. luridiflorum 1617
V. hasbenlii 1622
Consensus
Conservation
Sequence logo
2,260
2,280
Coffea canephora 1765
Coffea arabica 2186
Antirrhinum majus 1523
Veronicastrum sibiricum 1683
V. vulcanicum 1658
V. luridiflorum 1643
V. hasbenlii 1648
Consensus
Conservation
Sequence logo
2,300
Coffea canephora 1794
Coffea arabica 2186
Antirrhinum majus 1523
Veronicastrum sibiricum 1683
V. vulcanicum 1686
V. luridiflorum 1672
V. hasbenlii 1674
Consensus
Conservation
Sequence logo

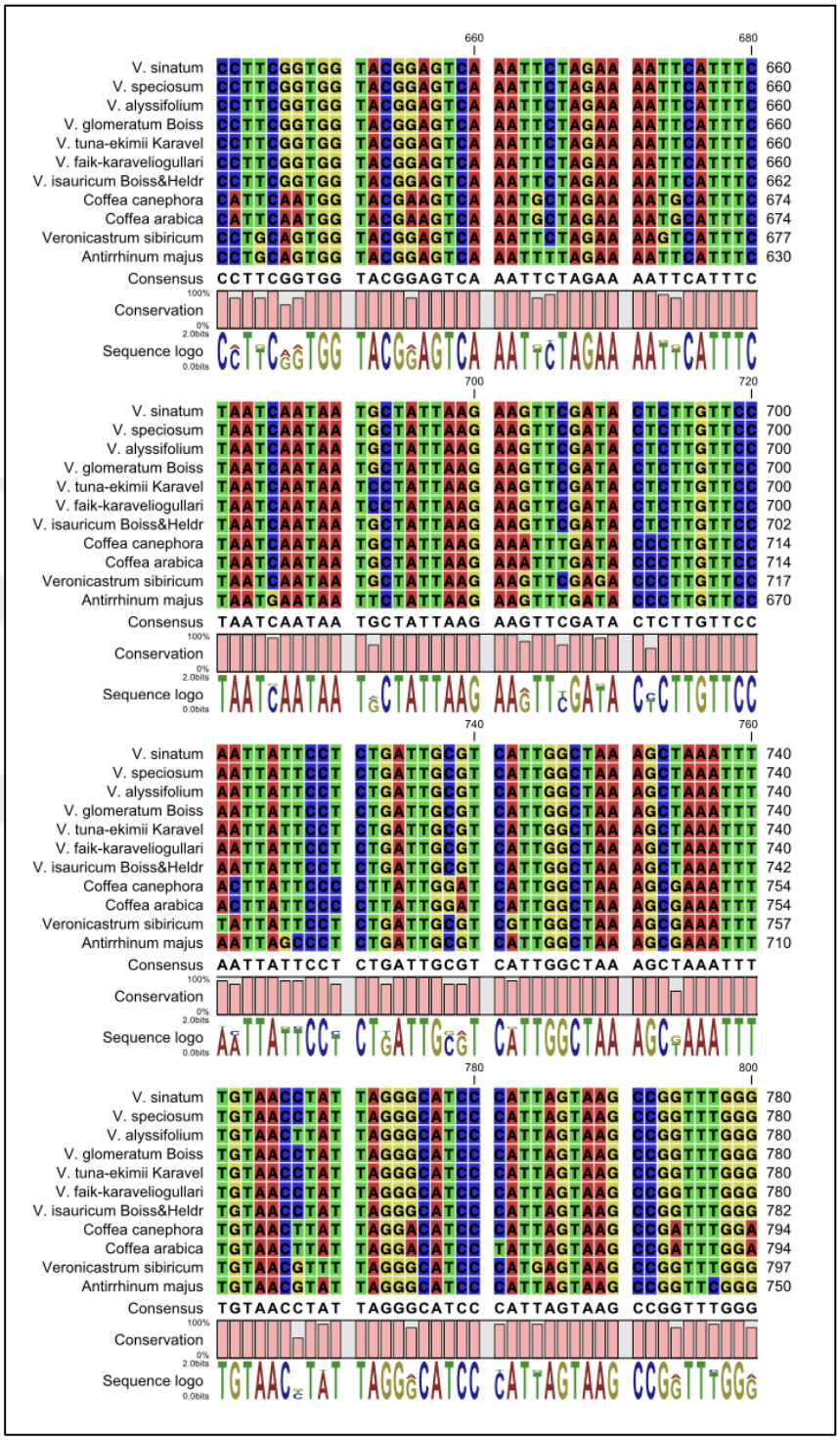
APPENDIX B: *ITS – matk* ASSEMBLED CHROMATOGRAM RESULTS OF NGS SEQUENCED *VERBASCUM* L. SAMPLES

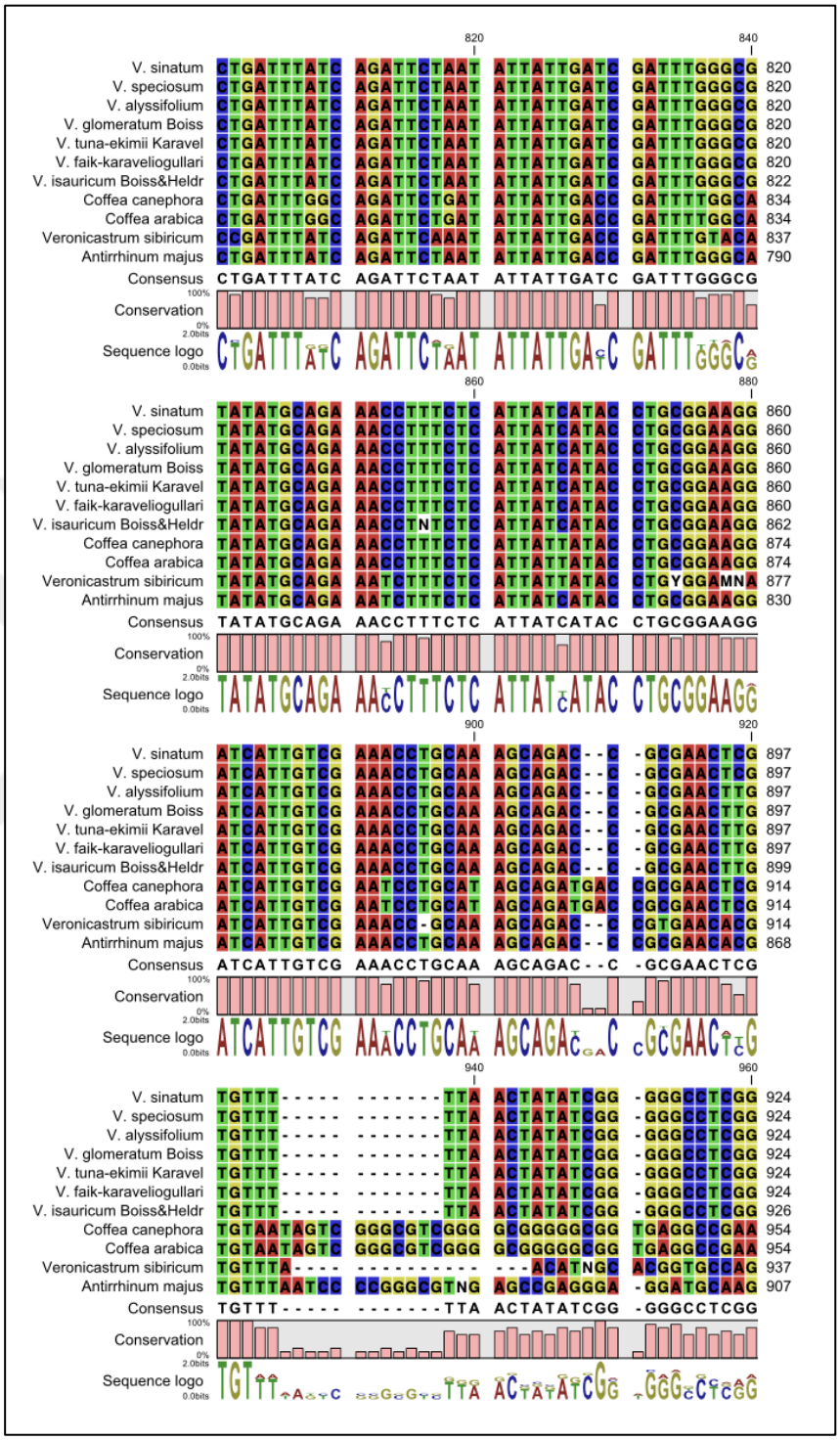


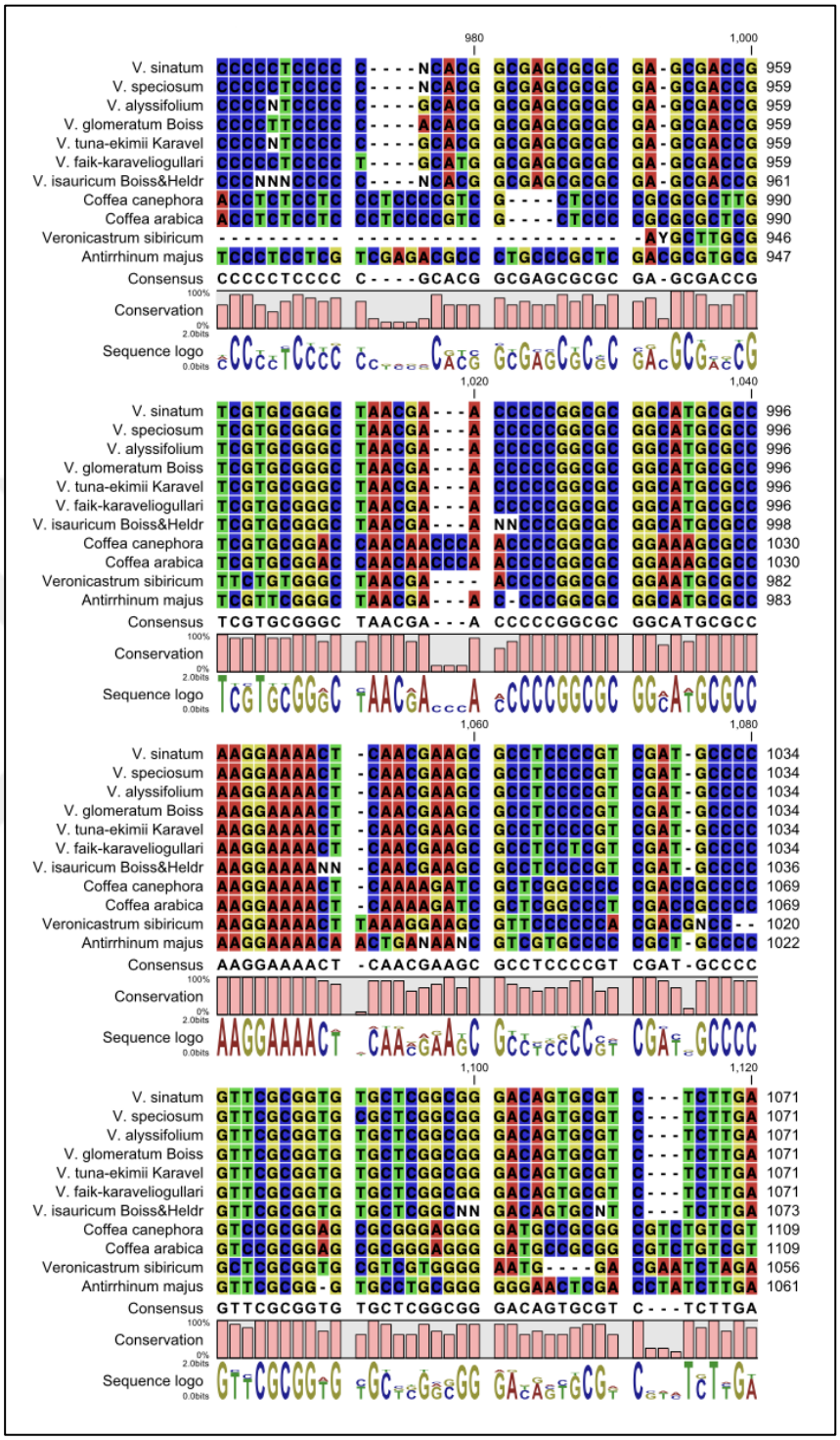

180
200
V. sinatum TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA 183
V. speciosum TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA 183
V. alyssifolium TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA 183
V. glomeratum Boiss TATTGTAATT GGAAAAGGCT TATTACTCCA AAGAAAGCCA 183
V. tuna-ekimii Karavel TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA 183
V. faik-karaveliogullari TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA 183
V. isauricum Boiss&Heldr TATTGTAATT GGAANAGGCT TATTACTCCA AAGAAAGCCA 185
Coffea canephora TATTGGAATT GGAATACTCT TATTGATAGA AAGAAACCCA 200
Coffea arabica TATTGGAATT GGAATACTCT TATTGATAGA AAGAAACCCA 200
Veronicastrum sibiricum TATTGTAATT CGAATAGTCT TATTACTACA AAGAAAGCCG 200
Antirrhinum majus TATTGTAATT GGAATAGTCT TATTCTTCCA AAGAAAGCCA 153
Consensus TATTGTAATT GGAATAGGCT TATTACTCCA AAGAAAGCCA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
220
240
V. sinatum GTTCCTCCTT TTCA---AAA AGGAATCAAA GGTTATTCTT 220
V. speciosum GTTCCTCCTT TTCA---AAA AGGAATCAAA GGTTATTCTT 220
V. alyssifolium GTTCCTCCTT TTCA---AAA AGAAATCAAA GGTTATTCTT 220
V. glomeratum Boiss GTTCCTCCTT TTCA---AAA AGGAATCAAA GGTTATTCTT 220
V. tuna-ekimii Karavel GTTCCTCCTT TTCA---AAA AGAAATCAAA GGTTATTCTT 220
V. faik-karaveliogullari GTTCCTCCTT TTCA---AAA AGAAATCAAA GGTTATTCTT 220
V. isauricum Boiss&Heldr GTTCCTCCTT TTCA---AAA AGAAATCAAA GGTTATTCTT 222
Coffea canephora GTTTTGATTT TTCACCAAAA AGAAATCAAA AATTATTCTT 240
Coffea arabica GTTTTGATTT TTCACCAAAA AGAAATCAAA AATTATTCTT 240
Veronicastrum sibiricum GTTTCTCTTT TTCA---AAA AGAAATCAAA GATTCTTTTT 237
Antirrhinum majus GTTTCTCTTG TTCA---AAA ATAAATCAAA GATTGTTTTT 190
Consensus GTTCCTCCTT TTCA---AAA AGAAATCAAA GGTTATTCTT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
260
280
V. sinatum ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 260
V. speciosum ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 260
V. alyssifolium ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 260
V. glomeratum Boiss ATTCTTATAT AATTATCATG TCGGTGAATA CGAATCCATT 260
V. tuna-ekimii Karavel ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 260
V. faik-karaveliogullari ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 260
V. isauricum Boiss&Heldr ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT 262
Coffea canephora CTTGTTATAT AATTCTCATG TATGTGAATA CGAATCCATT 280
Coffea arabica CTTGTTATAT AATTCTCATG TATGTGAATA CGAATCCATT 280
Veronicastrum sibiricum ATTCTTATAT AATTCTCATG TATGTGAGTA CGAATCCGTT 277
Antirrhinum majus ATTCTTATAT AATTCTCATG TATGTGAATA CGAATCCATT 230
Consensus ATTCTTATAT AATTATCATG TCTGTGAATA CGAATCCATT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
300
320
V. sinatum TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. speciosum TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. alyssifolium TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. glomeratum Boiss TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. tuna-ekimii Karavel TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. faik-karaveliogullari TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 300
V. isauricum Boiss&Heldr TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA 302
Coffea canephora TTCTTCTTTC TCCGTAAGCA ATCTTCTCAT TTGCGATCAA 320
Coffea arabica TTCTTCTTTC TCCGTAAGCA ATCTTCTCAT TTGCGATCAA 320
Veronicastrum sibiricum TTCGTCTTTT TACGTAACCA ATCTTCTCAT TTACGATCAA 317
Antirrhinum majus TTCGTCTTCC TACGTAACCA ATCTTCTCAT TTACTATCAA 270
Consensus TTCGTCTTTC TACGTAATCA ATCTTCTCAT TTAAGATCAA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

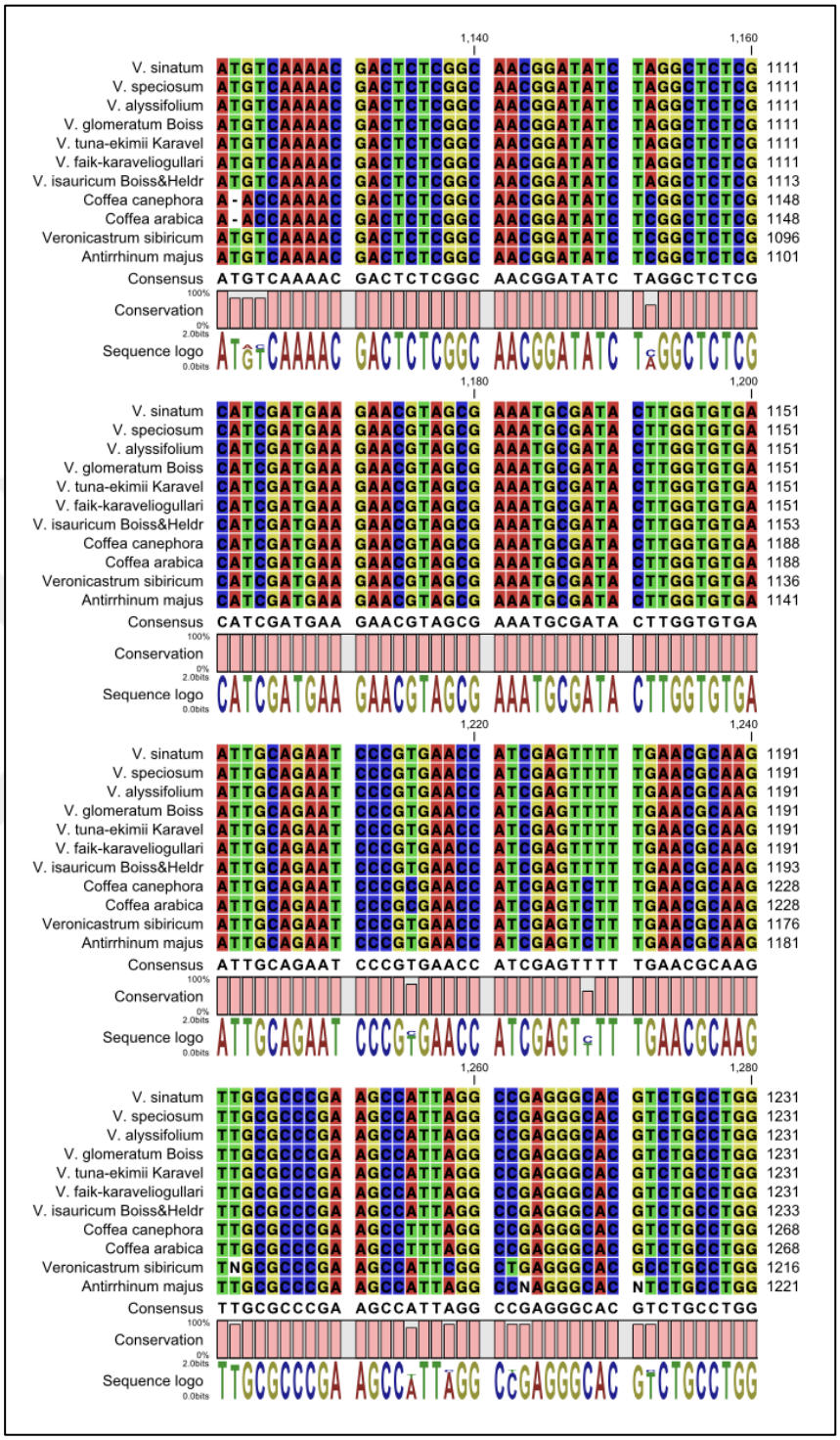

340
360
V. sinatum CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. speciosum CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. alyssifolium CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. glomeratum Boiss CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. tuna-ekimii Karavel CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. faik-karaveliogullari CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 340
V. isauricum Boiss&Heldr CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG 342
Coffea canephora CATCTTTTGG AGTCTTTCTT GAACGGATAT ATTTCTATGA 360
Coffea arabica CATCTTTTGG AGTCTTTCTT GAACGGATAT ATTTCTATGA 360
Veronicastrum sibiricum CATCTTCTGG AGCTCTTCTT GAACGAATCT ATTTCTATGG 357
Antirrhinum majus CATCTTCTGG AGTTCTTCTT GAACGAATCT ATTTCTATGG 310
Consensus CATCTTCTGG AGTTTTTCTT GAACGAATCT ATTTCTATGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
380
400
V. sinatum AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. speciosum AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. alyssifolium AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. glomeratum Boiss AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. tuna-ekimii Karavel AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. faik-karaveliogullari AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 380
V. isauricum Boiss&Heldr AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG 382
Coffea canephora AAAAAAAGAA CGTCTTGTAG AAGTCTTTG- ----- CTAAG 394
Coffea arabica AAAAAAAGAA CGTCTTGTAG AAGTCTTTG- ----- CTAAG 394
Veronicastrum sibiricum AAAAATAGAG CATCTTGTGA ACGTCTTTGG TAAGGTTAAG 397
Antirrhinum majus AAAAATAGAA TGTCTTGTGA ACGTCTTTGT TAAGGTTAAG 350
Consensus AAAAGTAGAA CGTCTTGTGA ACGTCTTTGG TAAAGTTAAG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
420
440
V. sinatum GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. speciosum GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. alyssifolium GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. glomeratum Boiss GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. tuna-ekimii Karavel GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. faik-karaveliogullari GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 420
V. isauricum Boiss&Heldr GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA 422
Coffea canephora GATTTTCAAG CCAGTCTATG GTTGTTCAAA GATCCTTTCA 434
Coffea arabica GATTTTCAAG CCAGTCTATG GTTGTTCAAA GATCCTTTCA 434
Veronicastrum sibiricum GACTGTCAGG TGAACCTATG GTTGGTCAAA GAACCTTGCA 437
Antirrhinum majus GATTTTCAGG CGAATCTATG GTTGGTCAAA GAACCTTGCA 390
Consensus GATTTTCAGG AGAACCTATG GTTGGTCAAG GAACCTTGCA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
460
480
V. sinatum TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. speciosum TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. alyssifolium TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. glomeratum Boiss TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. tuna-ekimii Karavel TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. faik-karaveliogullari TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 460
V. isauricum Boiss&Heldr TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC 462
Coffea canephora TACATTATGT TAGGTATCAA GGAAAATCCA TTCTGGTTTC 474
Coffea arabica TACATTATGT TAGGTATCAA GGAAAATCCA TTCTGGTTTC 474
Veronicastrum sibiricum TGCATTATGT TAGATATCAA AGAAAATCCA TTCTGGCTTT 477
Antirrhinum majus TGCATTATGT TAGGTATCAA AGAAAATCAA TTCTGGCTTC 430
Consensus TGCATTATGT TAGGTATCAA AGAAAATCCA TTCTGGCTTC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

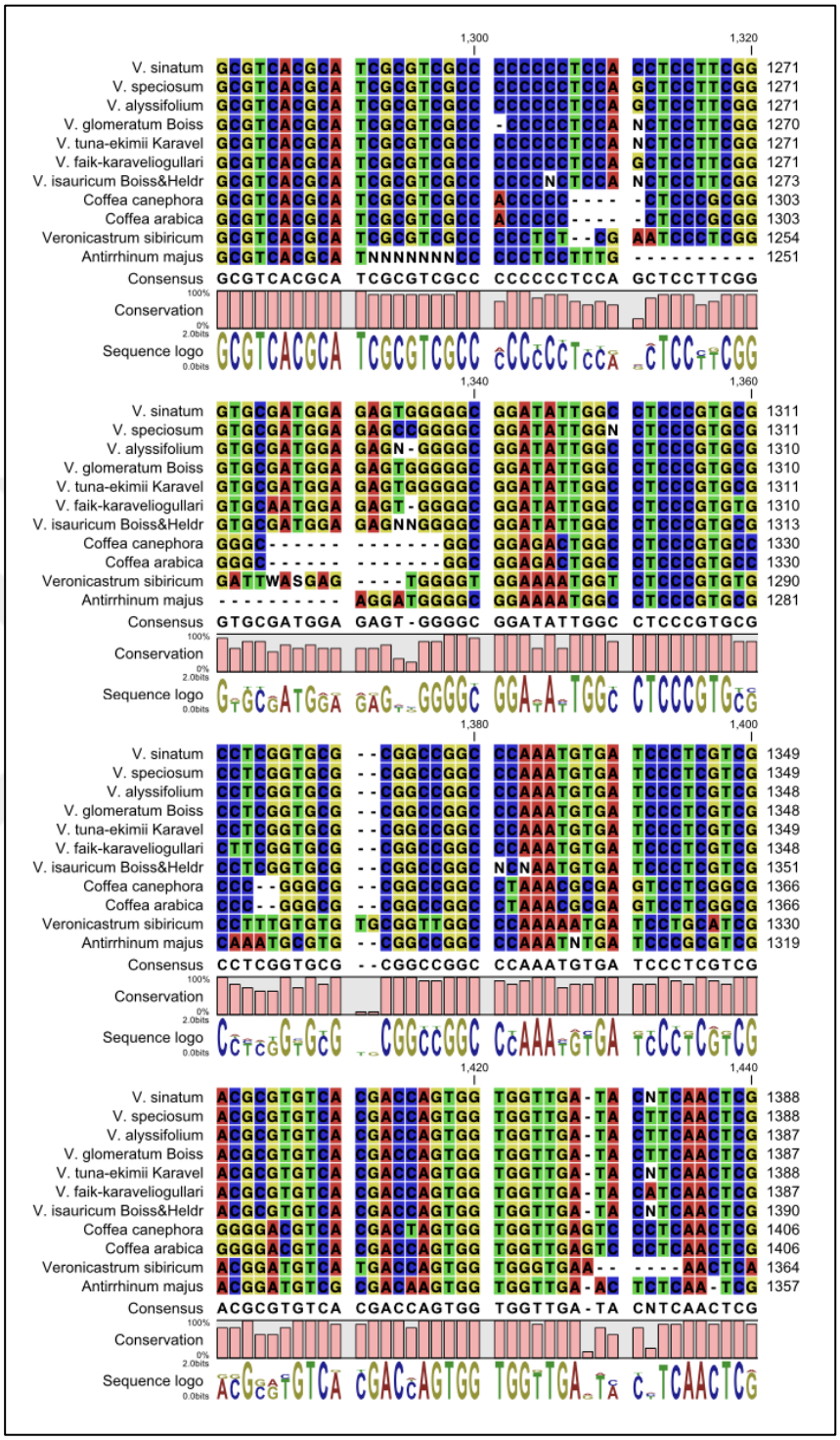

500
520
V. sinatum AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATCTTAC 500
V. speciosum AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 500
V. alyssifolium AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 500
V. glomeratum Boiss AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 500
V. tuna-ekimii Karavel AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 500
V. faik-karaveliogullari AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 500
V. isauricum Boiss&Heldr AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC 502
Coffea canephora AAATGGGACG CCTCTTTTGA TGAATAAATG GAAATCTTAC 514
Coffea arabica AAATGGGACG CCTCTTTTGA TGAATAAATG GAAATCTTAC 514
Veronicastrum sibiricum AAAGGGGACG TCTCTTTTTA TGAATAAATG GAAATGTTAC 517
Antirrhinum majus AAAAGGGACG TTTCTTTTAA TAAAAAAATG GAAATGTTAC 470
Consensus AAAAGGGACG TCTCTTTTAA TGAATAAATG GAAATGTTAC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
540
560
V. sinatum CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. speciosum CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. alyssifolium CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. glomeratum Boiss CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. tuna-ekimii Karavel CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. faik-karaveliogullari CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 540
V. isauricum Boiss&Heldr CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC 542
Coffea canephora CTTGTTAATT TTTGGCAATG TCATTTTGGT ATTTGGTTTC 554
Coffea arabica CTTGTTAATT TTTGGCAATG TCATTTTGGT ATTTGGTTTC 554
Veronicastrum sibiricum CTTGTCACTT TTTGGCAATG GCATTTTGCG CTCTGGTTTC 557
Antirrhinum majus CTTATGACTT TTTGGCAATG GCATTTTTTG CTTTGGTTTC 510
Consensus CTTGTCACTT TTTGGCAATG GTATTTTTCG CTTTGGTTTC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
580
600
V. sinatum ATCCAAGAAG GATTTATATA AACCAATTAT CCAATCATTC 580
V. speciosum ATCCAAGAAG GATTTATATA AACCAATTAT CCAATCATTC 580
V. alyssifolium ATCCAAGAAG GATTTATATA AACCAATTAT CCAATCATTC 580
V. glomeratum Boiss ATCCAAGAAG GATTTATATA AACCAATTAT CCAATCATTC 580
V. tuna-ekimii Karavel ATCCAAGAAG GATTTATATA AACCAATTAT CCGATCATTC 580
V. faik-karaveliogullari ATCCAAGAAG GATTTATATA AACCAATTAT CCGATCATTC 580
V. isauricum Boiss&Heldr ATCCAAGAAG GATTTATATA AACCAATTAT CCAACCATTC 582
Coffea canephora ACCCGGGAAG GGTCTATATA AACCAATTAC CCAATTATTC 594
Coffea arabica ACCCGGGAAG GGTCTATATA AACCAATTAC CCAATTATTC 594
Veronicastrum sibiricum ATCCAAGAAG GATTTCTATA AACCAATTAT CCAACCATTT 597
Antirrhinum majus AGCCAAGAAG GATTTATATA AACCAATTAT CCAACCATTT 550
Consensus ATCCAAGAAG GATTTATATA AACCAATTAT CCAATCATTC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
620
640
V. sinatum CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. speciosum CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. alyssifolium CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. glomeratum Boiss CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. tuna-ekimii Karavel CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. faik-karaveliogullari CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 620
V. isauricum Boiss&Heldr CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC 622
Coffea canephora CTTTAACTTT ATGGGCTATC TTTCAAGTGT GCGACTAAAC 634
Coffea arabica CTTTAACTTT ATGGGCTATC TTTCAAGTGT GCGACTAAAC 634
Veronicastrum sibiricum GCTTGAATTT GTGGGCTATC TTTCAAGCGT GCGAATGAAT 637
Antirrhinum majus CCTTGAATTT TTAGGCTATC TTTCAAACGT ACGAATGAAT 590
Consensus CCTTGAATTT TTGGGCTATC TTGCAAGCGT GCGAATGAAC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

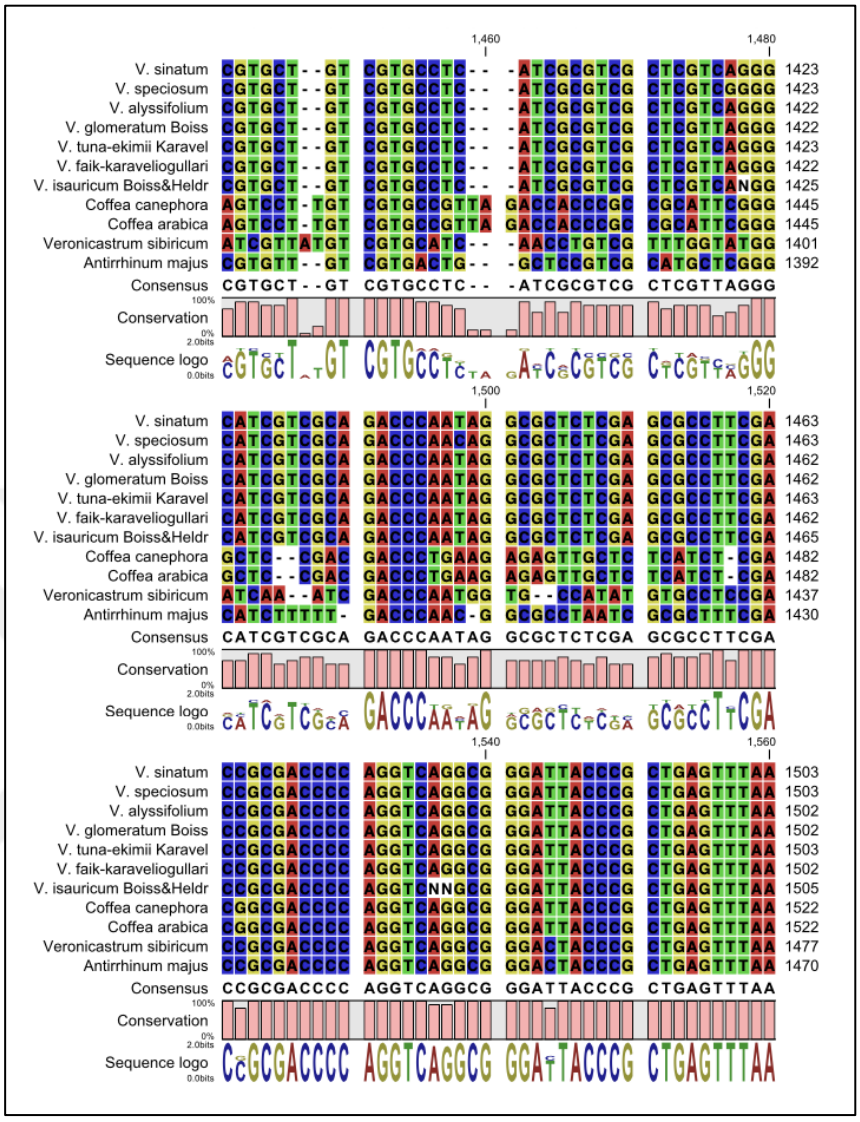

660
680
V. sinatum CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. speciosum CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. alyssifolium CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. glomeratum Boiss CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. tuna-ekimii Karavel CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. faik-karaveliogullari CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 660
V. isauricum Boiss&Heldr CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC 662
Coffea canephora CATTCAATGG TACGAAGTCA AATGCTAGAA AATGCATTTC 674
Coffea arabica CATTCAATGG TACGAAGTCA AATGCTAGAA AATGCATTTC 674
Veronicastrum sibiricum CCTGCAGTGG TACGGAGTCA AATTCTAGAA AAGTCATTTC 677
Antirrhinum majus CCTGCAGTGG TACGGAGTCA AATTTTAGAA AATTCATTTC 630
Consensus CCTTCGGTGG TACGGAGTCA AATTCTAGAA AATTCATTTC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
700
720
V. sinatum TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. speciosum TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. alyssifolium TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. glomeratum Boiss TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. tuna-ekimii Karavel TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. faik-karaveliogullari TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 700
V. isauricum Boiss&Heldr TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC 702
Coffea canephora TAATCAATAA TGCTATTAAG AAATTTGATA CGCTTGTTCC 714
Coffea arabica TAATCAATAA TGCTATTAAG AAATTTGATA CGCTTGTTCC 714
Veronicastrum sibiricum TAATCAATAA TGCTATTAAG AAGTTCGAGA CCCTTGTTCC 717
Antirrhinum majus TAATGAATAA TTCTATTAAG AAGTTTGATA CCCTTGTTCC 670
Consensus TAATCAATAA TGCTATTAAG AAGTTCGATA CTCTTGTTCC
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
740
760
V. sinatum AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. speciosum AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. alyssifolium AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. glomeratum Boiss AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. tuna-ekimii Karavel AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. faik-karaveliogullari AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 740
V. isauricum Boiss&Heldr AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT 742
Coffea canephora ACTTATTCCC CTTATTGGAT CATTGGCTAA AGCGAAATTT 754
Coffea arabica ACTTATTCCC CTTATTGGAT CATTGGCTAA AGCGAAATTT 754
Veronicastrum sibiricum TATTATTCCT CTGATTGCGT CGTTGGCTAA AGCGAAATTT 757
Antirrhinum majus AATTAGCCCT CTGATTGCGT CATTGGCTAA AGCGAAATTT 710
Consensus AATTATTCCT CTGATTGCGT CATTGGCTAA AGCTAAATTT
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
780
800
V. sinatum TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. speciosum TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. alyssifolium TGTAACTTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. glomeratum Boiss TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. tuna-ekimii Karavel TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. faik-karaveliogullari TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 780
V. isauricum Boiss&Heldr TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG 782
Coffea canephora TGTAACTTAT TAGGACATCC CATTAGTAAG CCGATTTGGA 794
Coffea arabica TGTAACTTAT TAGGACATCC TATTAGTAAG CCGATTTGGA 794
Veronicastrum sibiricum TGTAACGTTT TAGGGCATCC CATGAGTAAG CCGGTTTGGG 797
Antirrhinum majus TGTAACGTAT TAGGGCATCC CATTAGTAAG CCGGTTCGGG 750
Consensus TGTAACCTAT TAGGGCATCC CATTAGTAAG CCGGTTTGGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits


820
840
V. sinatum CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. speciosum CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. alyssifolium CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. glomeratum Boiss CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. tuna-ekimii Karavel CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. faik-karaveliogullari CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 820
V. isauricum Boiss&Heldr CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG 822
Coffea canephora CTGATTTGGC AGATTCTGAT ATTATTGACC GATTTTGGCA 834
Coffea arabica CTGATTTGGC AGATTCTGAT ATTATTGACC GATTTTGGCA 834
Veronicastrum sibiricum CCGATTTATC AGATTCAAAT ATTATTGACC GATTTGTACA 837
Antirrhinum majus CTGATTTATC AGATTCTAAT ATTATTGACC GATTTGGGCA 790
Consensus CTGATTTATC AGATTCTAAT ATTATTGATC GATTTGGGCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
860
880
V. sinatum TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. speciosum TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. alyssifolium TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. glomeratum Boiss TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. tuna-ekimii Karavel TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. faik-karaveliogullari TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG 860
V. isauricum Boiss&Heldr TATATGCAGA AACCTNTCTC ATTATCATAC CTGCGGAAGG 862
Coffea canephora TATATGCAGA AACCTTTCTC ATTATTATAC CTGCGGAAGG 874
Coffea arabica TATATGCAGA AACCTTTCTC ATTATTATAC CTGCGGAAGG 874
Veronicastrum sibiricum TATATGCAGA AATCTTTCTC ATTATTATAC CTGYGGAMNA 877
Antirrhinum majus TATATGCAGA AATCTTTCTC ATTATCATAC CTGCGGAAGG 830
Consensus TATATGCAGA AACCTTTCTC ATTATCATAC CTGCGGAAGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
900
920
V. sinatum ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTCG 897
V. speciosum ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTCG 897
V. alyssifolium ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTTG 897
V. glomeratum Boiss ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTTG 897
V. tuna-ekimii Karavel ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTTG 897
V. faik-karaveliogullari ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTTG 897
V. isauricum Boiss&Heldr ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTTG 899
Coffea canephora ATCATTGTCG AATCCTGCAT AGCAGATGAC CGCGAACTCG 914
Coffea arabica ATCATTGTCG AATCCTGCAT AGCAGATGAC CGCGAACTCG 914
Veronicastrum sibiricum ATCATTGTCG AAACC-GCAA AGCAGAC--C CGTGAACACG 914
Antirrhinum majus ATCATTGTCG AAACCTGCAA AGCAGAC--C CGCGAACACG 868
Consensus ATCATTGTCG AAACCTGCAA AGCAGAC--C -GCGAACTCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
940
960
V. sinatum TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. speciosum TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. alyssifolium TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. glomeratum Boiss TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. tuna-ekimii Karavel TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. faik-karaveliogullari TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 924
V. isauricum Boiss&Heldr TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG 926
Coffea canephora TGTAATAGTC GGGCGTCGGG GCGGGGGCGG TGAGGCCGAA 954
Coffea arabica TGTAATAGTC GGGCGTCGGG GCGGGGGCGG TGAGGCCGAA 954
Veronicastrum sibiricum TGTTTA---- ---------- ---ACATNGC ACGGTGCCAG 937
Antirrhinum majus TGTTTAATCC CCGGGCGTNG AGCCGAGGGA -GGATGCAAG 907
Consensus TGTTT----- -------TTA ACTATATCGG -GGGCCTCGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits


980
1,000
V. sinatum CCCCCTCCCC C----NCACG GCGAGCGCGC GA-GCGACCG 959
V. speciosum CCCCCTCCCC C----NCACG GCGAGCGCGC GA-GCGACCG 959
V. alyssifolium CCCCNTCCCC C----GCACG GCGAGCGCGC GA-GCGACCG 959
V. glomeratum Boiss CCCCTTCCCC C----ACACG GCGAGCGCGC GA-GCGACCG 959
V. tuna-ekimii Karavel CCCCNTCCCC C----GCACG GCGAGCGCGC GA-GCGACCG 959
V. faik-karaveliogullari CCCCCTCCCC T----GCATG GCGAGCGCGC GA-GCGACCG 959
V. isauricum Boiss&Heldr CCCNNNCCCC C----NCACG GCGAGCGCGC GA-GCGACCG 961
Coffea canephora ACCTCTCCCC CCTCCCCGTC G----CTCCC CGCGCGCTTG 990
Coffea arabica ACCTCTCCCC CCTCCCCGTC G----CTCCC CGCGCGCTCG 990
Veronicastrum sibiricum ---------- ---------- ---------- -AYGCTTGCG 946
Antirrhinum majus TCCCTCCTCG TCGAGACGCC CTGCCCGCTC GACGCGTGCG 947
Consensus CCCCCTCCCC C----GCACG GCGAGCGCGC GA-GCGACCG
Conservation
Sequence logo
1,020
1,040
V. sinatum TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. speciosum TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. alyssifolium TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. glomeratum Boiss TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. tuna-ekimii Karavel TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. faik-karaveliogullari TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC 996
V. isauricum Boiss&Heldr TCGTGCGGGC TAACGA---A NNCCCGGCGC GGCATGCGCC 998
Coffea canephora TCGTGCGGAC CAACAACCCA ACCCCGGCGC GGAAAGCGCC 1030
Coffea arabica TCGTGCGGAC CAACAACCCA ACCCCGGCGC GGAAAGCGCC 1030
Veronicastrum sibiricum TTCTGTGGGC TAACGA---- ACCCCGGCGC GGAATGCGCC 982
Antirrhinum majus TCGTTCGGGC TAACGA---A C-CCCGGCGC GGCATGCGCC 983
Consensus TCGTGCGGGC TAACGA---A CCCCCGGCGC GGCATGCGCC
Conservation
Sequence logo
1,060
1,080
V. sinatum AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1034
V. speciosum AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1034
V. alyssifolium AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1034
V. glomeratum Boiss AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1034
V. tuna-ekimii Karavel AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1034
V. faik-karaveliogullari AAGGAAAACT -CAACGAAGC GCCTCCTCGT CGAT-GCCCC 1034
V. isauricum Boiss&Heldr AAGGAAAANN -CAACGAAGC GCCTCCCCGT CGAT-GCCCC 1036
Coffea canephora AAGGAAAACT -CAAAAGATC GCTCGGCCCC CGACCGCCCC 1069
Coffea arabica AAGGAAAACT -CAAAAGATC GCTCGGCCCT CGACCGCCCC 1069
Veronicastrum sibiricum AAGGAAAACT TAAAGGAAGC GTTCCCCCCA CGACGNCC-- 1020
Antirrhinum majus AAGGAAAACA ACTGANAANC GTCGTGCCCC CGCT-GCCCC 1022
Consensus AAGGAAAACT -CAACGAAGC GCCTCCCCGT CGAT-GCCCC
Conservation
Sequence logo
1,100
1,120
V. sinatum GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. speciosum GTTCGCGGTG CGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. alyssifolium GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. glomeratum Boiss GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. tuna-ekimii Karavel GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. faik-karaveliogullari GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA 1071
V. isauricum Boiss&Heldr GTTCGCGGTG TGCTCGGCNN GACAGTGCNT C---TCTTGA 1073
Coffea canephora GTCCGCGGAG CGCGGGAGGG GATGCCGCGG CGTCTGTCGT 1109
Coffea arabica GTCCGCGGAG CGCGGGAGGG GATGCCGCGG CGTCTGTCGT 1109
Veronicastrum sibiricum GCTCGCGGTG CGTCGTGGGG AATG----GA CGAATCTAGA 1056
Antirrhinum majus GTTCGCGG-G TGCCTGCGGG GGGAACTCGA CCTATCTTGA 1061
Consensus GTTCGCGGTG TGCTCGGCGG GACAGTGCGT C---TCTTGA
Conservation
Sequence logo


1,140
1,160
V. sinatum ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. speciosum ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. alyssifolium ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. glomeratum Boiss ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. tuna-ekimii Karavel ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. faik-karaveliogullari ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1111
V. isauricum Boiss&Heldr ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG 1113
Coffea canephora A-ACCAAAAC GACTCTCGGC AACGGATATC TCGGCTCTCG 1148
Coffea arabica A-ACCAAAAC GACTCTCGGC AACGGATATC TCGGCTCTCG 1148
Veronicastrum sibiricum ATGTCAAAAC GACTCTCGGC AACGGATATC TCGGCTCTCG 1096
Antirrhinum majus ATGTCAAAAC GACTCTCGGC AACGGATATC TCGGCTCTCG 1101
Consensus ATGTCAAAAC GACTCTCGGC AACGGATATC TAGGCTCTCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,180
1,200
V. sinatum CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. speciosum CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. alyssifolium CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. glomeratum Boiss CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. tuna-ekimii Karavel CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. faik-karaveliogullari CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1151
V. isauricum Boiss&Heldr CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1153
Coffea canephora CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1188
Coffea arabica CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1188
Veronicastrum sibiricum CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1136
Antirrhinum majus CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA 1141
Consensus CATCGATGAA GAACGTAGCG AAATGCGATA CTTGGTGTGA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,220
1,240
V. sinatum ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. speciosum ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. alyssifolium ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. glomeratum Boiss ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. tuna-ekimii Karavel ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. faik-karaveliogullari ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1191
V. isauricum Boiss&Heldr ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG 1193
Coffea canephora ATTGCAGAAT CCCGCGAACC ATCGAGTCTT TGAACGCAAG 1228
Coffea arabica ATTGCAGAAT CCCGCGAACC ATCGAGTCTT TGAACGCAAG 1228
Veronicastrum sibiricum ATTGCAGAAT CCCGTGAACC ATCGAGTCTT TGAACGCAAG 1176
Antirrhinum majus ATTGCAGAAT CCCGTGAACC ATCGAGTCTT TGAACGCAAG 1181
Consensus ATTGCAGAAT CCCGTGAACC ATCGAGTTTT TGAACGCAAG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,260
1,280
V. sinatum TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. speciosum TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. alyssifolium TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. glomeratum Boiss TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. tuna-ekimii Karavel TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. faik-karaveliogullari TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1231
V. isauricum Boiss&Heldr TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG 1233
Coffea canephora TTGCGCCCGA AGCCTTTAGG CCGAGGGCAC GTCTGCCTGG 1268
Coffea arabica TTGCGCCCGA AGCCTTTAGG CCGAGGGCAC GTCTGCCTGG 1268
Veronicastrum sibiricum TNGCGCCCGA AGCCATTCGG CTGAGGGCAC GCCTGCCTGG 1216
Antirrhinum majus TTGCGCCCGA AGCCATTAGG CCNAGGGCAC NTCTGCCTGG 1221
Consensus TTGCGCCCGA AGCCATTAGG CCGAGGGCAC GTCTGCCTGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits


1,300
1,320
V. sinatum GCGTCACGCA TCGCGTCGCC CCCCCCTCCA GCTCCTTCGG 1271
V. speciosum GCGTCACGCA TCGCGTCGCC CCCCCCTCCA GCTCCTTCGG 1271
V. alyssifolium GCGTCACGCA TCGCGTCGCC CCCCCCTCCA GCTCCTTCGG 1271
V. glomeratum Boiss GCGTCACGCA TCGCGTCGCC -CCCCCTCCA NCTCCTTCGG 1270
V. tuna-ekimii Karavel GCGTCACGCA TCGCGTCGCC CCCCCCTCCA NCTCCTTCGG 1271
V. faik-karaveliogullari GCGTCACGCA TCGCGTCGCC CCCCCCTCCA GCTCCTTCGG 1271
V. isauricum Boiss&Heldr GCGTCACGCA TCGCGTCGCC CCCCNCTCCA NCTCCTTCGG 1273
Coffea canephora GCGTCACGCA TCGCGTCGCC ACCCCC---- -CTCCCGCGG 1303
Coffea arabica GCGTCACGCA TCGCGTCGCC ACCCCC---- -CTCCCGCGG 1303
Veronicastrum sibiricum GCGTCACGCA TCGCGTCGCC CCCTCT--GG AATCCCTCGG 1254
Antirrhinum majus GCGTCACGCA TNNNNNNNCC CCCTCCTTTG ---------- 1251
Consensus GCGTCACGCA TCGCGTCGCC CCCCCCTCCA GCTCCTTCGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,340
1,360
V. sinatum GTGCGATGGA GAGTGGGGGC GGATATTGGC CTCCCCGTGCG 1311
V. speciosum GTGCGATGGA GAGCCGGGGC GGATATTGGN CTCCCGTGCG 1311
V. alyssifolium GTGCGATGGA GAGN-GGGGC GGATATTGGC CTCCCGTGCG 1310
V. glomeratum Boiss GTGCGATGGA GAGTGGGGGC GGATATTGGC CTCCCGTGCG 1310
V. tuna-ekimii Karavel GTGCGATGGA GAGTGGGGGC GGATATTGGC CTCCCGTGCG 1311
V. faik-karaveliogullari GTGCAATGGA GAGT-GGGGC GGATATTGGC CTCCCGTGTG 1310
V. isauricum Boiss&Heldr GTGCGATGGA GAGNNGGGGC GGATATTGGC CTCCCGTGCG 1313
Coffea canephora GGGC------ -------GGC GGAGACTGGC CTCCCGTGCC 1330
Coffea arabica GGGC------ -------GGC GGAGACTGGC CTCCCGTGCC 1330
Veronicastrum sibiricum GATTWASGAG ----TGGGGT GGAAAATGGT CTCCCGTGTG 1290
Antirrhinum majus ---------- AGGATGGGGC GGAAAATGGC CTCCCGTGCG 1281
Consensus GTGCGATGGA GAGT-GGGGC GGATATTGGC CTCCCGTGCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,380
1,400
V. sinatum CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1349
V. speciosum CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1349
V. alyssifolium CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1348
V. glomeratum Boiss CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1348
V. tuna-ekimii Karavel CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1349
V. faik-karaveliogullari CTTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG 1348
V. isauricum Boiss&Heldr CCTCGGTGCG --CGGCCGGC NCNAATGTGA TCCCTCGTCG 1351
Coffea canephora CCC--GGGCG --CGGCCGGC CTAAACGCGA GTCCTCGGCG 1366
Coffea arabica CCC--GGGCG --CGGCCGGC CTAAACGCGA GTCCTCGGCG 1366
Veronicastrum sibiricum CCTTTGTGTG TGCGGTTGGC CCAAAAATGA TCCTGCATCG 1330
Antirrhinum majus CAAATGCGTG --CGGCCGGC CCAAATNTGA TCCCGCGTCG 1319
Consensus CCTCGGTGCG --CGGCCGGC CCAAATGTGA TCCCTCGTCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,420
1,440
V. sinatum ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CNTCAACTCG 1388
V. speciosum ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CTTCAACTCG 1388
V. alyssifolium ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CTTCAACTCG 1387
V. glomeratum Boiss ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CTTCAACTCG 1387
V. tuna-ekimii Karavel ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CNTCAACTCG 1388
V. faik-karaveliogullari ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CATCAACTCG 1387
V. isauricum Boiss&Heldr ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CNTCAACTCG 1390
Coffea canephora GGGGACGTCA CGACTAGTGG TGGTTGAGTC CCTCAACTCG 1406
Coffea arabica GGGGACGTCA CGACCAGTGG TGGTTGAGTC CCTCAACTCG 1406
Veronicastrum sibiricum ACGGATGTCA TGACCAGTGG TGGGTGAA-- ----AACTCA 1364
Antirrhinum majus ACGGATGTCG CGACAAGTGG TGGTTGA-AC TCTCAA-TCG 1357
Consensus ACGCGTGTCA CGACCAGTGG TGGTTGA-TA CNTCAACTCG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits


1,460
1,480
V. sinatum CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTCAGGG 1423
V. speciosum CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTCGGGG 1423
V. alyssifolium CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTCAGGG 1422
V. glomeratum Boiss CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTTAGGG 1422
V. tuna-ekimii Karavel CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTCAGGG 1423
V. faik-karaveliogullari CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTTAGGG 1422
V. isauricum Boiss&Heldr CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTCANGG 1425
Coffea canephora AGTCCT-TGT CGTGCCGTTA GACCACCCGC CGCATTCGGG 1445
Coffea arabica AGTCCT-TGT CGTGCCGTTA GACCACCCGC CGCATTCGGG 1445
Veronicastrum sibiricum ATCGTTATGT CGTGCATG--AACCTGTCG TTTGGTATGG 1401
Antirrhinum majus CGTGTT--GT CGTGACTG--GCTCCGTCG CATGCTCGGG 1392
Consensus CGTGCT--GT CGTGCCTC--ATCGCGTCG CTCGTTAGGG
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,500
1,520
V. sinatum CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1463
V. speciosum CATCGTCGCA GACCCAACAG GCGCTCTCGA GCGCCTTCGA 1463
V. alyssifolium CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1462
V. glomeratum Boiss CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1462
V. tuna-ekimii Karavel CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1463
V. faik-karaveliogullari CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1462
V. isauricum Boiss&Heldr CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA 1465
Coffea canephora GCTC--CGAC GACCCTGAAG AGAGTTGCTC TCATCT-CGA 1482
Coffea arabica GCTC--CGAC GACCCTGAAG AGAGTTGCTC TCATCT-CGA 1482
Veronicastrum sibiricum ATCAA--ATC GACCCAATGG TG--CCATAT GTGCCTCCGA 1437
Antirrhinum majus CATCTTTTT- GACCCAAC-G GCGCCTAATC GCGCTTTCGA 1430
Consensus CATCGTCGCA GACCCAATAG GCGCTCTCGA GCGCCTTCGA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits
1,540
1,560
V. sinatum CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1503
V. speciosum CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1503
V. alyssifolium CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1502
V. glomeratum Boiss CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1502
V. tuna-ekimii Karavel CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1503
V. faik-karaveliogullari CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1502
V. isauricum Boiss&Heldr CCGCGACCCC AGGTCNNGCG GGATTACCCG CTGAGTTTAA 1505
Coffea canephora CGGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1522
Coffea arabica CGGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA 1522
Veronicastrum sibiricum CCGCGACCCC AGGTCAGGCG GGACTACCCG CTGAGTTTAA 1477
Antirrhinum majus CCGCGACCCC AGGTCAGGCG GGACTACCCG CTGAGTTTAA 1470
Consensus CCGCGACCCC AGGTCAGGCG GGATTACCCG CTGAGTTTAA
100%
Conservation
0%
2.0bits
Sequence logo
0.0bits

APPENDIX C: FASTA FILES OF THE ANALYZED *VERBASCUM* L. SAMPLES SEQUENCES

ITS NGS

LOCUS *V. alyssifolium* 868 bp DNA linear

NNNNNNNNNNNNNNTTCNNNTNCNTNTTCCTNTNGNTNNTACGTNGNCNNTTC
ACTGCCANCAAAGTACGCAGCAGTCCACTGAACCTTATCATTTAGAGGAAGGA
GAAG
TCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAAACCTG
CAAAGCAGACCGCGAACTTGTGTTTTT AACTATATCG GGGGCCTCGG
CCCCNTCCCC
CGCACGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAACGAACCCCCGGCGC
GGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCCCGTCGATGCCCCGTTCG
CGGTGTGCTCGGCGGGACAGTGCGTCTCTTGAATGTCAAAACGACTCTCGGCA
ACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTG
GTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCC
CGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCATCGCGTCGCC
CCCCCCTCCAGCTCCTTCGGGTGCGATGGAGAGNGGGGCGGATATTGGCCTCC
CGTGCGCCTCGGTGCGCGGCCGGCCCAAATGTGATCCCTCGTCGACGCGTGTC
ACGACCAGTGGTGGTTGATACTTCAACTCGCGTGCTGTCGTGCCTCATCGCGTC
GCTCGTCAGGGCATCGTCGCAGACCCAATAGGCGCTCTCGAGCGCCTTCGACC
GCGACCCCAGGTCAGGCGGG
ATTACCCGCTGAGTTTAAGCATATCAATAAGCGGAGGAAAAGAAACTAACAA
GGATTCCC NTAGTAACGG NGNGNGNAAN NGNNGGAAAT ACTACNNN

LOCUS *V. faik-karaveliogullari* 867 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNTNTTNNNNTNNNGNNNACNTNNNCCNT
TCGCTTCCCTTTACGTNTCGTGCTGTCCACTGAACCTTATCATTTAGAGGAAGG
AGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAA

ACCTGCAAAGCAGACCGCGAACTTGTGTTTTTAACTATATCGGGGGCCTCGGC
CCCCTCCCCTGCATGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAACGAACC
CCCGGCGCGGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCTCGTCGATGC
CCCGTTCGCGGTGTGCTCGGCGGGACAGTGCGTCTCTTGAATGTCAAAACGAC
TCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATGC
GATACTTGGTGTGAATT
GCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCCGAAGCCAT
TAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCCCCTCC
AGCTCCTTCGGGTGCAATGGAGAGTGGGGCGGATATTGGCCTCCCGTGTGCTT
CGGTGCGCGGCCGGCCCAAATGTGATCCCTCGTCGACGCGTGTCACGACCAGT
GGTGGTTGATACATCAACTCGCGTGCTGTCGTGCCTCATCGCGTCGCTCGTTAG
GGCATCGTCGCAGACCCAATAGGCGCTCTCGAGCGCCTTCGACCGCGACCCCA
GGTCAGGCGGGATTACCCGCTGAGTTTAAGCATATCAATAAGCGGAGGAAAA
GAAACTAACAAGGATTCCCTTAGTAACGGCNANCNAACCGAACGAAGTANTA
CNGG

LOCUS *V. glomeratum Boiss* 874 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNTNTTNNNNTNNNGNNNACNTNNNCCNT
TCGCTTCCCTTTACGTNTCGTGCTGTCCACTGAACCTTATCATTTAGAGGAAGG
AGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAA
ACCTGCAAAGCAGACCGCGAACTTGTGTTTTTAACTATATCGGGGGCCTCGGC
CCCCTCCCCTGCATGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAACGAACC
CCCGGCGCGGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCTCGTCGATGC
CCCGTTCGCGGTGTGCTCGGCGGGACAGTGCGTCTCTTGAATGTCAAAACGAC
TCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATGC
GATACTTGGTGTGAATT
GCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCCGAAGCCAT
TAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCCCCTCC
AGCTCCTTCGGGTGCAATGGAGAGTGGGGCGGATATTGGCCTCCCGTGTGCTT
CGGTGCGCGGCCGGCCCAAATGTGATCCCTCGTCGACGCGTGTCACGACCAGT
GGTGGTTGATACATCAACTCGCGTGCTGTCGTGCCTCATCGCGTCGCTCGTTAG

GGCATCGTCGCAGACCCAATAGGCGCTCTCGAGCGCCTTCGACCGCGACCCCA
GGTCAGGCGGGATTACCCGCTGAGTTTAAGCATATCAATAAGCGGAGGAAAA
GAAACTAACAAGGATTCCCTTAGTAACGGCNANCNAACCGAACGAAGTANTA
CNGG

LOCUS *V. isauricum Boiss&Heldr* 871 bp DNA linear

NNNNNNNNNNNNNCNNNNNNNNGTTCNNNTNGNTNNNNNNTANGCATNNTT
CTGCATGANNNNNNAGCANCNGTCCACTGAACCTTATCATTTAGAGGAAGGA
GAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAAA
CCTGCAAAGCAGACCGCGAACTTGTGTTTTTAACTATATCGGGGGCCTCGGCC
CNNNCCCCCNCACGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAACGAANN
CCCGGCGCGGCATGCGCCAAGGAAAANNCAACGAAGCGCCTCCCCGTCGATG
CCCCGTTCGCGGTGTGCTCGGCNNGACAGTGCNTCTCTTGAATGTCAAAACGA
CTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATG
CGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCA
AGTTGCGCCCGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCAT
CGCGTCGCCCCCCCNCTCCANCTCCTTCGGGTGCGATGGAGAGNNGGGGCGGAT
ATTGGCCTCCCGTGCGCCTCGGTGCGCGGCCGGCNCNAATGTGATCCCTCGTC
GACGCGTGTCACGACCAGTGGTGGTTGAACNTCAACTCGCGTGCTGTCGTGCC
TCATCGCGTCGCTCGTCANGGCATCGTCGCAGACCCAATAGGCGCTCTCGAGC
GCCTTCGACCGCGACCCCAGGTCNNGCGGGATTACCCGCTGAGTTTAAGCATA
TCAATAAGCGGAGGAAAAGAAACTAACAAGGATTCCNNNNNNNNNNNNNNN
NNNNNNNNACNNNNNNNANNNNNNNNNN

LOCUS *V. sinatum* 875 bp DNA linear

NNNNNNNCCCNCNNTNCTTCGTTCAGTTACGTATTCGTATTGCGGTTAAGTTCC
TCGTGCAGTGTCCAAGAGATCAACACCTGTCCACTGAACCTTATCATTTAGAG
GAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATT
GTCGAAACCTGCAAAGCAGACCGCGAACTCGTGTTTTTAACTATATCGGGGGC

CTCGGCCCCCTCCCCCNCACGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAA
CGAACCCCCGGCGCGGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCCCG
TCGATGCCCCGTTCGCGGTGTGCTCGGCGGGACAGTGCGTCTCTTGAATGTCA
AAACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGC
GAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTG
AACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCGTC
ACGCATCGCGTCGCCCCCCCCTCCACCTCCTTCGGGTGCGATGGAGAGTGGGG
GCGGATATTGGCCTCCCGTGCGCCTCGGTGCGCGGCCGGCCCAAATGTGATCC
CTCGTCGACGCGTGTCACGACCAGTGGTGGTTGATACNTCAACTCGCGTGCTG
TCGTGCCTCATCGCGTCGCTCGTCAGGGCATCGTCGCAGACCCAATAGGCGCT
CTCGAGCGCCTTCGACCGCGACCCCAGGTCAGGCGGGATTACCCGCTGAGTTT
AAGCATATCAATAAGCGGAGGAAAAGAAACTAACAAGGATTCCCTTAGTAAC
GGCGAGCGAACCGAAGGAANTTCTANAGG

LOCUS *V. speciosum* 877 bp DNA linear

NNNNNNNNNNNNNNNNCTTCGTTCCNTTCCCTGTTCANTTCGCNGTTACGTG
TGNNGTTTGCTTCCCGAGAACTCATCAGCACGTCCACTGAACCTTATCATTTAG
AGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA
TTGTCGAAACCTGCAAAGCAGACCGCGAACTCGTGTTTTTAACTATATCGGGG
GCCTCGGCCCCCTCCCCCNCACGGCGAGCGCGCGAGCGACCGTCGTGCGGGCT
AACGAACCCCCGGCGCGGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCC
CGTCGATGCCCCGTTCGCGGTGCGCTCGGCGGGACAGTGCGTCTCTTGAATGT
CAAAACGACTCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTA
GCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTT
TGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCG
TCACGCATCGCGTCGCCCCCCCCTCCAGCTCCTTCGGGTGCGATGGAGAGCCG
GGGCGGATATTGGNCTCCCGTGCGCCTCGGTGCGCGGCCGGCCCAAATGTGAT
CCCTCGTCGACGCGTGTCACGACCAGTGGTGGTTGATACTTCAACTCGCGTGCT
GTCGTGCCTCATCGCGTCGCTCGTCGGGGCATCGTCGCAGACCCAACAGGCGC
TCTCGAGCGCCTTCGACCGCGACCCCAGGTCAGGCGGGATTACCCGCTGAGTT

TAAGCATATCAATAAGCGGAGGAAAAGAAACTAACAAGGATTCCCTTAGTAA
CGGTGAGCGAACAGGACGAAATTGTACAGG

LOCUS *V. tuna-ekimii Karavel* 868 bp DNA linear

NNNNNNNNNNNNNNNCNNNTNNNTGTTCGTNTTGCNGNTANGTGAGCNNCT
CATTGTCCNNNTNCTCTNCANCNGTCCACTGAACCTTATCATTTAGAGGAAGG
AGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTGTCGAA
ACCTGCAAAGCAGACCGCGAACTTGTGTTTTTAACTATATCGGGGGCCTCGGC
CCCNTCCCCCGCACGGCGAGCGCGCGAGCGACCGTCGTGCGGGCTAACGAACC
CCCGGCGCGGCATGCGCCAAGGAAAACTCAACGAAGCGCCTCCCCGTCGATGC
CCCGTTCGCGGTGTGCTCGGCGGGACAGTGCGTCTCTTGAATGTCAAAACGAC
TCTCGGCAACGGATATCTAGGCTCTCGCATCGATGAAGAACGTAGCGAAATGC
GATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTTTTTGAACGCAA
GTTGCGCCCGAAGCCATTAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCATC
GCGTCGCCCCCCCCTCCANCTCCTTCGGGTGCGATGGAGAGTGGGGGCGGATA
TTGGCCTCCCGTGCGCCTCGGTGCGCGGCCGGCCCAAATGTGATCCCTCGTCG
ACGCGTGTCACGACCAGTGGTGGTTGATACNTCAACTCGCGTGCTGTCGTGCC
TCATCGCGTCGCTCGTCAGGGCATCGTCGCAGACCCAATAGGCGCTCTCGAGC
GCCTTCGACCGCGACCCCAGGTCAGGCGGGATTACCCGCTGAGTTTAAGCATA
TCAATAAGCGGAGGAAAAGAAACTAACAAGGATTCCCTTAGTAACGGNGANN
GAACNNNNNGAANTANTACNNN

***ITS* Sanger**

LOCUS *V. blattaira* 776 bp DNA linear

TTCCGGTTTTATTATAGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTATTGATAT
GCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGCT
CGAGAGCGCCTATTGGGTCTGCGACGATGCCCTACGAGCGACGCGATGAGGCA
CGACAGCACGCGAGTTGAGGTATCAACCACCACTGGTCGTGACACGCGTCGAC
GAGGGATCACATTTGGGCCGGCCGCGCACCGAAGCGCACGGGAGGCCAATAT

CCGCCCCCGCTCTCCATCGCACCCGAAGGAGCAGGAGGAGGGGGGCGACGCG
ATGCGTGACGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACT
TGCGTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCG
CATTTCGCTACGTTCTTCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGT
CGTTTTGACATTCAGGAGACGCACTGCCCCGCCGAGCGCACCGCGAACGGGGC
ATCGACGGGGAGGCGCTTCGTTGAGTTTTCCTTGGCGCATGCCGCGCCGGGGG
TTCGTTAGCCCGCACGACGCTCGCTCGCGCGCACGCCGTGCGGGGGAGGGGGC
CGAGGCCCCCGGTATAGTTAAAAACACAAGTTCGCGGGTCTGCTTTGCAGGTT
TCGACAATGATCCTTCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCC
TTCCTCTAAATGATAAGGTTCCAGTGGACAAA

LOCUS *V. glomeratum* 717 bp DNA linear

GTTATCTAAGGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTATTGATATGCTTAA
ACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGCTCGAGAA
CGCCTATtGGGTTTGcGACGATGCCCTAACGAGCGACGCGATGAGGCACGACA
GCACGCGAGTTGAAGTATCAACCACCACTGGTCGTGACACGCGTCGACGAGGG
ATCACATTTGGGCCGGCCGCGCACCGAGGCGCACGGGAGGCCAATATCCGCCC
CCACTCTCCATCGCACCCGAAGGAGCTGGAGGGGGGGCGACGCGATGCGTGA
CGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCTTCGCTACGTTCT
TCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGTCGTTTTGACATTCAA
GAGACGCACTGTCTCCGCGCGAGTCACATCCGTCAAACGGGGCATCGACGGGG
AGGCGCTTCGTTGAGTTTTCCTGGGCGCATGCCGCGCCGGGGGTTCGTTACCCC
GCACAACGGCCGCTCGCGCGCTCGCCGTGCGGGGGAGGGGGCCAAGGCCCCC
GATATAGTAAAAAACACAAAGTTCGCGGTCTGCTTTGCAGGTTTCGACAATGA
TCCTTCCGCAGGTTCACCT

LOCUS *V. hasbenlii* 781 bp DNA linear

TTTTTTCCGTTTTATCTAGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTATTGATA
TGCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGC

TCAAGAGCACCTAATGGGTCTGCGACGATGCCCTGCGAGCGACGCGATGAGGC
ACGACAGCACGCGAGTTGAGGTATCAACCACCACTAGTCGTGACATGCGTCGA
CGCGGGATCACATTTGGGCCGGCCGCGCACCGAGGCGCACGGGAGGCCAATA
TCCGCCCCACTCTCCATCGCACCCGAAGGAGCTGGAGGGGGGGCGACGCGATG
CGTGACGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACTTGC
GTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCGCAT
TTCGCTACGTTCTTCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGTCGT
TATGACATTCAAGAGACGCACTGTCCCGCCGAGCACACCGCGAACGGGGCATC
GACGGGGAGGCGATTCGTTGAGTTTTCCTTGGCGCATGCCGCGCCGGGGGTTC
GTTAGCCCGCACAACGGTCGCTCGCGCGCACGCCGTGCGGGGGAGAGGGCCG
AGGCCCCCAATATAGTTAAAAACACGAGTTCGCGGTCTGCTTTGCAGGTTTCG
ACAATGATCCTTCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCCTTC
CTCTAAATGATAAGGTCCAAGGGGAAAAAACAAAGT

LOCUS *V. luridiflorum* 775 bp DNA linear

TTTTCCTTTTTTAATATAGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTATTGATA
TGCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGC
TCGAGAGCGCCTATTGGGTCTGCGACGATGCCCTGACGAGCGACGCGATGAGG
CACGACAGCACGCGAGTTGAAGTATCAACCACCACTGGTCGTGACACGCGTCG
ACGAGGGATCACATTTGGGCCGGCCGCGCACCGAAGCGCACGGGAGGCCAAT
ATCCGCCCCGCTCTCCATCGCACCCGAAGGAGCTGGAGGGGGGGCGACGCGAT
GCGTGACGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACTTG
CGTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCGCA
TTTCGCTACGTTCTTCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGTCG
TTTTGACATTCAAGAGACGCACTGTCCCGCCGAGCACACCGCGAACGGGGCAT
CGACGGGGAGGCGCTTCGTTGAGTTTTCCTTGGCGCATGCCGCGCCGGGGGTT
CGTTAGCCCGCACGACGGTCGCTCGCGCGCTCGCCGTGCGGGGGAGGGGGCCG
AGGCCCCCGATATAGTTAAAAACACAAGTTCGCGGTCTGCTTTGCAGGTTTCG
ACAATGATCCTTCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCCTTC
CTCTAAATGATAGGTTCCCGGGGAAAAAA

LOCUS *V. ovalifolium ssp thraricum (Velen) Murb* 738 bp DNA linear

TTTTTTCTCGCTTATTGATATGCTTAACTCAGCGGGTAATCCCGCCTGACCTGG
GGTCGCGGTCGAAGGCGCTCGAGAGCGCCTATTGGGTCTGCGACGATGCCCTG
CCAACGAACCCGTTAAGGAACACCAGCCCGCAGTTGAATAATCACCCCCCATG
GGCCGGGCCCCGGGCCAACAAGGAATCACTTTGGGGCCGGCCGCCCACCGAG
GGGCAGGGAAGGCCAATACCCGCCCCCACTCCCCTTCGCCCCCAAGGAATTTG
GGGGGGGGGGCGACGCGAGGGGGGACCCCGGGGAAAAGGCCCCCTCGCCATA
GGGTCTTGGGGCACATTTGGTTTAAAAAACCTCGGGGTTCCGGGGTTTTCGGA
TTTCCACCCAAATTACCATTTTCCTCTATTTTCTTCCTCGAGGGAAGACCCAAA
ATCCCTTTTCCAAAAAGTTTTTTTAATTTTAAAAAAACCCTTTGCCCCGCAAAC
CACCCCGAAAGGGGGTCTTCGGGGGAAAGGCTTTTTTTAAATTTTTCTTGGCCC
AACGCCCCCCGGGGTTTTCTTTACCCCCCACGAGGGCCCCTCGCGCCCTCCCGG
GGGGGAGGAGGGCCCAGACCCCCCCAATATATTAAAAAACACATTTCCGCGTT
CTGTTTTGCGGTTTTCAAAAATGATCTTTCCGGAGGTTCCCCTACGAAAACTTT
TGTCACATCTTCTCTTTCCTAAAATGAAAGTTTTGGGGGAAAAA

LOCUS *V. ovalifolium ssp thracicum Donn ex Sims* 775 bp DNA linear

TTTTTCCGTTTTTATTAGGAATCCTTGTTAGTTTCTTTTCTTCCGCTTATTGATAT
GCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGCc
CGAGAGCGCCTGTTGGGTCTGCGACGATGCCCCGACGAGCGACGCGATGAGG
CACGACAGCACGCGAGTTGAAGTATCAACCACCACTGGTCGTGACACGCGTCG
ACGAGGGATCACATTTGGGCCGGCCGCGCACCGAGGCACACGGGAGGCCAAT
ATCCGCCCCCGCTCTCCATCGCACCCGAAGGAGCTGGAGGGGGGGCGACGCG
ATGCGTGACGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACT
TGCGTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCG
CTTTCGCTACGTTCTTCATCGATGCGAGAGCCTAgATATCCGTTGCCGAGAGTC
GTTTTGACATTCAAGAgACGCACTGTCCCGCCGAgCACACCGCGAACGGGGCA
TCGACGGGGAGGCGCTTCGTTGAgTTTTCCTTGGCGCATGCCGCGCCGGGGGTT
CGTTAGCCCGCACGACGGTCGCTCGCGCGCTCGCCGTGCGGGGGAGGGGGCCG
AgGCCCCCGATATAGTTAAAAACACGAGTTCGCGGTCTGCTTTGCAGGTTTCGA

CAATGATCCTTCCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCCTTC
CTCTAAATGATAAGGTTCAAGTGGACAA

LOCUS *Verbascum vulcanicum Boiss & Hedr var vulvanicum Boiss & Hedr* 779 bp DNA linear

TTTGTCCGTTTTTTTAAGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTAtTGATAT
GCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGCT
CGAGAGCGCCTGTTGGGTCTGCGACGATGCCCTGACGAGCGACGCAATGAGGC
ACGACAGCACGCGAGTTGAAGTATCAACCACCACTGGTCGTGACACGCGTCGA
CgAGGGATCACATTTGGGCCGGCCGCGCACCGAGGCGCACGGGAGGCCAATAT
CCGCCCCGCTtTCCATCGCACCCGAAGGAGCTGGAGGGGGGGGCGACGCGATG
CGTGACGCCCAGGCAgACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACTTGC
GTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCGCAT
TTCGCTACGTTCTTCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGTCGT
TTTGACATTCAAGAGACGCACTGTCCCGCCgAGCACACCGCgAACGGGGCATC
GACGGGGAGGCGCTTCGTTGAGTTTTCCTTGGCGCATGCCGCGCCGGGGGTTC
GTTACCCCGCACgACGGTCGCTCGCGCGCTCGCCGTGCGGGGGAGGGGGCCGA
GGCCCCCGATATAGTTAAAAACACAAGTTCGCGGTCTGCTTTGCAGGTTTCGA
CAATGATCCTTCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCCTTCC
TCTAAATGATAAGTCTCAAGTGGAAAAAAGCG

LOCUS *V. vulcanicum* 791 bp DNA linear

TTTTCCGTTTTCTCTATAGGAATCCTTGTTAGTTTCTTTTCCTCCGCTTATTGATA
TGCTTAAACTCAGCGGGTAATCCCGCCTGACCTGGGGTCGCGGTCGAAGGCGC
TCGAGAGCGCCTATTGGGTCTGCGACGATGCCCTAACGAGCGACGCGATGAGG
CACGACAGCACGCGAGTTGAAGTATCAACCACCACTGGTCGTGACACGCGTCG
ACGAGGGATCACATTTGGGCCGGCCGCGCACCGAGGCGCACGGGAGGCCAAT
ATCCGCCCCCACTCTCCATCGCACCCGAAGGAGCTGGAGGGGGGGCGACGCG
ATGCGTGACGCCCAGGCAGACGTGCCCTCGGCCTAATGGCTTCGGGCGCAACT

TGCGTTCAAAAACTCGATGGTTCACGGGATTCTGCAATTCACACCAAGTATCG
CATTTCGCTACGTTCTTCATCGATGCGAGAGCCTAGATATCCGTTGCCGAGAGT
CGTTTTGACATTCAAGAGACGCACTGTCCCGCCGAGCACACCGCGAACGGGGC
ATCGACGGGGAGGCGCTTCGTTGAGTTTTCCTTGGCGCATGCCGCGCCGGGGG
TTCGTTAGCCCGCACGACGGTCGCTCGCGCGCTCGCCGTGCGGGGGAGGGGGC
CGAGGCCCCCGATATAGTTAAAAACACAAGTTCGCGGTCTGCTTTGCAGGTTT
CGACAATGATCCTTCCGCAGGTTCACCTACGGAAACCTTGTTACGACTTCTCCT
TCCTCTAAATGATAAGTTCCAAGGGGAAAAAACAGCAAAAGCAACC

***matK* NGS**

LOCUS *V. alyssifolium* 1125 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNCTTCNTNCNNTTNNNTTTTTCGNAGTTAAANTTTCTCTNACA
TGTATCAATACCCTACCCAGTCCATCTGGAAATCTTGGTTCAAACTCTTCGTTA
TTGGTTAAAAGACGCCTCTTCTTTGCATTTATTACGATTCTTTCTCAATGAGTAT
TGTAATTGGAATAGGCTTATTACTCCAAAGAAAGCCAGTTCCTCCTTTTCAAAA
AGAAATCAAAGGTTATTCTTATTCTTATATAATTATCATGTCTGTGAATACGAA
TCCATTTTCGTCTTTCTACGTAATCAATCTTCTCATTTAAGATCAACATCTTCTG
GAGTTTTTCTTGAACGAATCTATTTCTATGGAAAAGTAGAACGTCTTGTGAACG
TCTTTGGTAAAGTTAAGGATTTTCAGGAGAACCTATGGTTGGTCAAGGAACCT
TGCATGCATTATGTTAGGTATCAAAGAAAATCCATTCTGGCTTCAAAAGGGAC
GTCTCTTTTAATGAATAAATGGAAATGTTACCTTGTCACTTTTTGGCAATGGTA
TTTTTCGCTTTGGTTTCATCCAAGAAGGATTTATATAAACCAATTATCCAATCA
TTCCCTTGAATTTTTGGGCTATCTTGCAAGCGTGCGAATGAACCCTTCGGTGGT
ACGGAGTCAAATTCTAGAAAATTCATTTCTAATCAATAATGCTATTAAGAAGT
TCGATACTCTTGTTCCAATTATTCCTCTGATTGCGTCATTGGCTAAAGCTAAAT
TTTGTAACTTATTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTGATTTATCAG
ATTCTAATATTATTGATCGATTTGGGCGTATATGCAGAAACCTTTCTCATTATC
ATAGCGGATCTTCCAAAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGA
TTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGTGCTGCT

TTGNATACATNAGGCAGTGGTTTAACCTTAGCAATACGTATGANGGAAGANNN
NNGNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNN

LOCUS *V. faik-karaveliogullari* 1123 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTN
NNNNNTTNNNNNNNNNTNTTTTTNNTNNNCCCCCCNTNNCTTCGTTCNGTTAGT
ATGCTACCCAGTCCATCTGGAAATCTTGGTTCAAACTCTTCGTTATTGGTTAAA
AGACGCCTCTTCTTTGCATTTATTACGATTCTTTCTCAATGAGTATTGTAATTGG
AATAGGCTTATTACTCCAAAGAAAGCCAGTTCCTCCTTTTCAAAAAGAAATCA
AAGGTTATTCTTATTCTTATATAATTATCATGTCTGTGAATACGAATCCATTTTC
GTCTTTCTACGTAATCAATCTTCTCATTTAAGATCAACATCTTCTGGAGTTTTTC
TTGAACGAATCTATTTCTATGGAAAAGTAGAACGTCTTGTGAACGTCTTTGGTA
AAGTTAAGGATTTTCAGGAGAACCTATGGTTGGTCAAGGAACCTTGCATGCAT
TATGTTAGGTATCAAAGAAAATCCATTCTGGCTTCAAAAGGGACGTCTCTTTTA
ATGAATAAATGGAAATGTTACCTTGTCACTTTTTGGCAATGGTATTTTTCGCTT
TGGTTTCATCCAAGAAGGATTTATATAAACAATTATCCGATCATTCCCTTGAAT
TTTTGGGCTATCTTGCAAGCGTGCGAATGAACCCTTCGGTGGTACGGAGTCAA
ATTCTAGAAAATTCATTTCTAATCAATAATCCTATTAAGAAGTTCGATACTCTT
GTTCCAATTATTCCTCTGATTGCGTCATTGGCTAAAGCTAAATTTTGTAACCTA
TTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTGATTTATCAGATTCTAATATT
ATTGATCGATTTGGGCGTATATGCAGAAACCTTTCTCATTATCATAGCGGATCT
TCCAAAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGATTTTCTTGTGCT
AGAACTTTGGCTCGTAAACACAAAAGTACTGTACGAGCAATACNGAACGAAG
NAANGGGGGGGTTNNNGTNNNNNNTACNTANNNANNANGNANNNNNNNNNN
NN

LOCUS *V. glomeratum Boiss* 1098 bp DNA linear

NNNNNNNNNCNNCNNNTNNNTTTTTNCNTTTNNNTTTTNCTAAGTTAANTTTA
CTCCNAGNNNNCCTCACNCCCTACCCAGTCCATCTGGAAATCTTGGTTCAAAC
TCTTCGTTATTGGTTAAAAGACGCCTCTTCTTTGCATTTATTACGATTCTTTCTC
AATGAGTATTGTAATTGGAAAAGGCTTATTACTCCAAAGAAAGCCAGTTCCTC
CTTTTCAAAAAGGAATCAAAGGTTATTCTTATTCTTATATAATTATCATGTCGG
TGAATACGAATCCATTTTCGTCTTTCTACGTAATCAATCTTCTCATTTAAGATC
AACATCTTCTGGAGTTTTTCTTGAACGAATCTATTTCTATGGAAAAGTAGAACG
TCTTGTGAACGTCTTTGGTAAAGTTAAGGATTTTCAGGAGAACCTATGGTTGGT
CAAGGAACCTTGCATGCATTATGTTAGGTATCAAAGAAAATCCATTCTGGCTT
CAAAAGGGACGTCTCTTTTAATGAATAAATGGAAATGTTACCTTGTCACTTTTT
GGCAATGGTATTTTTCGCTTTGGTTTCATCCAAGAAGGATTTATATAAACCAAT
TATCCAATCATTCCCTTGAATTTTTGGGCTATCTTGCAAGCGTGCGAATGAACC
CTTCGGTGGTACGGAGTCAAATTCTAGAAAATTCATTTCTAATCAATAATGCTA
TTAAGAAGTTCGATACTCTTGTTCCAATTATTCCTCTGATTGCGTCATTGGCTA
AAGCTAAATTTTGTAACCTATTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTG
ATTTATCAGATTCTAATATTATTGATCGATTTGGGCGTATATGCAGAAACCTTT
CTCATTATCATAGCGGATCTTCCAAAAAAAAGAGTTTGTATCGAATAAAGTAT
ATACTTCGATTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTA
CGAGGTGCTGCGAGGAGGTTTGGNTAGTAATTAACCTTAGCAATACGTANGAA
NGAAGAANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNN

LOCUS *V. isauricum Boiss&Heldr* 1042 bp DNA linear

NNNNNNNNNNNNNNTNNTTNCNNTTNCNTATTGCTANGTTAAGCATGTTCTG
CATGATGGGTTANNNNCCNACCCAGTCCATCTGGAAATCTTGGTTCAAACTCT
TCGTTATTGGTTAAAAGACGCCTCTTCTTTGCATTTATTACGATTCTTTCTCAAT
GAGTATTGTAATTGGAANAGGCTTATTACTCCAAAGAAAGCCAGTTCCTCCTTT
TCAAAAAGAAATCAAAGGTTATTCTTATTCTTATATAATTATCATGTCTGTGAA
TACGAATCCATTTTCGTCTTTCTACGTAATCAATCTTCTCATTTAAGATCAACAT

CTTCTGGAGTTTTTCTTGAACGAATCTATTTCTATGGAAAAGTAGAACGTCTTG
TGAACGTCTTTGGTAAAGTTAAGGATTTTCAGGAGAACCTATGGTTGGTCAAG
GAACCTTGCATGCATTATGTTAGGTATCAAAGAAAATCCATTCTGGCTTCAAA
AGGGACGTCTCTTTTAATGAATAAATGGAAATGTTACCTTGTCACTTTTTGGCA
ATGGTATTTTTCGCTTTGGTTTCATCCAAGAAGGATTTATATAAACCAATTATC
CAACCATTCCCTTGAATTTTTGGGCTATCTTGCAAGCGTGCGAATGAACCCTTC
GGTGGTACGGAGTCAAATTCTAGAAAATTCATTTCTAATCAATAATGCTATTA
AGAAGTTCGATACTCTTGTTCCAATTATTCCTCTGATTGCGTCATTGGCTAAAG
CTAAATTTTGTAACCTATTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTGATT
TATCAGATTCTAATATTATTGATCGATTTGGGCGTATATGCAGAAACCTNTCTC
ATTATCATAGCGGATCTTCCAAAAAAAAGAGTTTGTATCGAATAAAGTATATA
CTTCGATTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGA
GNTNNNNNNNNNNNNNNNNNNNNAANNNTNNNNANNNNTN
NNNNNNNNNNANNNNNNNNNNNNNNNNNNN

LOCUS *V. sinatum* 1223 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NTTCNTNCTCCNCNTNACTTCNTTCAGTTACGTATTGCTAAGGTTAAGTTCCTC
GTAGAGAGATCAANAACCTACCCAGTCCATCTGGAAATCTTGGTTCAAACTCT
TCGTTATTGGTTAAAAGACGCCTCTTCTTTGCATTTATTACGATTCTTTCTCAAT
GAGTATTGTAATTGGAATAGGCTTATTACTCCAAAGAAAGCCAGTTCCTCCTTT
TCAAAAAGGAATCAAAGGTTATTCTTATTCTTATATAATTATCATGTCTGTGAA
TACGAATCCATTTTCGTCTTTCTACGTAATCAATCTTCTCATTTAAGATCAACAT
CTTCTGGAGTTTTTCTTGAACGAATCTATTTCTATCGAAAAGTAGAACGTCTTG
TGAACGTCTTTGGTAAAGTTAAGGATTTTCAGGAGAACCTATGGTTGGTCAAG
GAACCTTGAATGAATAAATGGAAATCTTACCTTGTCACTTTTTGGCAATGGTAT
TTTTCGCTTTGGTTTCATCCAAGAAGGATTTATATAAACCAATTATCCAATCAT
TCCCTTGAATTTTTGGGCTATCTTGCAAGCGTGCGAATGAACCCTTCGGTGGTA
CGGAGTCAAATTCTAGAAAATTCATTTCTAATCAATAATGCTATTAAGAAGTTC
GATACTCTTGTTCCAATTATTCCTCTGATTGCGTCATTGGCTAAAGCTAAATTTT

GTAACCTATTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTGATTTATCAGATT
CTAATATTATTGATCGATTTGGGCGTATATGCAGAAACCTTTCTCATTATCATA
GCGGATCTTCCAAAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGATTT
TCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGTGCTGNTTTN
NTGACACTGCACGAGGAATTAACCTTAGCAATACGTANGAAGNAAGAAGAGG
GGGGGGGNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

LOCUS *V. speciosum* 1094 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNCNNNTTACTTCTTTCA
GTTACGTATTGCTAAGGTTAATTTCCTGTTAGAACTCATCAGCACCTACCCAGT
CCATCTGGAAATCTTGGTTCAAACTCTTCGTTATTGGTTAAAAGACGCCTCTTC
TTTGCATTTATTACGATTCTTTCTCAATGAGTATTGTAATTGGAATAGGCTTATT
ACTCCAAAGAAAGCCAGTTCCTCCTTTTCAAAAAGGAATCAAAGGTTATTCTT
ATTCTTATATAATTATCATGTCTGTGAATACGAATCCATTTTCGTCTTTCTACGT
AATCAATCTTCTCATTTAAGATCAACATCTTCTGGAGTTTTTCTTGAACGAATC
TATTTCTATGGAAAAGTAGAACGTCTTGTGAACGTCTTTGGTAAAGTTAAGGA
TTTTCAGGAGAACCTATGGTTGGTCAAGGAACCTTGCATGCATTATGTTAGGTA
TCAAAGAAAATCCATTCTGGCTTCAAAAGGGACGTCTCTTTTAATGAATAAAT
GGAAATGTTACCTTGTCACTTTTTGGCAATGGTATTTTTCGCTTTGGTTTCATCC
AAGAAGGATTTATATAAACCAATTATCCAATCATTCCCTTGAATTTTTGGGCTA
TCTTGCAAGCGTGCGAATGAACCCTTCGGTGGTACGGAGTCAAATTCTAGAAA
ATTCATTTCTAATCAATAATGCTATTAAGAAGTTCGATACTCTTGTTCCAATTA
TTCCTCTGATTGCGTCATTGGCTAAAGCTAAATTTTGTAACCTATTAGGGCATC
CCATTAGTAAGCCGGTTTGGGCTGATTTATCAGATTCTAATATTATTGATCGAT
TTGGGCGTATATGCAGAAACCTTTCTCATTATCATAGCGGATCTTCCAAAAAA
AAGAGTTTGTATCGAATAAAGTATATACTTCGATTTTCTTGTGCTAGAACTTTG
GCTCGTAAACACAAAAGTACTGTACGTGCTGTTGTGTTATNATTAAGGACGCA
ATTAACCTNAGAAATACGTATGAAGGAAGAAGNNGGNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNN

LOCUS *V. tuna-ekimii Karavel* 1206 bp DNA linear

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNNNNNNN
NNTTTTTTNTNTNNTATTTTTNNNNTTAANAGCCTCTCATTGTTCTCTNNNGCN
CCTACCCAGTCCATCTGGAAATCTTGGTTCAAACTCTTCGTTATTGGTTAAAAG
ACGCCTCTTCTTTGCATTTATTACGATTCTTTCTCAATGAGTATTGTAATTGGAA
TAGGCTTATTACTCCAAAGAAAGCCAGTTCCTCCTTTTCAAAAAGAAATCAAA
GGTTATTCTTATTCTTATATAATTATCATGTCTGTGAATACGAATCCATTTTCGT
CTTTCTACGTAATCAATCTTCTCATTTAAGATCAACATCTTCTGGAGTTTTTCTT
GAACGAATCTATTTCTATGGAAAAGTAGAACGTCTTGTGAACGTCTTTGGTAA
AGTTAAGGATTTTCAGGAGAACCTATGGTTGGTCAAGGAACCTTGCATGCATT
ATGTTAGGTATCAAAGAAAATCCATTCTGGCTTCAAAAGGGACGTCTCTTTTA
ATGAATAAATGGAAATGTTACCTTGTCACTTTTTGGCAATGGTATTTTTCGCTT
TGGTTTCATCCAAGAAGGATTTATATAAACCAATTATCCGATCATTCCCTTGAA
TTTTTGGGCTATCTTGCAAGCGTGCGAATGAACCCTTCGGTGGTACGGAGTCA
AATTCTAGAAAATTCATTTCTAATCAATAATCCTATTAAGAAGTTCGATACTCT
TGTTCCAATTATTCCTCTGATTGCGTCATTGGCTAAAGCTAAATTTTGTAACCT
ATTAGGGCATCCCATTAGTAAGCCGGTTTGGGCTGATTTATCAGATTCTAATAT
TATTGATCGATTTGGGCGTATATGCAGAAACCTTTCTCATTATCATAGCGGATC
TTCCAAAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGATTTTCTTGTGC
TAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGGTGCTGTAGAGAAACAA
TGAGGCTCTTAANTATNGCAANAANTACGTANGAANGAAGAAGNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

***matK* Sanger**

LOCUS *V. hasbenlii* 905 bp DNA linear

TTTCCCGACACAATATACTTGTGTTTACGAGCCAAAGTTCTAGCACAAGAAAA
TCGAAGTATATACTTTATTCGATACAAACTCTTTTTTTTGGAAGATCCGCTATG
ATAATGAGAAAGGTTTCTGCATATACGCCCAAATCGATCAATAATATTAGAAT

CTGATAAATCAGCCCAAACCGGCTTACTAATGGGATGCCCTAATAGGTTACAA
AATTTAGCTTTAGCCAATGACGCAATCAGAGGAATAATTGGAACAAGAGTATC
GAACTTCTTAATAGCATTATTGATTAGAAATGCATTTTCTAGAATTTGACTCCG
TACCACCGAAGGGTTCATTCGCACGCTTGCAAGATAGCCCAAAAATTCAAGGG
AATGATTGGATAATTGGTTTATATAAATTCTTCTTGGATGAAACCAAAGCGAA
AAATACCATTGCCAAAAAGTGACAAGGTAACATTTCCATTTATTCATTAAAAG
AGACGTCCCTTTTGAAGCCAGAATGGATTTTCTTTGATACCTAACATAATGCAT
GCAAGGTTCCTTGATCAACCATAGGTTCTCCTGAAAATCCTTAACTTTACCAAA
GACGTTCACAAGACGTTCTACTTTCCCATAGAAATAGATTCGTTCAAGAAAAA
CTCCAGAAGATGTTGATCTTAAATGAGAAGATTGATTACGTAGAAAGACGAAA
ATGGATTCGTATTCACATACATGATATTATATAAGAATAAGAATAACCTTTGAT
TTCTTTTTGAAAAGGAGGAACTGGCTTTCTTTGGAGTAATAAGCCTATTCCAAT
TACAATACTCATTGAGAAAGAATCGTAATAAATGCAAAGAAGAGGCGTCTTTT
AACCAATAACGAAGAGTTTGAACCAAGATTTCCAGATGGGACTGGGTAAA

LOCUS *V. luridiflorum* 914 bp DNA linear

AATTTTCGTTACCAGTTACTTGTGTTTACGAGCCAAAGTTCTAGCACAAGAAAA
TCGAAGTATATACTTTATTCGATACAAACTCTTTTTTTTGGAAGATCCGCTATG
ATATGAGAAAGGTTTCTGCATATACGCCCAAATCGATCAATAATATTAGAATC
TGATAAATCAGCCCAAACCGGCTTACTAATGGGATGCCCTAATAGGTTACAAA
ATTTAGCTTTAGCCAATGACGCAATCAGAGGAATAATTGGAACAAGAGTATCG
AACTTCTTAATAGCATTATTGATTAGAAATGAATTTTCTAGAATTTGACTCCGT
ACCACCGAAGGGTTCATTCGCACGCTTGCAAGATAACCCAAAAATTCAAGGGA
ATGATTGGATAATTGGTTTATATAAATCCTTCTTGGATGAAACCAAAGCGAAA
AATACCATTGCCAAAAAGTGACAAGGTAACATTTCCATTTATTCATTAAAAGA
GACGTCCCTTTTGAAGCCAGAATGGATTTTCTTTGATACCTAACATAATGCATG
CAAGGTTCCTTGACCAACCATAGGTTCTCCTGAAAATCCTTAACTTTACCAAAG
ACGTTCACAAGACGTTTTACTTTTCCATAGAAATAGATTCGTTCAAGAAAAACT
CCAGAAGATGTTGATCTTAAATGAGAAGATTGATTACGTAGAAAGACGAAAAT
GGATTCGTATTCACAGACATGATAATTATATAAGAATAAGAAAACCTTTGATT
TCTTTTTGAAAAGGAGGAACTGGCTTTCTTTGGAGTAATAAGCCTATTCCAATT

ACAATACTCATTGAGAAAGAATCGTAATAAATGCAAAGAAGAGGCGTCTTTTA
ACCAATAACGAAGAGTTTGAACCAAGGATTTCCAGAAGGGAACTGGGGTAAA
ATCC

LOCUS *V. vulcanicum* 909 bp DNA linear

TTCCGTCACAGAAATCTTGTGTTTACGAGCCAAAGTTCTAGCACAAGAAAATC
GAAGTATATACTTTATTCGATACAAACTCTTTTTTTTGGAAGATCCGCTATGAT
AATGAGAAAGGTTTCTGCATATACGCCCAAATCGATCAATAATATTAGAATCT
GATAAATCAGCCCAAACCGGCTTACTAATGGGATGCCCTAATAGGTTACAAAA
TTTAGCTTTAGCCAATGACGCAATCAGAGGAATAATTGGAACAAGAGTATCGA
ACTTCTTAATAGCATTATTGATTAGAAATGAATTTTCTAGAATTTGACTCCGTA
CCACCGAAGGGTTCATTCGCACGCTTGCAAGATAGCCCAAAAATTCAAGGGAA
TGATTGGATAATTGGTTTATATAAATCCTTCTTGGATGAAACCAAAGCGAAAA
ATACCATTGCCAAAAAGTGACAAGGTAACATTTCCATTTATTCATTAAAAGAG
ACGTCCCTTTTGAAGCCAGAATGGATTTTCTTTGATACCTAACATAATGCATGC
AAGGTTCCTTGACCAACCATAGGTTCTCCTGAAAATCCTTAACTTTACCAAAGA
CGTTCACAAGACGTTCTACTTTTCCATAGAAATAGATTCGTTCAAGAAAAACTC
CAGAAGATGTTGATCTTAAATGAGAAGATTGATTACGTAGAAAGACGAAAAT
GGATTCGTATTCACCGACATGATAATTATATAAGAATAAGAATAACCTTTGAT
TCCTTTTTGAAAAGGAGGAACTGGCTTTCTTTGGAGTAATAAGCCTTTTCCAAT
TACAATACTCATTGAGAAAGAATCGTAATAAATGCAAAGAGAGGCGTCtTTTA
ACCAATAACGAAGAGTTTGAACCAAGATTTCCAGTGGGAAACTGGGGTAAAG
CCA

Outgroup *ITS*

LOCUS *Antirrhinum majus* 720 bp DNA linear

CCTGCGGAAGGATCATTGTCGAAACCTGCAAAGCAGACCCGCGAACACGTGTT
TAATCCCCGGGCGTNGAGCCGAGGGAGGATGCAAGTCCCTCCTCGTCGAGACG
CCCTGCCCGCTCGACGCGTGCGTCGTTCGGGCTAACGAACCCCGGCGCGGCAT

GCGCCAAGGAAAACAACTGANAANCGTCGTGCCCCCGCTGCCCCGTTCGCGGG
TGCCTGCGGGGGGAACTCGACCTATCTTGAATGTCAAA ACGACTCTCG
GCAACGGATA
TCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAA
TTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCC
ATTAGGCCNAGGGCACNTCTGCCTGGGCGTCACGCATNNNNNNNCCCCCTCCT
TTGAGGATGGGGCGGAAAATGGCCTCCCGTCGCAAATGCGTGCGGCCGGCCCA
AATNTGATCCCGCGTCGACGGATGTCGCGACAAGTGGTGGTTGAACTCTCAAT
CGCGTGTTGTCGTGACTGGCTCCGTCGCATGCTCGGGCATCTTTTTGACCCAAC
GGCGCCTAATCGCGCTTTCGACCGCGACCCCAGGTCAGGCGGGACTACCCGCT
GAGTTTAAGCATATCAATAAGCGGAGGAAAAGAAACTTACAAGGATTCCCCTA
GTAACGGCGAGCGAACCGGGAACA

LOCUS *Coffea arabica* 2011 bp DNA linear

TAAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGG
AGTATGGTCGCAAGGCTGAAACTTAAAGGGATTGACGGAAGGGCACCACCAG
GAGTGGAGCCTGCGGCTTAATTTGACCCAACACGGGGAAACTTACCAGGTCCA
GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTG
GTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACG
AACGAGACCTCAGCCTGCTAACTAGCTATGCGGAGGAATCCCTCCGCAGCTAG
CTTCTTAGAGGGACTACGGCCTTTTAGGCCGCGGAAGTTTGAGGCAATAACAG
GTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGTATTC
AACGAGTCTATAGCCTTGGCCGACAGGCCCGGGTAATCTTTGAAATTTCATCG
TGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAGGAATTCCTAGTAA
GCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACACACCGCCC
GTCGCTCCTACCGATTGAATGGTCCGGTGAAGTGTTCGGATCGCGGCGACGTG
AGCGGTTCGCCGCCCGCGACGTCGCGAGAAGTCCACTGAACCTTATCATTTAG
AGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA
TTGTCGAATCCTGCATAGCAGATGACCGCGAACTCGTGTAATAGTCGGGCGTC
GGGGCGGGGGCGGTGAGGCCGAAACCTCTCCTCCCTCCCCGTCGCTCCCCGCG
CGCTCGTCGTGCGGACCAACAACCCAACCCCGGCGCGGAAAGCGCCAAGGAA

AACTCAAAAGATCGCTCGGCCCTCGACCGCCCCGTCCGCGGAGCGCGGGAGG
GGATGCCGCGGCGTCTGTCGTAACCAAAACGACTCTCGGCAACGGATATCTCG
GCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCA
GAATCCCGCGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCTTTAG
GCCGAGGGCACGTCTGCCTGGGCGTCACGCATCGCGTCGCCACCCCCCTCCCG
CGGGGGCGGCGGAGACTGGCCTCCCGTGCCCCCGGGCGCGGCCGGCCTAAAC
GCGAGTCCTCGGCGGGGGACGTCACGACCAGTGGTGGTTGAGTCCCTCAACTC
GAGTCCTTGTCGTGCCGTTAGACCACCCGCCGCATTCGGGGCTCCGACGACCC
TGAAGAGAGTTGCTCTCATCTCGACGGCGACCCCAGGTCAGGCGGGATTACCC
GCTGAGTTTAAGCATATCAATAAGCGGAGGAAAAGAAACTAACAAGGATTCC
CCTAGTAACGGCGAGCGAACCGGGAACAGCCCAAGCTTAGAATCGGGCGGCT
CCGCCGTCCGAATTGTAGCCTGGAGAAGCGTCCTCAGCGGCGGACCGGGCCCA
AGTCCCCTGGAATGGGGCACCGGAGAGGGTGACAGTCCCGTCGTGCCCGGACC
CTGTCGCACCACGAGGCGCTGTCGGCGAGTCGGGTTGTTTGGGAATGCAGCCC
CAATCGGGCGGTAAATTCCGTCCAAGGCTAAATACCGGCGAGAGACCGATAG
CAAACAAGTACCGCGAGGGAAAGATGAAAAGGACTTTGAAAAAAAAAAAAA
AAAAAAAAAAAAAAAACCTGGCCGCCCCCCCCGCCCCCCAAAAACTTTTCAA
ACCCTTTCTTTGGGCCCGGGGCCCCATAAGGAAATGAACCTGGGCCTAACTTTT
TCCCAAGAAAACCTGGGCCTGAATAAGGGTTGGATTTCCCCCCATAAAAAACC
CCCGGGGGCCACCCAACGTTCCTAACAAATCCCAAAGGGATTTTTAAGGCCTT
TCTGGATTTTTCCACCGGAATCCCAACCAACTCCATTTCCAATTTCCCCCCCCC
CTG

LOCUS *Coffea canephora* 784 bp DNA linear

GACGTGAGCGGTTCGCCGCCCGCGACGTCGCGAGAAGTCCACTGAACCTTATC
ATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAA
GGATCATTGTCGAATCCTGCATAGCAGATGACCGCGAACTCGTGTAATAGTCG
GGCGTCGGGGCGGGGGCGGTGAGGCCGAAACCTCTCCTCCCTCCCCGTCGCTC
CCCGCGCGCTTGTCGTGCGGACCAACAACCCAACCCCGGCGCGGAAAGCGCCA
AGGAAAACTCAAAAGATCGCTCGGCCCCCGACCGCCCCGTCCGCGGAGCGCG
GGAGGGGATGCCGCGGCGTCTGTCGTAACCAAAACGACTCTCGGCAACGGAT

ATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGA
ATTGCAGAATCCCGCGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGC
CTTTAGGCCGAGGGCACGTCTGCCTGGGCGTCACGCATCGCGTCGCCACCCCC
CTCCCGCGGGGGCGGCGGAGACTGGCCTCCCGTGCCCCCGGGCGCGGCCGGCC
TAAACGCGAGTCCTCGGCGGGGGACGTCACGACTAGTGGTGGTTGAGTCCCTC
AACTCGAGTCCTTGTCGTGCCGTTAGACCACCCGCCGCATTCGGGGCTCCGAC
GACCCTGAAGAGAGTTGCTCTCATCTCGACGGCGACCCCAGGTCAGGCGGGAT
TACCCGCTGAGTTTAAGCATATCAATAAGCGGAGGAAAAGAAACT

LOCUS *Veronicastrumsibiricum* 625 bp DNA linear

TTTCCGTAGGTGAACCTGYGGAMNAATCATTGTCGAAACCGCAAAGCAGACC
CGTGAACACGTGTTTAACATNGCACGGTGCCAGAYGCTTGCGTTCTGTGGGCT
AACGAACCCCGGCGCGGAATGCGCCAAGGAAAACTTAAAGGAAGCGTTCCCC
CCACGACGNCCGCTCGCGGTGCGTCGTGGGGAATGGACGAATCTAGAATGTCA
AAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGC
GAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTG
AACGCAAGTNGCGCCCGAAGCCATTCGGCTGAGGGCACGCCTGCCTGGGCGTC
ACGCATCGCGTCGCCCCCTCTCGAATCCCTCGGGATTWASGAGTGGGGTGGAA
AATGGTCTCCCGTGTGCCTTTGTGTGTGCGGTTGGCCCAAAAATGATCCTGCAT
CGACGGATGTCATGACCAGTGGTGGGTGAAAACTCAATCGTTATGTCGTGCAT
CAACCTGTCGTTTGGTATGGATCAAATCGACCCAATGGTGCCATATGTGCCTCC
GACCGCGACCCCAGGTCAGGCGGGACTACCCGCTGAGTTTAA

Outgroup *matK*

LOCUS *Antirrhinum majus* 819 bp DNA linear

TGTTAGATATATTAATACCCTACCCTGTCCATGTGGAAGTCTTGGTTCAAACTC
TTCGATATTGGGTAAAAGATGCCTCTTCTTTGCATTTAGTACGATTCTTTCTCA
ACGAGTATTGTAATTGGAATAGTCTTATTCTTCCAAAGAAAGCCAGTTTCTCTT
GTTCAAAAATAAATCAAAGATTGTTTTTATTCTTATATAATTCTCATGTATGTG

AATACGAATCCATTTTCGTCTTCCTACGTAACCAATCTTCTCATTTACTATCAA
CATCTTCTGGAGTTCTTCTTGAACGAATCTATTTCTATGGAAAAATAGAATGTC
TTGTGAACGTCTTTGTTAAGGTTAAGGATTTTCAGGCGAATCTATGGTTGGTCA
AAGAACCTTGCATGCATTATGTTAGGTATCAAAGAAAATCAATTCTGGCTTCA
AAAGGGACGTTTCTTTTAATAAAAAAATGGAAATGTTACCTTATGACTTTTTGG
CAATGGCATTTTTTGCTTTGGTTTCAGCCAAGAAGGATTTATATAAACCAATTA
TCCAACCATTTCCTTGAATTTTTAGGCTATCTTTCAAACGTACGAATGAATCCT
GCAGTGGTACGGAGTCAAATTTTAGAAAATTCATTTCTAATGAATAATTCTATT
AAGAAGTTTGATACCCTTGTTCCAATTAGCCCTCTGATTGCGTCATTGGCTAAA
GCGAAATTTTGTAACGTATTAGGGCATCCCATTAGTAAGCCGGTTCGGGCTGA
TTTATCAGATTCTAATATTATTGACCGATTTGGGCATATATGCAGAAATCTTTC
TCATTATCATA

LOCUS *Coffea arabica* 1518 bp DNA linear

ATGGAGGAAATTCAAGGATATTTACAGCTTGATAGATCTCAACAACACGGCTT
TCTATATCCACTTATCTTTCAGGAGTATATTTATGCACTTGCTCATGATCATAGT
TTAAACCGATCGATTTTGTTAGAAAATCCAGTTTATGATAATAAATCCAGTTTC
CTAATTGTGAAACGTTTAATTACTCGAATTTATCAACAAAATCATTTTATTATT
TTTGCTAATAATTCTAATCAAAATCTATTTTTTGGTCACAACAAGAATTTATAT
TCTCAAACAATATCAGAAGGACTTGCATTTATTGTGGAAATTCCATTTTATATA
CGATTAATATCTTCTCAAGGGGGGAAGGGATATTAAAATTTCGTAATTTACGA
TCAATTCATTCACCATTTCCTTTCTTAGAGAACAATTTTTCACATTTAAATTCTG
TGTTAGATATACTAATACCCCATCCCGTACATCTGGAAATCTTGGTTCAAAACC
TTCGTTATTGGGTAAAAGACGCCTCCTCCTTGCATTTATTACGATTCTTTTTCCA
CGAATATTGGAATTGGAATACTCTTATTGATAGAAAGAAACCCAGTTTTGATTT
TTCACCAAAAAGAAATCAAAAATTATTCTTCTTGTTATATAATTCTCATGTATG
TGAATACGAATCCATTTTCTTCTTTCTCCGTAAGCAATCTTCTCATTTGCGATCA
ACATCTTTTGGAGTCTTTCTTGAACGGATATATTTCTATGAAAAAAAAGAACGT
CTTGTAGAAGTCTTTGCTAAGGATTTTCAAGCCAGTCTATGGTTGTTCAAAGAT
CCTTTCATACATTATGTTAGGTATCAAGGAAAATCCATTCTGGTTTCAAATGGG
ACGCCTCTTTTGATGAATAAATGGAAATCTTACCTTGTTAATTTTTGGCAATGT

CATTTTGGTATTTGGTTTCACCCGGGAAGGGTCTATATAAACCAATTACCCAAT
TATTCCTTTAACTTTATGGGCTATCTTTCAAGTGTGCGACTAAACCATTCAATG
GTACGAAGTCAAATGCTAGAAAATGCATTTCTAATCAATAATGCTATTAAGAA
ATTTGATACCCTTGTTCCACTTATTCCCCTTATTGGATCATTGGCTAAAGCGAA
ATTTTGTAACTTATTAGGACATCCTATTAGTAAGCCGATTTGGACTGATTTGGC
AGATTCTGATATTATTGACCGATTTTGGCATATATGCAGAAACCTTTCTCATTA
TTATAGCGGATCTTCCAAAAAAAAGAGTTTGTATCGAATAAAATATATACTTC
GACTTTCTTGTGCTAAAACTTTGGCTCGGAAACACAAAAGTACTGTGCGTGCTT
TTTTGAAAGGATTAGGCTCGGAATTTTTGGAAAAATTCCTCACGTCGGAAGAA
GAAGCTCTTTCTTTGACCTTCTCAAGAACTTCTTCTACTTTTCGGGGGGTATAT
AGAAATCGAATTTGGTATTTGGATATTATTTATATCAATGATCTGACCAATTAT
CAATGA

LOCUS *Coffea canephora* 2424 bp DNA linear

GTGCGGCTAGCATCTTTTACACATTTAGATGAAGCAGGGGATTCATCTATACC
ATCGGTAAAGTTTGAAAGACCACGACTGATCCTGAAAGGGAATGAATGGAAA
AAACAGCATGTCGGATTGAGGGAGAGTTCTGAGAATGTTTCATTTTTACCGGA
TCAGTAGAAAAGAAAAACCCGGTTCGAATTCTTGGAATGGAAGAAAATAAAG
TTGGGTTGAATGAATAAATGGATAGGGCCTCGCGGCTTCAATTAATTATAGGG
AAAGATAAAGAAATGAGCTTCTGTTCTGAATTTGGAATAATTTCCTGATCTAAT
TAGACGTTAAAAATATATTAGTGCCCGGTGCAGGAAAGCTTTCACTCCCATGA
GTGAATTCTCAATTTTTTCAACGAATCCTAACTATTCTAACTCTCCATTATGAA
ACGGAGATGTGTGTGTAGAAGAAGCAGTATATTGATAAAGAAAATTTTTCCGA
AATCAAAAGAGCGATTGGGTTGAAAAAATAAAAGATTTCTAACCATCTTGTTA
TCCTATAACATAGACTAAAAATGTAACAAAGAGAGGATAGAGAATCTGTTTAT
AAGTTTATACTCGTCCCCGAGGTATCTTTTATTACTAAAATACTTTGTTTTGACT
GTATCGCACTATGTATCATTTGATAACCCCCAAATCTTCTACCTTTGGTTCAAA
TTTAAATCAAATGGAGGAAATTCAAGGATATTTACAGCTTGATAGATCTCAAC
AACACGGCTTTCTATATCCACTTATCTTTCAGGAGTATATTTATGCACTTGCTC
ATGATCATAGTTTAAACCGATCGATTTTGTTAGAAAATCTAGTTTATGATAATA
AATCCAGTTTCCTAATTGTGAAACGTTTAATTACTCGAATTTATCAACAAAATC

ATTTTATTATTTTTGCTAATAATTCTAATCAAAATCTATTTTTTGGTCACAACAA
GAATTTATATTCTCAAACAATATCAGAAGGACTTGCAT
TTATTGTGGAAATTCCATTTTATATACGATTAATATCTTCTCAAGGGGGGAAAG
GGATATTAAAATATCGTAATTTACGATCAATTCATTCACCATTTCCTTTCTTAG
AGAACAATTTTTCACATTTTAATTCTGTGTTAGATATACTAATACCCCATCCCG
TACATCTGGAAATCTTGGTTCAAAACCTTCGTTATTGGGTAAAAGACGCCTCCT
CCTTGCATTTATTCGATTCTTTTTCCACGAATATTGGAATTGGAATACTCTTATT
GATAGAAAGAAACCCAGTTTTGATTTTTCACCAAAAAGAAATCAAAAATTATT
CTTCTTGTTATATAATTCTCATGTATGTGAATACGAATCCATTTTCTTCTTTCTC
CGTAAGCAATCTTCTCATTTGCGATCAACATCTTTTGGAGTCTTTCTTGAACGG
ATATATTTCTATGAAAAAAAAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTT
CAAGCCAGTCTATGGTTGTTCAAAGATCCTTTCATACATTATGTTAGGTATCAA
GGAAAATCCATTCTGGTTTCAAATGGGACGCCTCTTTTGATGAATAAATGGAA
ATCTTACCTTGTTAATTTTTGGCAATGTCATTTTGGTATTTGGTTTCACCCGGGA
AGGGTCTATATAAACCAATTACCCAATTATTCCTTTAACTTTATGGGCTATCTT
TCAAGTGTGCGACTAAACCATTCAATGGTACGAAGTCAAATGCTAAAAATGCA
TTTCTAATCAATAATGCTATTAAGAAATTTGATACCCTTGTTCCACTTATTCCCC
TTATTGGATCATTGGCTAAAGCGAAATTTGTAACTTATTAGGACATCCCATTAG
TAAGCCGATTTGGACTGATTTGGCAGATTCTGATATTATTGACCGATTTTGGCA
TATATGCAGAAACCTTTCTCATTATTATAGCGGATCTTCCAAAAAAAAGAGTTT
GTATCGAATAAAGTATATACTTCGACTTTCTTGTGCTAAAACTTTGGCTCGGAA
ACACAAAAGTACTGTGCGTGCTTTTTTGAAAGGATTAGGCTCGGAATTTTTGG
AAAAATTCCTCACGTCGGAAGAAGAAGCTCTTTCTTTGACCTTCTCAAGAACTT
CTTCTACTTTTCAGGGGGTATATAGAAATCGAATTTGGTATTT
GGATATTATTTATATCAATGATCTGACCAATTATCAATGATTTATCCTGAGGTC
TTGTAAATGCCAATTTTCCTAAGAGATAATAAAAATTTTTTACTTCTACTATTC
TGAAATGTTAATGTAATAAGGGTTAAATCAACCGATTATTCTTAAAAGTTTTTC
TAAAAAAAAAAGAACTGATGTATACATAGAGAAAGCCGTGTGCAATGAAAAA
TGCAAGCACGGCTTGGGGAGGGATGTTTAC

LOCUS *Veronicastrum sibiricum* 1533 bp DNA linear

ATGGAGGAAATCTTACAGCTAGAGAGATCTCAACAACACGACTTTCTATACCA
CTTATCTTTCAGGAGTATATTTATGCATTTGCTCATGATCGTACTTTCAGTAGAT
CTATTTTGTCGGAAAATATGGATTCGGAAAATCTGGGTTATCAGAATAAATCC
AGTTTACTGATTGTGAAACGATTAATTGCTCAAATGTATCAACAGAATCATTTT
CTTATTTCTCCTAATGATTCTAACCAAAATCCATTTTGGGGTCGAAAGAAGAAT
TTGTATTCTCAAATCATATCAGAGGGGTTTGCATTTATTGTGGAAATTCCATTT
TCTGTACAATTAATATCTTCTCTAGAAGGAAAAAAGAAAAAGATAGTAAAATA
TCAGAATTTACGCTCAATCCATTCAATATTTCCCTTTTTAGAGGACAATTTTTC
ACATTTAAATTTTGTATTAGATATACTAATACCCCACCCTGTCCATGTGGAAAT
CTTGGTTCAAACTCTTCGCTATTGCGTAAAAGATGCCTCTTCTTTGCATTTAGT
ACGATTTTTTCTCAACGAGTATTGTAATTCGAATAGTCTTATTACTACAAAGAA
AGCCGGTTTCTCTTTTTCAAAAAGAAATCAAAGATTCTTTTTATTCTTATATAA
TTCTCATGTATGTGAGTACGAATCCGTTTTCGTCTTTTTACGTAACCAATCTTCT
CATTTACGATCAACATCTTCTGGAGCTCTTCTTGAACGAATCTATTTCTATGGA
AAAATAGAGCATCTTGTGAACGTCTTTGGTAAGGTTAAGGACTGTCAGGTGAA
CCTATGGTTGGTCAAAGAACCTTGCATGCATTATGTTAGATATCAAAGAAAAT
CCATTCTGGCTTTAAAGGGGACGTCTCTTTTTATGAATAAATGGAAATGTTACC
TTGTCACTTTTTGGCAATGGCATTTTGCGCTCTGGTTTCATCCAAGAAGGATTT
CTATAAACCAATTATCCAACCATTTGCTTGAATTTGTGGGCTATCTTTCAAGCG
TGCGAATGAATCCTGCAGTGGTACGGAGTCAAATTCTAGAAAAGTCATTTCTA
ATCAATAATGCTATTAAGAAGTTCGAGACCCTTGTTCCTATTATTCCTCTGATT
GCGTCGTTGGCTAAAGCGAAATTTTGTAACGTTTTAGGGCATCCCATGAGTAA
GCCGGTTTGGGCCGATTTATCAGATTCAAATATTATTGACCGATTTGTACATAT
ATGCAGAAATCTTTCTCATTATTATAGCGGATCTTCCAAAAAAAAGAGTTTATA
TCGAATAAAGTATATACTTCGACTTTCTTGTGCTAGAACTTTGGCTCGGAAACA
CAAAAGTACTGTACGTGCTTTTTTGAAAACATTAGGATCCGGATTATTGGAAG
AATTTCTTATATCGGAAGAAGACGTTCTTTGTTGGACCCTCCCCAAAACTTCTT
CCGCTTTGCGGGGATTATATAAAAGTCGGATTTGGTATTTGGATATTATTTGTA
TCAATGATCTGGCCACTCACAAATAA

APPENDIX D: SANGER SEQUENCING CHROMATOGRAM RESULTS OF *MATK* AND *ITS* PRIMERS ON *VERBASCUM LURİDİFLORUM*

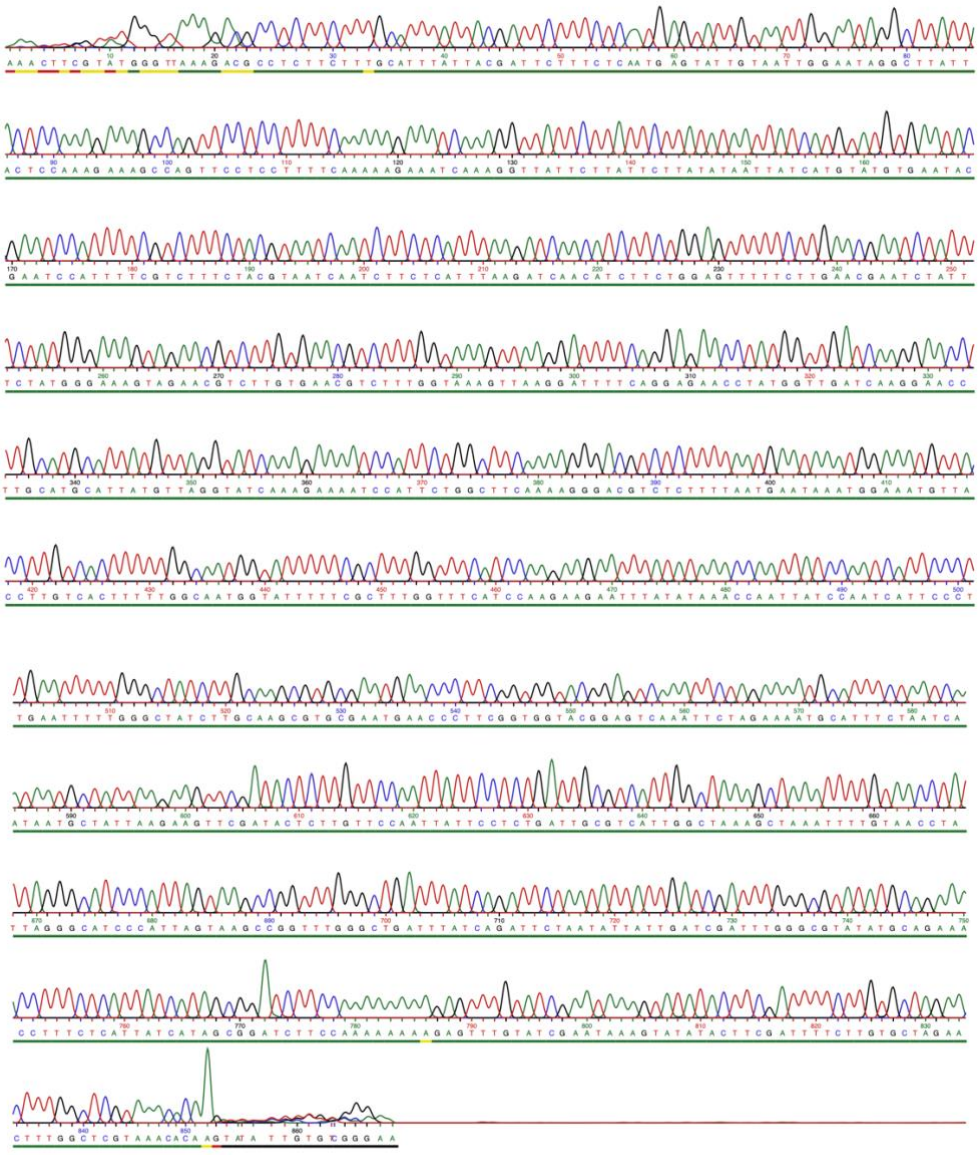


Figure D.1. Chromatogram Results of Forward *matK* Primer on *Verbascum luridiflorum*

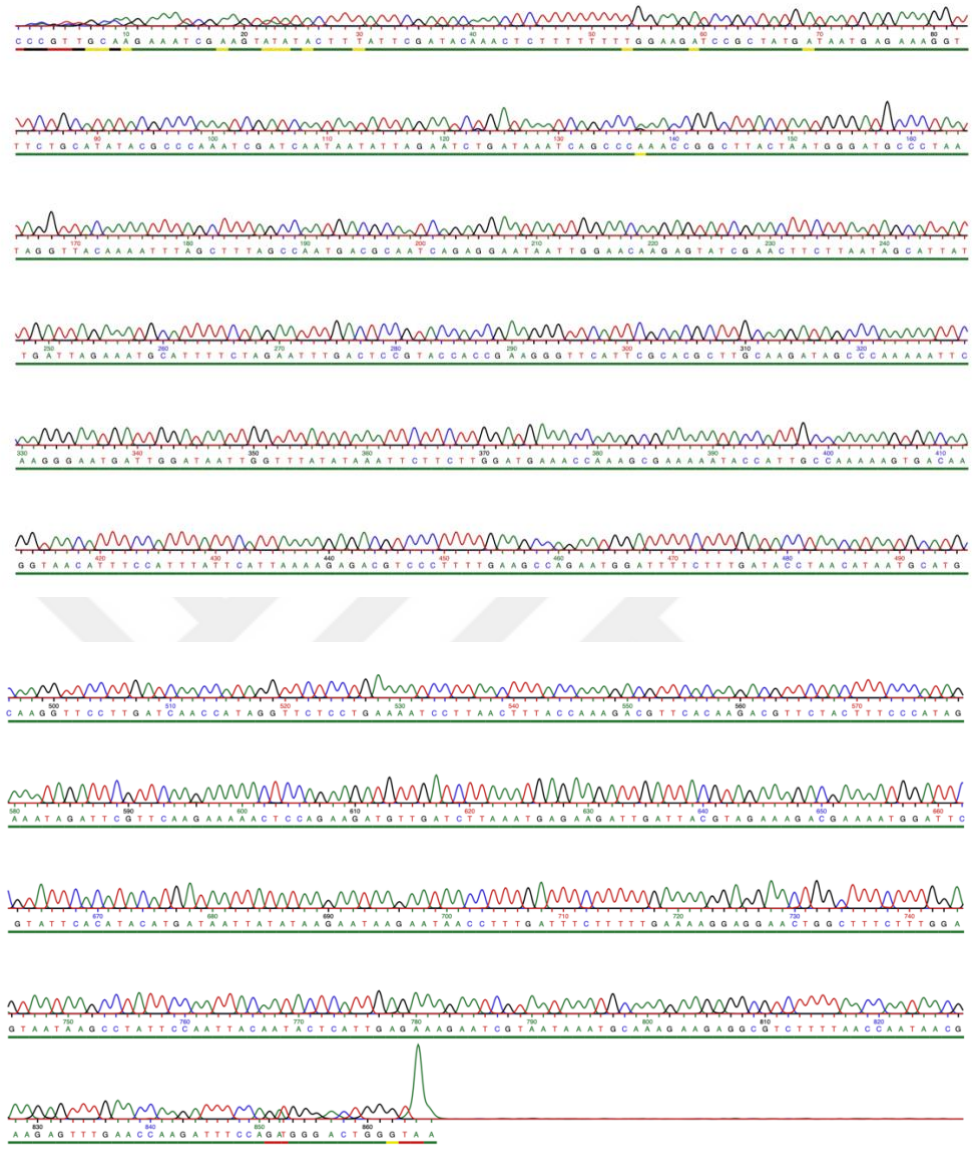


Figure D.2. Chromatogram Results of Reverse *matK* Primer on *Verbascum luridiflorum*

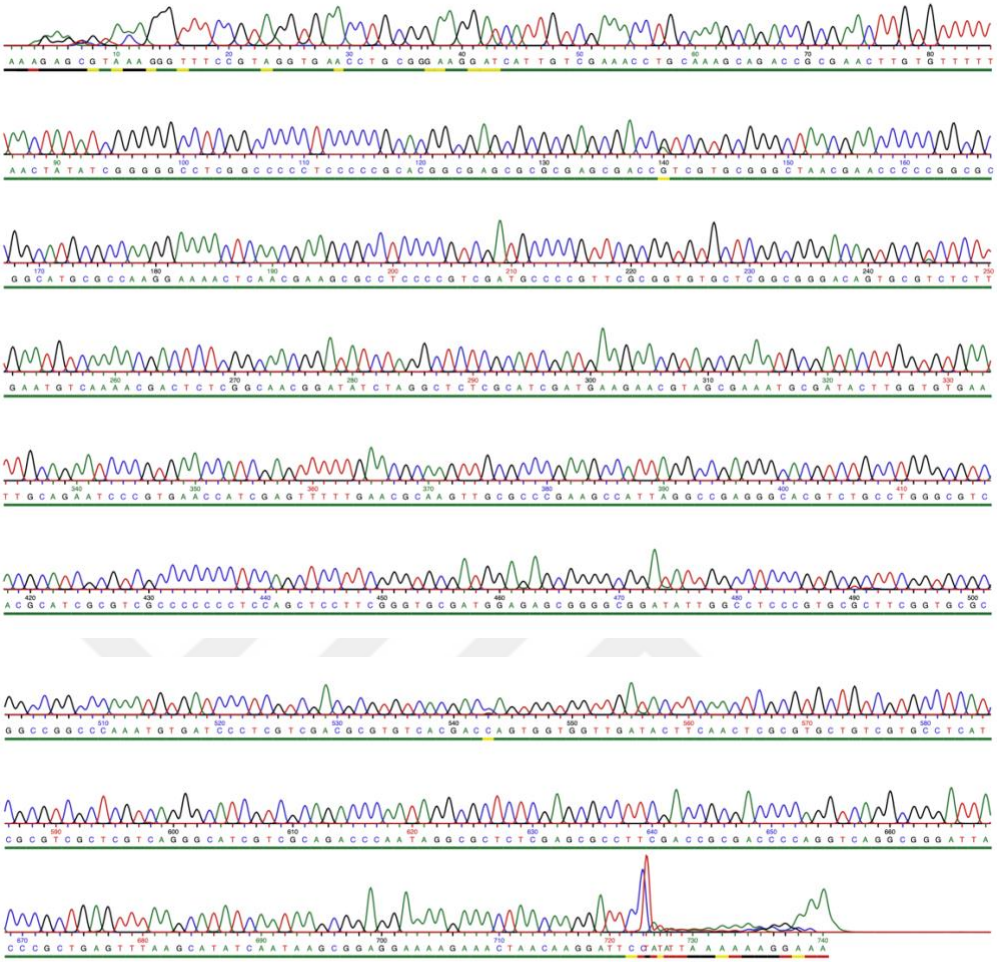


Figure D.3. Chromatogram Results of Forward *ITS* Primer on *Verbascum luridiflorum*

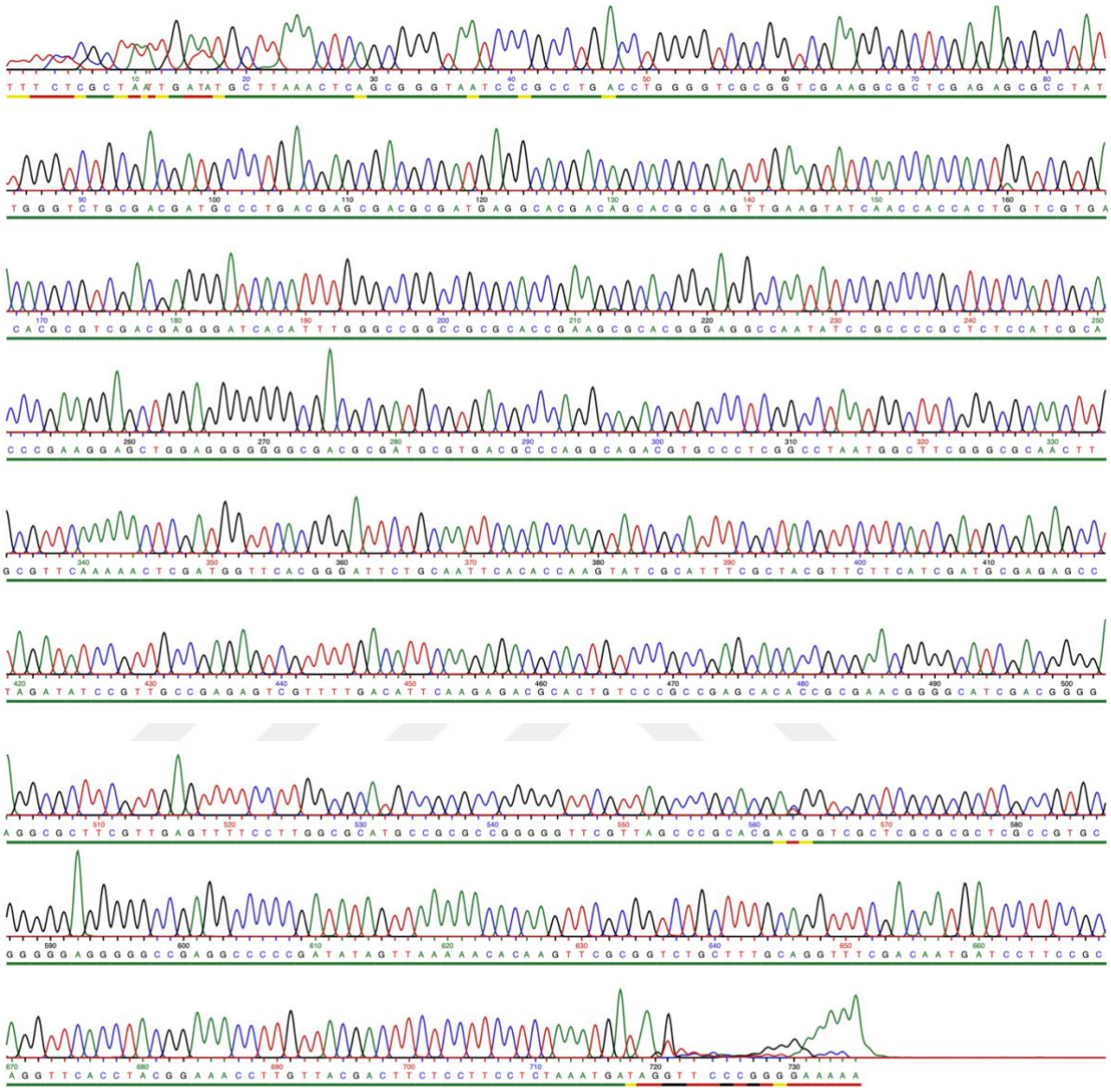


Figure D.4. Chromatogram Results of Reverse *ITS* Primer on *Verbascum luridiflorum*

APPENDIX E: GEL ELECTROPHORESIS RESULTS OF CTAB PROTOCOL ISOLATED *VERBASCUM* L. SAMPLES