UNIVERSITY OF NAIROBI

DNA barcoding, mineral and phytochemical analysis of Cape gooseberry (*Physalis* spp.) accessions in Kenya

Katherine Pere (BSc, MSc)

A thesis submitted in fulfillment of the requirements for the award of Doctor of Philosophy Degree (Biochemistry) of the University of Nairobi

July, 2024

DECLARATION

I, the undersigned declare that this is my original work and that it has not been presented to any institution of learning for academic credit. All the sources used herein are duly acknowledged.

Signature:  **Date:** 16th August 2024

Name: Katherine Pere

Registration No.: H56/54538/2019

Supervisors:

This thesis has been submitted with our approval as the University Supervisors:

Prof. Evans N. Nyaboga

Department of Biochemistry, University of Nairobi

Signature  **.....** **Date**16/08/2024**.....**

Prof. Edward K. Muge

Department of Biochemistry, University of Nairobi

Signature  **.....** **Date**19/08/2024**.....**

Prof. Maina J. Wagacha

Department of Biology, University of Nairobi

Signature  **.....** **Date**18/08/2024**.....**

DEDICATION

I dedicate this work firstly to God, for the faithfulness and grace he has provided me with throughout my studies until this level. Secondly, I dedicate this work to my parents, husband and siblings that have been supportive in so many ways and have always encouraged me to focus on the final goal despite all the challenges. I thank everyone, friends and family for their never-ending prayers and encouragement.

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my Supervisors Prof. Evans Nyaboga, Prof. Edward Muge and Prof. Maina Wagacha for their advice and academic guidance throughout my PhD studies including but not limited to proposal drafting, laboratory research work, manuscript drafting and submission, as well as compiling this thesis. I would also like to thank my main Supervisor for purchasing some consumables that I used in the laboratory work as well as funding the sequencing of samples studied in my research work.

I would like to appreciate the former director Centre for Bioinformatics and Biotechnology (CEBIB) Prof. George Obiero, UoN for allowing me to perform part of my research work in CEBIB laboratories. I would like to also appreciate Pheris Namakwa and Nicholas Kipkiror for their assistance. Finally, I would like to appreciate Mr. Thomas Arani, Mr. George Okumu, Mrs Phyllis Mwai and Mr. Francis Mbaya from the Department of Biochemistry, Mr. Mwangi from the Department of Chemistry and Mrs. Anne Owiti from Centre for Biotechnology and Bioinformatics (CEBIB), University of Nairobi for their assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

12S:	12 Svedberg
16S:	16 Svedberg
AAS:	Atomic absorption spectrophotometer
ABGD:	Automatic barcode gap discovery
AFLP:	Amplified fragment length polymorphism
ANOVA:	Analysis of Variance
AtpF:	Adenosine triphosphate (ATP) synthase F
AtpH:	Adenosine triphosphate (ATP) synthase H
BMC:	BioMed Central
BLAST:	Basic local alignment search tool
BLASTn:	Basic local alignment search tool nucleotide
CAPS:	Cleaved amplified polymorphic sequences
CBOL:	Consortium for the Barcode of Life
CO1:	Cytochrome oxidase 1
CoEHM:	Center of Excellence in HIV Medicine
COS:	Conserved ortholog set
cpDNA:	Chloroplast DNA
CTAB:	Cetyltrimethylammonium bromide
CV:	Coefficient of Variance
Cytb:	Cytochrome b
DAF-DNA:	Amplified fingerprinting deoxyribonucleic acid
DnaSP:	Deoxyribonucleic acid sequence polymorphism
DPPH:	2, 2-diphenyl-2-picrylhydrazyl
DW:	Dry weight
Dxy:	Average nucleotide substitution per site between populations
GAE:	Gallic acid equivalent
GTR:	Generalized Time Reversal
HDS:	Honestly significant difference
HIV:	Human immunodeficiency virus
HRS:	Hydrogen peroxide radical scavenging

InDels:	Insertions and deletions
IRG:	Immunity regulated genes
ISSR:	Inter-simple sequence repeats
ITS:	Internal transcribed spacer
ITS1:	Internal transcribed spacer 1
ITS2:	Internal transcribed spacer 2
K2P:	Kimura 2 parameter
K80:	Kimura 80
mat-K:	Maturase K
MCMC:	Markov chain Monte Carlo
MEGA:	Molecular evolutionary genetic analysis
MUSCLE:	MUltiple Sequence Comparison by Log-Expectation
MSA:	Multiple sequence alignment
NA:	Nutrient agar
NCBI:	National Center for Biotechnology Information
OD:	Optical density
PCR:	Polymerase chain reaction
Pi:	Nucleotide diversity
RAPD:	Random amplified polymorphic DNA
rbcL:	Ribulose-1,5-bisphosphate carboxylase large
RFLP:	Restriction fragment length polymorphism
RAM:	Random amplified microsatellites
RNase:	Ribonuclease
rRNA:	Ribosomal Ribonucleic acid
SCARs:	Sequence characterized amplified regions
SD:	Standard deviation
SNPs:	Single nucleotide polymorphisms
SSR:	Simple sequence repeats
TFC:	Total flavonoid content
TPC:	Total phenolic acid content
TTC:	Total tannic acid content

USA:	United States of America
UV:	Ultra-violet
ycf 1:	Yeast cadmium factor protein 1

ABSTRACT

Physalis genus belongs to the *Solanaceae* family and is mostly a wild self-propagating orphan fruit plant. *Physalis* fruit contains nutritional and phytochemical compounds of importance to public health and is a potential ingredient for fortification of foods and beverages. In Kenya, commercial cultivation of *Physalis* has significantly increased for the last decade and the farmers are concerned about the reliability and identity of the planting material. Challenges facing the production of planting material include difficulty in precise identification of the *Physalis* species due to lack of discriminatory morphological features. DNA barcoding is an effective tool in identification and discrimination of plant species. Therefore, the objectives of this study were to identify and discriminate the different *Physalis* species in the wild and cultivated in Kenya using DNA barcoding as well as determine the nutritional and biochemical profile of selected *Physalis* accessions. To identify the *Physalis* species from selected Counties in Kenya, ribulose-1,5-bisphosphate carboxylase large (*rbcL*) gene and internal transcribed spacer 2 (ITS2) DNA barcodes were used. Genomic DNA was extracted from leaves and polymerase chain reaction (PCR) amplification was carried out using ITS2 and *rbcL* barcodes followed by Sanger sequencing of the amplicons. Sequences were curated and aligned along their reference sequences based on MUltiple Sequence Comparison by Log-Expectation (MUSCLE). The generated multiple sequence alignments were used to prepare phylogenetic trees for *Physalis* species discrimination based on *rbcL* and ITS2 genes by Bayesian inference based on MrBayes software. Genetic diversity, distance and polymorphism were also assessed for the *Physalis* species through the use of DNA sequence polymorphism (DnaSP) software. Additionally, assessment of *Physalis* species mineral content for nine representative accessions was performed on an atomic absorption spectrophotometer and phytochemical content determined by UV-Visible spectrophotometry. Further, the radical scavenging activity of *Physalis* accessions was determined for 2,2-diphenyl-2-picrylhydrazyl (DPPH) and hydrogen peroxide radicals. Success rate of PCR amplification and sequencing of *rbcL* and ITS2 genes ranged from 82% to 88% and 65% to 67%, respectively. The phylogenetic tree generated for the *rbcL* sequences was unable to distinguish between *Physalis* species. However, the phylogenetic tree generated for the ITS2 sequences successfully identified three species of *Physalis*: *Physalis peruviana*, *Physalis purpurea* and *Physalis cordata*. There was low nucleotide diversity and genetic distance of 0.01333 and 0.04, respectively, for the *rbcL* genes while high genetic diversity and distance was noted for the ITS2 sequences. The highest genetic diversity and distance between *Physalis* species was noted between *P. peruviana* and *P. cordata* at 0.36923 and 0.703, respectively based on ITS2 sequences. The highest genetic nucleotide diversity and distance within species was noted for the *P. peruviana* at 0.26324 and 0.46, respectively, for the ITS2 gene sequences. Tajima D values obtained indicated low frequency mutations among the *Physalis* sequences based on *rbcL* and ITS2 barcodes. The mineral content of the *Physalis* accessions was significantly different ($p < 0.05$) for calcium, copper, zinc, nickel and lithium between accessions. For the phytochemicals of *Physalis* fruit accessions, tannic acid content was significantly different ($p < 0.05$), while the phenolic acid and flavonoid contents were not significantly different ($p > 0.05$). *Physalis* accessions DPPH radical scavenging activity was significantly different ($p < 0.05$) while no significant difference ($p > 0.05$) was noted for hydrogen peroxide scavenging capacity. Therefore, this study identified Kenyan *Physalis* accessions based on ITS2 barcode region as *P. purpurea*, *P. peruviana* and *P. cordata*. The findings also demonstrated that *Physalis* accessions in Kenya are rich in mineral and phytochemical contents as well as high antioxidant properties.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the study

Physalis genus is a wild plant that belongs to the nightshade (*Solanaceae*) family (Zhang and Tong, 2016). It is native to South American Andes either Peru or Chile, hence referred to as the Peruvian gooseberry. *Physalis* is now cultivated in various regions of the world and the fruits are exported from several countries including Colombia, Australia, New Zealand, Great Britain, Zimbabwe, Kenya, Egypt, South Africa, Madagascar and South East Asia (Zhang *et al*., 2013; Ramadan and Moersel, 2003). Colombia is the leading producer, consumer and exporter of *Physalis* fruits (Zhang *et al*., 2013). In Kenya, the fruits are vastly seen as wild and self-propagating plants, commonly in farms during and after harvesting of maize. *Physalis* has golden yellow edible fruits and inside it has a juicy pulp with many small yellowish seeds. The fruits are protected by papery husks against insects, birds, and adverse conditions. There is growing commercial interest in this orphan fruit crop in Kenya because of its nutritional and pharmacological properties and health benefits.

Nutritional and phytochemical profile of *Physalis* has been studied to a moderate level in the Andes (Ramadhan and Morsel, 2007). The *Physalis* fruit contains fat and water-soluble vitamins (B-complex, C, A, E and K), minerals (mainly phosphorus, magnesium, potassium, zinc and calcium), sugars (e.g., sucrose, glucose and fructose), fatty acids (palmitate, oleic and gamma linoleic acid), phytosterols (sitosterol and stigmasterol) as well as fiber (pectin) (Puente *et al*., 2011; Ramadhan and Morsel, 2003). The fruit pomace (skin and seed) contains 19.3% oil, 3.1% ash, 17.8% protein, 24.5% carbohydrates and 28.7% crude fiber (Ramadhan and Morsel, 2007). Phytochemicals found in *Physalis* include withanolides, physalins, carotenoids, phenolics and flavonoids (Puente *et al*., 2011). The most studied phytochemicals are the poly phenolic compounds and they have been shown to have anti-oxidant properties and medicinal properties (Somani *et al*., 2015). The key polyphenolics include flavonoids, stilebenes, phenolic acids, coumarins and tannins (Bayir *et al*., 2019). The high content of vitamins, minerals and antioxidant phytochemicals in *Physalis* explains its medicinal properties such as anti-cancer, antibacterial and anti-inflammatory activity (Hong *et al*., 2015). *P. alkengi* has been used in China to treat sore throats, tumors, urinary problems, hepatitis, eczema, leishmaniasis and cough (Chinese Pharmacopoeia Editorial Committee, 2015). However, there is lack of information on cultivar-specific mineral and phytochemical profiling of *Physalis* species growing in Kenya. The characterization of *Physalis* phytochemical content has been done in Uganda and they were able to identify several types of phytochemicals present in the plant such as phenolic

compounds and flavonoids (Kasali *et al*., 2021). Characterization of *Physalis* species and major cultivars in terms of their nutritional and chemical composition, functional and physicochemical properties could provide important knowledge for future utilization. The information could accelerate the processing of value-added products from gooseberry fruits.

Due to their ethnobotanical, medical and economic applications and a wide array of species there is a need to authenticate and identify various *Physalis* species for future species-specific use. *Physalis* species identification ensures their appropriate utilization, germplasm conservation and formation of future breeding programs (Feng *et al*., 2018). Identification of *Physalis* species using traditional identification systems such as morphological appearance has resulted in misidentification due to high similarities in phenotypic characteristics (Feng *et al*., 2016). *Physalis minima* and *Physalis pubescens* are morphologically similar and present a clear challenge in their distinction using their phenotypic characteristics only (Feng *et al*., 2016). Morphological identification is also affected by environmental and physiological factors which can affect phenotypic characteristics (Vargas-Ponce *et al*., 2011; Menzel, 1951). Misidentification of *Physalis* can cause the loss of genetic information due to lack of genetic conservation (Feng *et al*., 2018). Since morphological identification of *Physalis* species has proven to be inefficient there is a need to use other robust means of genus and species identification (Yu *et al*., 2021). In the *Solanaceae* family, identification of the *Physalis* genus is also important because plants in this genus are sometimes confused with other plants such *Nicandra physalodes* (Feng *et al*., 2016). Molecular identification of plant species has proven to be more efficient than the use of morphology. This has been done successfully using species specific SCAR markers and DNA barcoding (Feng *et al*., 2018; Feng *et al*., 2016; Simbaqueba *et al*., 2011).

DNA barcoding entails a rapid and reliable method of genus and species identification using short universal standardized DNA sequences (Saddhe and Kumar, 2018). DNA barcording has widely been utilized and accepted in the identification of plants and animals (Kress, 2005). Therefore, DNA barcoding has proved to be an effective tool in taxon identification and also in the identification of medicinal plants (Yu *et al*., 2021; Dormontt *et al*., 2018). There are several known DNA barcodes utilized in the identification of plants such as chloroplast genes and the nuclear ribosomal DNA (rDNA) genes. The chloroplast DNA barcode genes including ribulose bisphosphate carboxylase large (*rbcL*) and maturase (*matK*) have been reported to be ideal for plant barcoding studies (Kang *et al*., 2017). Intergenic sequence *psbA-trnH* gene and the nuclear rRNA barcodes including 5.8S, 26S, 18S rRNA and internal transcriber spacer (ITS) 1 and 2 are used as supplements in plant barcoding (Kang *et al*., 2017). Proper selection of plant DNA barcodes is crucial for the successful

identification of plant genus and species. Among the chloroplast DNA barcodes, the *rbcL* region, which was the first plant gene to be sequenced, has proved to be highly effective and successful in species identification for some plant genera (Ismail *et al*., 2020). It has a higher amplification rate during polymerase chain reaction (PCR), better universality, and produces high quality sequences with high success rate during sequencing (Newmaster *et al*., 2006). It is considered the best characterized DNA barcode gene (Duan *et al*., 2019). The nuclear DNA barcodes ITS2 genes are considered the best candidate to use in DNA barcoding due to their high species discrimination, ability to identify diversity at the interspecies and intra-species level, high rate of success in amplification and sequencing in plants (Kang *et al*., 2017). It is however important to note that the debate on the best DNA barcode in plant species and taxon identification is still ongoing as certain barcodes work better in the identification of specific plants (Ralte and Singh, 2021; Tran *et al*., 2021).

DNA barcoding has not been previously used in the identification of *Physalis* species in Kenya. Molecular markers such as simple sequence repeat (SSR) have been used in the genetic characterization of *Physalis* in Kenya (Muraguri *et al*., 2021; Simbaqueba *et al*., 2011), with the assumptions that the identity of the species was *P. peruviana*. Therefore, there is need to identify and discriminate the *Physalis* species in Kenya for species-specific uses, genetic conservation of vulnerable species and for use in breeding programs. The efficiency of *rbcL* and ITS2 DNA barcode genes in the identification of *Physalis* species in Kenya was evaluated in the current study.

1.2 Problem statement

Physalis is a wild fruit producing plant that grows without human intervention in Kenyan fields after harvesting of maize or even in forests and on the sides of the roads. It has wide application in ethnomedicine where it is used in the deterrence and treatment of many diseases due to a rich phytochemical profile (Hong *et al*., 2015). *Physalis* plants also have a rich nutritional profile and can be a source of both macro- and micro-nutrients (Musinguzi *et al*., 2007). Most *Physalis* species are morphologically similar and difficult to distinguish. *P. minima* and *P. pubescens* are highly similar in appearance, though they have different applications in ethnomedicine. Improper identification of these species can lead to improper utilization of the plants (Feng *et al*., 2016). There is a lack of a database of *Physalis* species and accessions growing in Kenya as well as studies on their genetic diversity. The lack of genetic diversity studies of the plant in Kenya is also a risk to genetic loss of this important fruit. This can lead to the extinction of important species and accessions due to climate change and urbanization as the plant mostly grows without human intervention in bushes and forests.

There is a need to identify and genetically characterize *Physalis* species to initiate proper utilization and conservation as well as its genetic improvement through breeding programs.

1.3 Justification

Physalis fruits have high nutritional value, and pharmacological properties due to their rich phytochemical profile and many health benefits. This plant is rich in vitamins, minerals and soluble sugars (Puente *et al*., 2011; Ramadhan and Morsel, 2007). The medicinal usages indicate that it might be rich in a plethora of phytochemicals (Puente *et al*., 2011). However, it is underutilized in Kenya due to a lack of information on the nutritional and biochemical profiles of the accessions growing in the Country. Assessment of the nutritional and phytochemical content of *Physalis* accessions growing in Kenya could support the utilization of the plants as sustainable resource for the development of biofortified foods and also promote their use as natural source of antioxidants. Molecular identification and characterization of *Physalis* species has majorly been done in Colombia and China (Chacón *et al*., 2016; Feng *et al*., 2016). Precise identification of *Physalis* species is crucial to ensure safe utilization especially for medicinal applications, like in any other medicinal plants; it is paramount to prevent improper use (Feng *et al*., 2016; Menzel *et al*., 1951). Molecular identification could also ensure the maintenance and conservation of *Physalis* genetic resources to prevent the loss/extinction of important species and accessions and also help in the development of breeding programs to create hybrids. For breeding purposes, information on accurate identification of the different species is fundamental. DNA barcoding is the most effective and emerging method of molecular identification of different species in plants.

1.4 Objectives

1.4.1 Main objective

The main objective of the study was to determine the mineral and phytochemical profiles of *Physalis* accessions and the genetic relationships of *Physalis* species using DNA barcoding.

1.4.2 Specific objectives

1. To determine the mineral composition, phytochemical content, antioxidant activity and identity of indigenous *Physalis* accessions from Sorget forest, Kericho County, Kenya.
2. To determine the identity and genetic diversity of *Physalis* accessions in Kenya using their DNA sequences and RNA secondary structural information of their ITS2 barcode.
3. To determine the identity and genetic diversity of *Physalis* accessions in Kenya using cpDNA gene sequences.

4. To investigate the species resolution ability of nuclear (ITS2) and chloroplast DNA (*rbcL*) loci in 64 accessions of *Physalis* collected in Kenya.

1.5 Null Hypotheses

1. *Physalis* accessions growing in Kenya are not rich in nutrients, minerals, phytochemicals and lack anti-oxidant activity.
2. *Physalis* species growing in Kenya cannot be discriminated based on sequences and RNA secondary structures of their ITS2 barcode.
3. *Physalis* species growing in Kenya cannot be discriminated based on cpDNA gene sequences.
4. *Physalis* species and their accessions are not genetically diverse and have low level of nucleotide polymorphism based on nuclear and cpDNA sequences.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 The Genus *Physalis*

Physalis is a genus that belongs to the Solanaceae (nightshade) plant family. *Physalis* was first described by Linnaeus in 1753. There are more than 90 known species of *Physalis* that grow in the temperate and tropical regions of the world (Zimmer *et al*., 2019). Most species of *Physalis* are found in the Andes and America with few originating in Europe and Asia (Seleem and Nassar, 2021; Whitson and Manos, 2005). In China five *Physalis* species and two varieties have been identified (Chinese Academy of Sciences, 1978). In Mexico and China, *P. peruviana*, *P. pubescens*, *Physalis alkekengi* and *Physalis philadelphica* are extensively cultivated for their edible fruits which are rich in nutrients and phytochemicals that confer them with medicinal properties (Zamoras-Tavares *et al*., 2015; Wei *et al*., 2012).

Physalis species are considered annual or perennial flowering herbs that can also be woody at their base and are creeping plants (Zimmer *et al*., 2019). The leaves of *Physalis* species have soft texture, are alternate, petioled and angled. Their flowers are mostly yellowish, whitish or blue with extra axillary pedicles. They have a bladder like calyx that is five-toothed and encloses yellowish-green berries that are pulpy in the inside with small, flat and rounded seeds (Figure 2.1.) (Bailey, 1949).

Figure 2.1. Plant morphology of *Physalis* accessions.

A: represents a photograph of mature *Physalis* fruits; and B: represents a *Physalis* plant

2.2 Distribution and economic importance of *Physalis*

The plant being native to the Andes has been naturalized in other parts of the world such as Asian and African countries. The *Physalis* plants are mainly grown, consumed and exported for income in Colombia. The economic value of *Physalis* in Colombia is linked to the high demand for fruits in the European market (Álvarez-Flórez *et al*., 2017; Ordoñez *et al*., 2017). Other exporters of *Physalis* plants include Madagascar, New Zealand, Uganda, Australia, Zimbabwe and Kenya (Zhang *et al*., 2013; Ramadan and Moersel, 2003). *Physalis* plants are considered important for income generation, as food and for medicinal purposes.

2.3 Nutritional profile of *Physalis* fruits

The nutritional profile of *Physalis* fruits has shown that they contain unique bioactive compounds. These compounds include withanolides, physalins, vitamins such as A, C, E, K and B complex, carbohydrate sugars and particularly rich in sucrose, glucose and fructose (Zhang *et al*., 2013; Ramadhan, 2011; Puente *et al*., 2011; Zhao, 2007). Sucrose levels are usually higher than glucose. Fructose is less than sucrose and glucose but is important for diabetic patients as a source of energy (Puente *et al*., 2011). It is also rich in fiber such as pectin which is an intestinal regulator and is also used in the preparation of jams and jellies (Gironés-Vilaplana *et al*., 2014). Other studies have indicated that *Physalis* is a good source of polyunsaturated fatty acids which are essential fats. It has been noted to have 0.2% fatty acids of different types such as palmitic acid, palmitoleic acid, oleic acid, linoleic acid and gamma-linolenic acid (Rodrigues *et al*., 2009; Ramadhan and Morsel, 2003). Foods that have high linoleic acid content prevent the development of hypertension and cardiovascular diseases (Ramadhan and Morse, 2003). The fruit also contains phytosterols such as ergosterol (9.23 g/kg) and stigmasterol (6.23 g/kg) (Ramadhan and Morsel, 2003). Another study showed that the most abundant phytosterol is campesterol but the fruits also contain stigmasterols and sitosterols. The phytosterol content is thought to play an important role in the ability of the fruit to lower cholesterol levels (Puente *et al*., 2011). The leaves of *Physalis angulata* have a high concentration of fatty acids (Ogundajo *et al*., 2015). The ascorbic acid content of the fruit is higher compared to those of other common fruits (Rosa *et al*., 2023). However, the ascorbic acid level of *Physalis* is comparable to that of orange (50 mg/100ml) and strawberry (60 mg/100ml) fruits (Ramadan and Mörsel, 2003; Belitz and Grosch, 1999). *P. peruviana* in particular has a high amount of vitamins C, A and B complex (Puente *et al*., 2011).

Physalis plants have been studied for the presence of essential trace elements (minerals). *Physalis* contains phosphorus, zinc, iron, potassium, magnesium and calcium (Rodrigues *et al*., 2009;

Musinguzi *et al*., 2007). The iron content of *P. peruviana* was noted to be higher than in regular sources such as beans (Rodrigues *et al*., 2009). These minerals are thought to be important for fruit health and also confer the fruit with some of its functional properties. Zinc for example is thought to be a non-enzymatic antioxidant and is thought to attribute to the anti-oxidant activity of the fruit even at low concentrations (Puente *et al*., 2011).

2.4 Phytochemical profile of *Physalis* fruits

Phytochemical activity of *Physalis* has majorly been attributed to the presence of withanolides, physalins, phenolics, flavonoids and carotenoids (Puente *et al*., 2011; Wu *et al*., 2006). Withanolides are steroidal lactones that have biological properties such as insect repellant and anti-feedant, anti-tumor, anti-inflammatory, antibacterial, immune-modulatory and cytotoxic activity (Ahmad *et al*., 1999). Lan *et al*. (2009) demonstrated the presence of seventeen withanolides found in *Physalis* including withanolide E and 4-beta-hydroxy-withanolides that have anticancer activity (Lan *et al*., 2009). Carotenoids found in *Physalis* majorly include all-trans-beta-carotene as well as others such as 9-cis-beta-carotene and all-trans-alpha-cryptoxanthin (Puente *et al*., 2011). Physalins which are simply immunosuppressive substances are also found in *Physalis*. Those isolated include; physalins A, B, D and glycosides which have been credited with anti-cancer activity (Wu *et al*., 2004). Due to the high antioxidant properties, the fruit has been used in folk medicine to treat ailments such as malaria, asthma, dermatitis, hepatitis and rheumatism (Wu *et al*., 2004). Phytochemicals such as polyphenols and terpenes have been identified in *P. minima*, *P. angulata* and *P. peruviana* (Kasali *et al*., 2021; Ogundajo *et al*., 2015; Usaizan *et al*., 2014). Among the Malayali tribe in Kolli Hills of India the whole *P. peruviana* plant extract is used in the treatment of skin diseases (Anjalam *et al*., 2016). Other Indian tribes such as Manjoor, Thiashola and Western Ghats use the seeds and leaves of *P. peruviana* to treat glaucoma and jaundice (Sharmila *et al*., 2014). Studies have shown that the fruit of *Physalis* also has antihelminthic properties as well as playing an important role in the strengthening of the optic nerve due to its high vitamin A content (Arun and Asha, 2007).

2.4 Genetic characterization of plants

Molecular characterization studies in plants help to identify and discriminate the species of plants, assess the genetic relatedness of species and describe their germplasm. Genetically diverse plants have better chances of survival in changing environmental conditions (Chacón *et al*., 2016). Any information achieved from genetic diversity studies of natural and cultivated crops helps to identify variation in plants. This is essential for crop breeding programs and improves the value of crops

(Grandillo, 2014). This information also prevents loss of diversity as a result of the extinction of plant species due to the effects of climate change.

2.4.1 Use of molecular markers in genetic characterization of plants

The development and utilization of molecular markers to determine plant genetic diversity is important for molecular genetic studies (Idrees and Irshad, 2014). Markers majorly show polymorphism which can arise due to mutations in the genome loci (Amiteye, 2021). Plant molecular markers are used in the identification of cultivars, genetic diversity assessment and marker assisted crop breeding (Collard *et al*., 2005). The properties of good molecular markers include easy to use and rapid results, easily available, co-dominant inheritance, reproducible, highly polymorphic, recurrent occurrence in the genome and they should be selectively neutral to environmental changes (Idrees and Irshad, 2014). Genetic characterization and diversity can be assessed using dominant molecular markers, co-dominant molecular markers and DNA barcoding techniques. Some of the dominant molecular marker techniques include DNA amplified fingerprinting (DAF), inter-simple sequence repeats (ISSR), random amplified polymorphic DNA (RAPD) as well as amplified fragment length polymorphism (AFLP). Co-dominant molecular marker techniques include simple sequence repeats (SSR), sequence characterized amplified regions (SCARs), cleaved amplified polymorphic sequence (CAPS) as well as restriction fragment length polymorphism (RFLP) (Govindaraj *et al.*, 2015). Co-dominant markers are preferred in genetic diversity studies over dominant markers because they show heterozygosity (Idrees and Irshad, 2014).

2.4.2 DNA barcoding in genetic characterization of plants

DNA barcoding is a novel technique that uses short regions on a DNA molecule to classify species (Hebert *et al*., 2004). It is an important technique for understanding how to conserve the biodiversity of plants (de Vere *et al*., 2015). DNA barcoding of plants is not only important in species discrimination but can also be used to assess intraspecific and interspecific variation during genetic diversity studies (Li *et al*., 2015). DNA barcoding for assessment of genetic diversity requires the presence of individual plant barcodes studies to have successful genetic variation analysis (Li *et al*., 2015).

Terrestrial plants have their DNA barcode markers in the coding regions of chloroplast and mitochondria, especially the part of the genes *rbcL* and *matK* DNA sequences (Li *et al*., 2015). The ideal barcode locus for identification and discrimination of plants is still debatable as the

performance of each barcode differs among different genera of plants. The two chloroplast barcodes *rbcL* and *matK* are considered ideal loci for the identification of plants and were proposed as the preferred plant barcode loci by the Consortium for the Barcode of Life (CBOL) (CBOL Plant Working Group A, 2009). Although *rbcL* and *matK* are considered standard barcodes for plants they do not perform well in species discrimination due to their high rate of conservation within and between species (Li *et al*., 2014). Their low sequence variation provides poorly resolved phylogenetic trees, which means they cannot discriminate species but can authenticate the plant genus of the plants (Li *et al*., 2014). Other plant barcodes include chloroplast barcodes *psbA-trnH*, *atpF*, *atpH* and *ycf 1* (Chase *et al*., 2005, Kress, 2005).

The mitochondrial DNA barcode cytochrome oxidase 1 (CO 1) can be used in the identification of both plant and animal species (Kress, 2005). The DNA barcode CO 1 is largely preserved across species that use oxidative phosphorylation for energy production (Hebert *et al*., 2003). This barcode is rarely used in plant species discrimination due to low genetic variation associated with a low rate of mutation (Kress, 2005). It is however considered a standard barcode in the discrimination of animals (Yao *et al*., 2010). In animals, mitochondrial genes are preferred in species discrimination due to their lack of introns (Yao *et al*., 2010). Other mitochondrial barcodes for animals include *cytb*, 12S ribosomal ribonucleic acid (rRNA) and 16S rRNA (Yang *et al*., 2018).

Nuclear barcodes for plants are ITS1 and ITS2 obtained from the ITS region (Yao *et al*., 2010). ITS2 is considered a secondary barcode for the identification of plants and is preferred over the whole ITS region as it has a short length and high efficiency in PCR amplification (Han *et al*., 2013; Chen *et al*., 2010). It has been proposed that the ITS2 region be used as the standard barcode for phylogenetic analysis and species discrimination of plants (Zhao *et al*., 2015; Chen *et al*., 2010; Shultz and Wolf, 2009). This is due to its high interspecific and low intraspecific variation, available conserved regions for universal barcode design and ease of amplification (Yao *et al*., 2010).

Differences in the efficiency of DNA barcodes can be demonstrated whereby in the study of the genus *Aquilaria* there was more genetic variation observed for the *matK* loci than *rbcL*, and this made species identification using *matK* more robust (Thitikornpong *et al*., 2018). The barcode *matK* has also shown better species discrimination of the genus *Dioscorea* as compared to *rbcL* and *psbA-trnH* (Sun *et al*., 2012). The *matK* barcode is considered strong but not perfect for the identification of plant species (Sun *et al*., 2012). In another study for the species discrimination of jewel orchid accessions *rbcL* gave better species discrimination than *matK* or a combination of the two genes (Ho

et al., 2021). A study on the discrimination of *Calligonum* species using four DNA barcodes (*rbcL*, *matK*, *psbA-trnH* and *trnLF*) showed an inability for any of the individual barcodes to discriminate species (Li *et al*., 2014). However, in a study of the same species, a combination of the three barcodes *rbcL*, *matK* and *trnLF* provided a slight improvement in species discrimination (Li *et al*., 2014). The barcode *psbA-trnH*, a non-coding chloroplast marker within the intergenic region, has effectively identified species rich genera *Pedicularis*, *Primula*, *Rhododendron* and *Parnassia* (Pang *et al*., 2012). A combination of *psbA-trnH* and ITS2 showed better species discrimination than the use of the p*sbA-trnH* barcode alone (Pang *et al*., 2012). ITS2 has mostly shown low amplification and sequencing efficiency as compared to chloroplast barcodes but better resolution of species due to its high interspecific variation (Acharya *et al*., 2022; Tripathi *et al*., 2013). ITS2 can also help assess secondary structures for RNA which helps in understanding the mechanism of RNA molecules and its ability to demonstrate polyphyletic phylogeny can be used in genetic divergence studies (Acharya *et al*., 2022).

2.5 Molecular markers in identification and genetic characterization of *Physalis* plants

Physalis accessions have been characterized using different types of molecular markers such as SNPs, InDels, SSR markers and DNA barcoding (Ralte and Singh, 2021; Muraguri *et al*., 2021; Feng *et al*., 2016; Garzón-Martínez *et al*., 2015; Simbaqueba *et al*., 2011). Few studies have used SSR markers in the characterization of *Physalis* species (Feng *et al*., 2023; Muraguri *et al*., 2021; Simbaqueba *et al*., 2011). A study has shown that identification of plants based on morphological characteristic is more reliable and authentic than the use of SSR markers (Yin *et al*., 2023). This study indicated that SSR molecular markers cannot replace morphological markers but a combination of the two would be more accurate and reliable in plant variety identification (Yin *et al*., 2023). Therefore, SSR markers are not very efficient in species discrimination but can provide robust data on genetic diversity of plants.

Physalis is categorized as an orphan fruit crop, therefore, there is very little information on its genetic diversity. It is important to understand the diversity of *Physalis* as this information will assist in developing crossbreeding strategies and conservation methods for the various species identified (Garzón-Martínez *et al*., 2015). Genetic diversity of *Physalis* has been studied in Colombia where the plant mainly grows as a wild fruit and also cultivated for commercial purposes (Garzón-Martínez *et al*., 2015; Chacón *et al*., 2014). In Colombia, the genetic diversity of the *Physalis* was assessed using single nucleotide polymorphisms (SNPs) and InDels derived from immunity regulated genes

(IRG) and conserved ortholog set (COS) II markers. The studies demonstrated that there were moderate to high genetic diversity in *P. peruviana* and related taxa species in Colombia (Garzón-Martínez *et al*., 2015). Recommendations indicated a need to further analyze the genetic diversity of the fruit in other parts of the world where it is available (Garzón-Martínez *et al*., 2015). Other markers that have been used to assess diversity of *Physalis* include random amplified microsatellites (RAM) for genetic diversity studies of Colombian *Physalis* spp and showed high heterozygosity (Morilla-Paz *et al*., 2011). Molecular markers used in tomato characterization such as SSRs, InDels and COS have also been utilized in determining *Physalis* genetic diversity. However, due to the fact that *Physalis* and tomatoes are only related at the family level, high polymorphism detected (Garzón-Martínez *et al*., 2015). Genetic diversity of *Physalis* spp., from Indonesia has been assessed based on SSR markers and revealed high intraspecies variation amongst the *Physalis* accessions studied (Sadiya *et al*., 2021). The SSR markers have also been used to assess the genetic diversity of *Physalis* in Kenya and identified a high diversity amongst accessions studied (Muraguri *et al*., 2021). However, the species identity of the *Physalis* accessions used in the study was not known. In China, the recent use of SSR markers in assessment of genetic diversity of *P. angulata* has shown considerable diversity among the accessions within the species (Feng *et al*., 2023). There are limited studies that have focused on assessing the genetic diversity of *Physalis* species and their accessions using DNA barcode genes. More work needs to be done on determining interspecific and intraspecific genetic diversity of *Physalis* accessions based on chloroplast and nuclear DNA barcodes. Genetic diversity of *Physalis* plants has also been assessed using morphological and physicochemical traits (Santos *et al*., 2022; Usaizan *et al*., 2018). Understanding the genetic diversity of plants is key in establishing proper breeding programs and ensuring conservation of rare and endangered species and cultivars.

2.6 DNA barcoding in species discrimination and genetic characterization of *Physalis* accessions

The identification of *Physalis* accessions and assessment of their genetic diversity based on DNA barcodes has not been widely studied. However, much of the work that has been done has shown that DNA barcodes have a high efficiency in species discrimination based on BLASTn analysis and phylogenetic analysis (Feng *et al*., 2018; Feng *et al*., 2016). DNA barcodes such as ITS2, *rbcL*, *matK* and *psbA-trnH* have been utilized in the characterization of *Physalis* accessions (Ralte and Singh, 2021; Rosario *et al*., 2019; Feng *et al*., 2018; Feng *et al*., 2016). In one study the barcode *psbA-trnH* was able to facilitate the identification of *Physalis* accessions from China as *P. angulata*, *Physalis alkekengi* var. *franchetti*, *P. pubescens* and *P. peruviana* (Feng *et al*., 2018). In another study of

Physalis accessions from China, the ITS2 RNA secondary structures facilitated the identification of *Physalis greenmanii*, *Physalis hintonii*, *P. alkekengi* and *P. alkekengi* var. *franchetti* (Feng *et al*., 2016). DNA barcoding has been considered more robust in species discrimination of plants than other molecular markers. Combining two DNA barcode markers in species discrimination provides more robust species identification (Parmentier *et al*., 2013).

2.6.1 DNA barcoding in species discrimination of *Physalis* accessions

Physalis accessions in China have been discriminated using the psbA-trnH barcode marker which showed high efficiency at discrimination based on BLASTn analysis and phylogenetic analysis (Feng *et al*., 2018; Feng *et al*., 2016). In a study by Feng et al. (2016), species like *P. greenmanii* and *P. hintonii* were identified. Furthermore, the identification of *Physalis* as an ethno-medicinal plant within the *Solanaceae* family has been documented based on ITS2 and *rbcL* genes (Ralte and Singh, 2021). Another study conducted in China utilized the ITS2 barcode to effectively identify *Physalis* accessions through BLASTn analysis and near distance methods (Feng *et al*., 2016).

2.6.2 DNA barcodes in assessment of genetic diversity of *Physalis* accessions

Due to the status of *Physalis* as an orphan plant, few studies are focusing on its genetic diversity. Information on genetic diversity of *Physalis* plant would provide important details to breeders for selection parents for breeding programs and conservation of endangered species and accessions from extinction (Delgado-Bastidas *et al*., 2019). Conservation of the species *Physalis peruviana* L. in Colombia entails the preservation of domesticated and wild type germplasms in research centers and Universities (Simbaqueba *et al*., 2011). Molecular characterization of *Physalis peruviana* in Colombia has largely been done using SSR markers which are ideal in population genetics studies due to their high co-dominant and polymorphic inheritance (Juyó *et al*., 2015). The intraspecific and interspecific genetic divergence of *Physalis* accessions in China has been studied with the results indicating that the accessions had higher interspecific variation than intraspecific variation based on the *psbA-trnH* barcode (Feng *et al*., 2018). Intraspecific divergence of *Physalis* accessions from China was noted to be low using the ITS2 barcode (Feng *et al*., 2016).

2.7 ITS2 RNA secondary structures in discrimination of plant species

The ITS2 barcode is a spacer nuclear gene in plants that has a high level of interspecific and intraspecific variation and can be used in species discrimination among various plant genera (Anaz *et al*., 2021). This gene is important in ribosome synthesis and its secondary and tertiary structures are

crucial for this function (Zhang *et al*., 2020). Despite its variation among species and its rapid evolution it must maintain specific RNA secondary structures that provide the functionality of the ribosome (Zhang *et al*., 2020). The RNA secondary structure of the ITS2 gene can also facilitate species discrimination above the generic level and this makes ITS2 efficient in species discrimination (Coleman, 2003). ITS2 secondary structure comparison among plant species can also help divulge the level of divergence of genes within species and also provide visual differences of various plant species (Anaz *et al*., 2021).

ITS2 gene sequences and RNA secondary structures have been used in species discrimination of genus *Achyranthes* with 100% success in correct identification of species (Singh *et al*., 2020). ITS1 and ITS2 secondary structures have been used to successfully discriminate *Nepenthes* species with enhanced resolution (Saidon *et al*., 2023). Therefore, the use of ITS2 secondary structure in species discrimination has been widely successful for different plant genera. ITS2 secondary structure discrimination for *Physalis* has not been widely studied but in the few studies done species discrimination was successful. The ITS2 secondary structures have been successfully used in discrimination of *Physalis* species from China such *P. greenmanii*, *P. alkekengi*, *P. hintonii* and *P. alkekengi* var. *franchetti* (Feng *et al*., 2016). Identification of plants species using ITS barcode sequences and their secondary structures can therefore provide an efficient and robust analysis which can be used as a basis for the conservation and utilization of plant genetic resources. No study has been done to use the ITS2 RNA secondary structures to discriminate *Physalis* species in Kenya.

CHAPTER THREE

3.0 Determination of mineral composition, phytochemical content, antioxidant activity and identity of *Physalis* accessions from Sorget forest, Kericho County, Kenya

3.1 Introduction

Indigenous varieties adapted to a particular region are ideal resilient crops for climate change adaptation but are neglected and often lost due to the rapid domestication of commercial cultivars. Restoring such plant species would empower local farmers and provide huge economic and nutritional benefits. One such underutilized wild plant species is of the genus *Physalis* and belongs to the Nightshade (*Solanaceae*) family (Afroz *et al*., 2020). *Physalis* species are native to the Peruvian and Ecuadorian Andes region of South America; hence, it is referred to as the Peruvian gooseberry (Feng *et al*., 2016), although some are also native to Southeast Asia and Eurasia (Wei *et al*., 2012). Studies of *Physalis* in China have identified five species including *P. alkekengi*, *P. angulata*, *P. pubescens*, *P. peruviana*, and *P. minima* and two variants of *P. alkekengi* (Chinese Academy of Sciences, 1978). *Physalis* species such as *P. philadelphica*, *P. peruviana*, and *P. pubescens* are grown in various parts of the world (Sang-Ngern *et al*., 2016). *Physalis* fruits are exported from several countries including Colombia, Australia, New Zealand, Great Britain, Zimbabwe, Kenya, Egypt, South Africa, Madagascar, and South East Asia (Zhang *et al*., 2013; Ramadan and Moersel, 2003). The largest producer, consumer, and exporter of *Physalis* fruits is Colombia (Zhang *et al*., 2013). In Kenya, the fruits are seen in vast numbers as wild and self-propagating plants, commonly in farms during and after the harvesting of maize. Currently, there is growing commercial interest in this fruit crop because of its nutritional and pharmacological properties and health benefits. The key steps involved in the proper utilization of this indigenous plant species are accurate identification and authentication.

Taxonomic identification and morphological characterization are the most common methods of plant authentication, although they are limited by environmental or physiological factors and the developmental phase of the plant species (Menzel, 1951). Employing molecular tools such as DNA barcoding could be more successful in species identification. DNA barcoding requires a short universal DNA sequence that exhibits a sufficient level of variation to discriminate species (Barcaccia *et al*., 2015, Hebert *et al*., 2003). The proposed plant DNA core barcodes from the Consortium for the Barcode of Life (CBOL) Plant Working Group comprise the chloroplast gene large subunit of ribulose bisphosphate carboxylase (*rbcL*) and *matK* with the *psbA-trnH* intergenic sequence and internal transcribed spacer (ITS), and a nuclear gene as the supplement barcode (Kang

et al., 2017). Chloroplast DNA barcodes such as *matK* and *rbcL* have been used in many phylogenetic and plant species identification analyses (Feng *et al*., 2018). The nuclear internal transcribed spacer 2 (ITS2) barcode has also been demonstrated to exhibit 100% species identification and discrimination statistics in plants due to its high intra- and interspecific divergence (Zhao *et al*., 2018). Currently, there is lack of information on the genetic characterization of indigenous *Physalis* plants from the wild in Kenya.

The nutritional and phytochemical profiles of *Physalis* have been studied to a moderate level in the Andes (Ramadan and Moersel, 2007). The *Physalis* fruit contains fat and water-soluble vitamins (B-complex, C, A, E, and K), minerals (mainly phosphorus, magnesium, potassium, zinc, and calcium), sugars (e.g., sucrose, glucose, and fructose), fatty acids (palmitate, oleic, and gamma linoleic acid), phytosterols (sitosterol and stigmasterol), and fiber (pectin) (Puente *et al*., 2011; Ramadan and Moersel, 2003). The fruit pomace (skin and seed) contains 19.3% oil, 3.1% ash, 17.8% protein, 24.5% carbohydrates, and 28.7% crude fiber (Ramadan and Moersel, 2007). The phytochemicals found in *Physalis* include withanolides, physalins, carotenoids, phenolics, and flavonoids (Puente *et al*., 2011). A review study states that the most studied phytochemicals are the polyphenolics, which have antioxidant properties and many health-related benefits (Somani *et al*., 2015). The key polyphenolics include flavonoids, stilebenes, phenolic acids, coumarins, and tannins (Bayir *et al*., 2019). The *Physalis* plant's high content of vitamins, minerals, and antioxidant phytochemicals gives it medicinal properties such as anti-inflammatory activity and this has been specifically determined for *P. alkekengi* species (Hong *et al*., 2015). The characterization of the nutritional and bioactive properties of cultivated and wild fruits of *P. peruviana* growing in the northern Argentinian region identified several types of phytochemicals present in the plant, such as flavonoids and tannins (Bazalar Pereda *et al*., 2019). However, the mineral composition, phytochemical profile, and antioxidant activities of wild fruits of *Physalis* growing in Kenya have not yet been characterized.

The current study aimed to identify and assess the *rbcL* and ITS2 gene barcodes to discriminate indigenous *Physalis* accessions collected from the forest. The mineral and phytochemical content as well as antioxidant activities of ripe fruits were also evaluated to determine if they are nutrient-rich so that they could be used as sustainable resources for the development of biofortified crops and also promoted as a natural source of antioxidants.

3.2 Materials and methods

3.2.1 Sampling of plant material

Leaves and mature fruits of 10 *Physalis* plants were collected based on purposive sampling in April 2019 from different spots in Sorget forest in Londiani area of Kericho County, Kenya (Figure 3.1.). The area is located at an elevation of 2528 m above sea level and a latitude of 0.0684° S and a longitude of 35.5548° E (Appendix 1A).

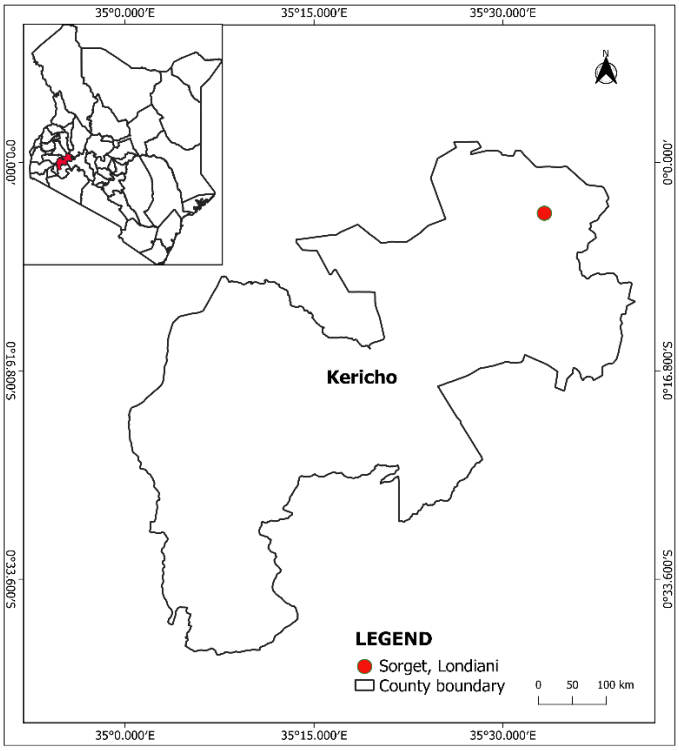


Figure 3.1. Map showing the location of *Physalis* accessions sampling in Londiani, Kericho County, Kenya.

The collected *Physalis* plant samples were identified by the taxonomist Mr. Patrick Mutiso and the samples were preserved in the University of Nairobi herbarium in the Department of Biology (Codes of Voucher Specimens: KP/UON2019/001- KP/UON2019/010). Ten *Physalis* plants were sampled for their fruits and leaves based on the availability and ripeness of the fruits. The maturity of the fruits was assessed based on the color of the fruit. Ripe fruits had a yellow to orange color. The leaves were used for the molecular identification of the plant species, while the fruits were used for the study of the nutritive value and functional attributes. The collected leaf and fruit samples were wrapped with aluminum foil, kept in an ice box and transferred to the Molecular Biology Laboratory in the Department of Biochemistry, University of Nairobi. The leaf samples were kept at −80 °C prior to genomic DNA extraction. The fruit samples were utilized immediately upon arrival at the laboratory for the extraction of phytochemicals and the determination of the mineral content.

3.2.2 Molecular authentication of *Physalis* plants

3.2.2.1 Isolation of genomic DNA from leaves of *Physalis* accessions

Isolation of genomic DNA from leaves of *Physalis* accessions was performed using the cetyltrimethylammonium bromide (CTAB) method Dellaporta *et al*., 1983). Ribonuclease A (RNase, 0.6 mg/mL) was added to the DNA accessions followed by incubation at 37 °C in a water bath for 30 min to eliminate any contaminating RNA. The integrity of the extracted genomic DNA was verified using 0.8% (*w/v*) agarose gel stained with ethidium bromide (0.5 µg/mL) and viewed under UV transilluminator in Gel Doc™ EZ Imaging System (BioRad, Hercules, CA, USA). DNA was stored at −20 °C until use in the molecular analysis.

3.2.2.2 Polymerase chain reaction (PCR) amplification and sequencing

Polymerase chain reaction (PCR) amplification was performed using the DNA barcoding primers *rbcL* and ITS2 (Table 3.1). DNA amplification was conducted using a One Taq® Hot start 2× master mix with standard buffer (New England Biolabs, Ipswich, MA, USA) following the manufacturer's instructions. Amplification was conducted in a Veriti, 96-well Thermal Cycler (Thermo Fischer Scientific, Waltham, MA, USA). Optimization was performed in order to acquire the best conditions for PCR amplification. The annealing temperature for both primers was optimized at the following temperatures: 50 °C, 51 °C, 52 °C, 54 °C, 56 °C, and 58 °C. The best optimum cycling conditions for both primers were used for the PCR amplification of the DNA samples (Table 3.1). The amplicons were confirmed using 1% agarose gel stained with ethidium bromide (0.5 µg/mL) under a UV transilluminator in the Gel Doc™ EZ Imaging System (BioRad, Hercules, CA, USA). Amplicons were cleaned using a gel clean up kit (Applied Biosystems, Thermo Fischer Scientific, Waltham, MA, USA) and sent for sanger sequencing at the University of Nairobi (UoN) Center of Excellence in HIV Medicine (CoEHM) using a 3730 s DNA analyzer (Thermo Fischer, Waltham, MA, USA).

Table 3.1. Oligonucleotide primers used for PCR amplification and optimum PCR cycling conditions

Barcode Region	**Primer Name**	**Primer Sequence (5′ to 3′)**	**PCR Conditions**
ITS2	ITS2-F	CCTTATCATTTAGAGGAAGGAG	1 cycle of 94 °C 5 min; 30 cycles of 94 °C 30 s, 58 °C 45 s, and 72 °C 1 min; 72 °C 7 min
	ITS2-R	TCCTCCGCTTATTGATATGC	
rbcL	*rbcL*-1-F	ATGTCACCACAAACAGAA	1 cycle of 94 °C 5 min; 30 cycles of 94 °C 30 s, 58 °C 45 s, 72 °C 1min; 72 °C 7 min
	rbcL-74-R	TCGCATGTACCTGCAGTAGC	

3.2.2.3 Sequence and phylogenetic analysis

The sequences of each barcode were edited manually in the BioEdit software version 7.2.5.0 (Hall, 1999). The BioEdit software successfully recognized and opened the auto sequencer trace files for the sequenced ITS2 and *rbcL* genes from the *Physalis* accessions. In every instance of sequence editing the chromatogram trace was adjusted with the horizontal and vertical scale bars so as to observe the peaks of the trace. The editable notepad file comprising the sequence was opened and copied to get a duplicate that was edited without changing the original sequence. The called bases on the chromatogram trace were then visualized and editing of the sequence for pseudobases was done by observing the color-coded peaks on the chromatogram trace below the pseudobase and replacing the pseudobase with the correct base. Edited sequences were then saved under the FAST All (FASTA) format in preparation for Basic Local Alignment Search Tool for nucleotides (BLASTn) analysis.

The edited sequences were compared with the available nucleotide sequences in the GenBank database. The sequences were blasted in the NCBI GenBank BLASTn database to determine the sequence homology with other deposited ITS2 and *rbcL* sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi accessed on 10 February 2023). The identification of *Physalis* species was based on the least expected value (E-value), highest query coverage, and similarity percentage. The obtained sequences were also assembled and aligned using the MUSCLE algorithm. Multiple sequence alignment (MSA) of ITS2 and *rbcL* sequences in this study and the reference sequences retrieved from the NCBI database was performed using the MUSCLE software version 3.8 (Edgar,

2004). This MSA was used in the preparation of a phylogenetic tree. The ITS2 and *rbcL* sequences were also aligned separately using MUSCLE, and viewed and trimmed on Jalview version 2.11.2.6 to obtain uniform sequence lengths (Troshin *et al*., 2018; Troshin *et al*., 2011). Two MSAs were prepared separately for ITS2 and *rbcL* sequences without use of their reference sequences and were used in the genetic diversity, nucleotide polymorphism, neutrality test, and automatic barcode gap discovery (ABGD) analysis. All MSAs attained were compressed using ESPript 3 (http://espript.ibcp.fr accessed on 12 February 2023) (Robert and Gouet, 2014). The ITS2 and *rbcL* sequences were submitted to NCBI GenBank through a web-based sequence submission tool and accession numbers were assigned.

Phylogenetic trees were constructed based on the Bayesian inference (BI) method using MrBayes version 3.2.7 (https://nbisweden.github.io/MrBayes/ accessed on 12 February 2023). Statistical analysis was performed using the posterior distribution of the model parameter, which was estimated using the Markov chain Monte Carlo (MCMC) method (Ronquist *et al*., 2012; Ronquist and Huelsenbeck, 2003; Huelsenbeck and Ronquist, 2001). MCMC sampling was performed over 18,000,000 generations at a sampling frequency of 1000 and the first 25% (relburnin = yes burninfrac = 0.25) of samples were discarded when estimating the posterior probabilities of the trees. After 18,000,000 generations, the analysis was stopped when the average standard deviation of the split frequencies was less than 0.01 and tree parameters were summarized. The constructed phylogenetic trees were visualized and modified using the FigTree software version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/ accessed on 12 February 2023).

3.2.3 Analysis of genetic divergence

DNA divergence within *Physalis* accession populations based on ITS2 and *rbcL* sequences was determined using the DnaSP software version 6.12.03 (Kartartsev, 2011). The multiple sequence alignment (MSA) for *Physalis* accessions based on either the ITS2 or *rbcL* marker was uploaded into the software and various parameters for the divergence were determined. The number of polymorphic segregating sites (S), nucleotide diversity, and the total number of substitutions were assessed as outlined by the Jukes and Cantor algorithm on DnaSP.

3.2.4 Determination of genetic distance within *Physalis* accessions

Intraspecific genetic distances and the overall mean distance of *Physalis* accessions based on the ITS2 and *rbcL* sequences were determined using the Kimura 2 parameter (K2P) model with the gamma distribution and a gamma parameter of 0.27 using MEGA version 11.0 (Kumar *et al*., 2018).

Sequence genetic distance was determined using multiple sequence alignments for *Physalis* accessions based on ITS2 and *rbcL* markers.

3.2.5 Nucleotide polymorphism

DNA polymorphisms of the ITS2 and *rbcL* sequences were assessed in all the *Physalis* accessions. The DNA sequence Polymorphism (DnaSP) software version 6.12.03 was utilized in the DNA polymorphism analysis for ITS2 and *rbcL* sequences of all *Physalis* accessions. The DNA polymorphism parameters determined were polymorphic segregating sites, singleton and parsimony informative sites, the nucleotide diversity, and the average number of nucleotide differences.

3.2.6 Tajima's neutrality tests

Tajima's neutrality test for both ITS2 and *rbcL* sequences of *Physalis* accessions were determined to estimate the frequency of mutations among species (Tajima, 1989). The Tajima's neutrality test determined the Tajima D value among the ITS2 and *rbcL* sequences of *Physalis* accessions using the MEGA 11.0 software (Tamura *et al*., 2021; Nei and Kumar, 2000). The analysis involved nine and ten ITS2 and *rbcL* sequences of *Physalis* accessions, respectively. The codon positions included were 1st + 2nd + 3rd + noncoding for the *rbcL* gene sequences. All ambiguous positions were eliminated for each sequence pair (pairwise deletion option) in both the analysis based on ITS2 and *rbcL* genes. There was a total of 399 and 614 positions in the final dataset for both the ITS2 and *rbcL* genes, respectively. The MSAs utilized in this analysis were similar those utilized for genetic diversity and DNA polymorphism studies based on ITS2 and *rbcL* sequences.

3.2.7 Analysis of DNA barcoding gap and intraspecific distance

In order to delimit the *Physalis* species based on their intraspecific divergence within a population, the automatic barcode gap discovery (ABGD) method described by (Puillandre *et al*., 2012) was utilized in this study. Multiple sequence alignments utilized in genetic diversity analysis were also utilized for the ABGD analysis of ITS2 and *rbcL* sequences of *Physalis* accessions. The ITS2 and *rbcL* multiple sequence alignments were separately inputted into the ABGD website (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html accessed on 17 February 2023) and the distance analysis was performed based on the K80 Kimura measure of distance. The default value for the relative gap width (X) was set at 1.5. Moreover, *p* values of intraspecific divergence were set at a prior minimum (P_{min}) and prior maximum (P_{max}) divergence of intraspecific diversity of 0.001 and 0.1, respectively. Default settings were utilized for all other parameters.

3.2.8 Analysis of mineral content in ripe fruits

The analysis of macrominerals (calcium (Ca), magnesium (Mg), potassium (K), and sodium (Na)) and trace elements (iron (Fe), zinc (Zn), manganese (Mn), copper (Cu), lithium (Li), and nickel (Ni)) was performed according to the method described by (Hernández *et al*., 2005). All the analyses were performed in triplicate. Each fruit sample (1 g) was digested in 5 mL of nitric acid, which was made up to 25 mL using distilled water. The mixture was heated on a hot plate until a third of the volume was left, which was filtered with Whatman filter paper No. 1. The filtrate was appropriately diluted and analyzed for mineral elements using an atomic absorption spectrophotometer (AAS) (Shimadzu, Kyoto, Japan). The wavelengths used for the analysis of each mineral were as follows: Ca—422.42 nm; Zn—213.52 nm; Cu—324.53 nm; Na—588.88 nm; Mg—285.04 nm; Fe—248.23 nm; K—766.74 nm; Ni—231.90 nm; Mn—27.03 nm; Li—670.85 nm. The results were expressed in ppm (1 ppm = 1 mg/L) of a sample of dry weight (DW). All the experiments were carried out three times with different fruit accessions.

3.2.9 Determination of phytochemical content

3.2.9.1 Estimation of total polyphenol content (TPC)

The estimation of phenol in fruit extracts was assayed using the Folin Ciocalteau method (Singleton *et al*., 1999). In a test tube with 2.25 mL of 10% Folin Ciocalteau reagent, 1 mL of filtrate of ethanolic fruit extract was added, mixed thoroughly, and allowed to settle at 23 ± 2 °C for 5 min. To the mixture, 2.25 mL of sodium bicarbonate solution (60 g/L) was added, vortexed, and incubated at 23 ± 2 °C for 90 min. The absorbance was measured at 725 nm using a spectrophotometer (Shimadzu, Kyoto, Japan). A blank was prepared using the same method, but the fruit extract was replaced with sterile distilled water. A standard curve was used to determine the phenol content using gallic acid. All the experiments were carried out three times with different fruit accessions. The TPC was assessed as mg/mL of gallic acid equivalents per gram of fruit extract.

3.2.9.2 Estimation of Total Tannin Content (TTC)

The content of tannins in the fruit samples was assayed using the Folin Ciocalteau method (Singleton *et al*., 1999). In a test tube, 0.25 mL of 10% Folin Ciocalteau reagent was added to 0.1 mL of ethanolic fruit extracts and vortexed. The mixture was allowed to settle at 23 ± 2 °C for 5 min, and then 1.25 mL of sodium hydroxide was added and the mixture was incubated at room temperature for 40 min. Absorbance was measured at 725 nm using a spectrophotometer. A blank was prepared using the same method, but the fruit sample was replaced with sterile distilled water. A standard curve was

then prepared for the estimation of the tannin content from *Physalis* fruits using tannic acid. All the experiments were carried out three times with different fruit accessions. The TTC was assessed as mg/mL of tannic acid equivalents per gram of fruit extract.

3.2.9.3 Estimation of total flavonoid content (TFC)

The determination of the flavonoid content in the fruit extracts was performed using the aluminum chloride ($AlCl_3$) colorimetric method (Chang *et al*., 2002). In a test tube, 1 mL of ethanolic *Physalis* fruit extract was added to 4 mL of water and 0.3 mL of 5% sodium nitrate solution. The content was mixed thoroughly followed by the addition of 0.6 mL of aluminum chloride. The mixture was mixed thoroughly and incubated for 6 min at 23 ± 2 °C before the addition of 2 mL sodium hydroxide. A precipitate was formed on mixing, which was centrifuged and the absorbance of the supernatant was measured at 510 nm. A blank was prepared using the same method, but the fruit extract was replaced with sterile distilled water. A standard curve was prepared for the estimation of the flavonoid content using rutin. All the experiments were carried out three times with different fruit accessions. The TFC was assessed as mg/mL of rutin equivalents per gram of fruit extract.

3.2.10 Estimation of antioxidant activity

The antioxidant activity was measured using 2, 2-diphenyl-2-picrylhydrazyl (DPPH) and hydrogen peroxide radical scavenging in vitro assays. The DPPH radical scavenging (RS) assay was performed based on the radical degradation method described by Brand-Williams *et al*. (1995) with some modifications. The sample extracts (0.5 mL) were mixed with 0.1 mM DPPH radical solution (0.3 mL) prepared in an ethanol solution. Color change from deep violet to light yellow was observed and the absorbance was measured at 517 nm using a UV-Visible spectrophotometer after 100 min of reaction in the dark. The blank was prepared using 3.3 mL of ethanol and 0.5 mL of the sample. A control was prepared using 3.5 mL of absolute ethanol and 0.3 mL of the DPPH radical solution and the absorbance was measured. The percentage inhibition of the DPPH radical of the sample extract relative to the control was used to determine the antioxidant capacity using the equation:

$$AA\% = 100 - \left(\frac{\text{Abs sample} - \text{Abs blank}}{\text{Abs control}}\right) * 100$$

The hydrogen peroxide scavenging assay was performed using a method determined by Ruch *et al*. (1989) with a few modifications. In a test tube, 0.5 mL of the sample extract was mixed with 4 mL of 4 mM hydrogen peroxide solution prepared in 0.1 M phosphate buffer (pH 7.4). The mixture was incubated for 10 min at room temperature and the absorbance was measured at 230 nm using a UV-Visible spectrophotometer. A blank was prepared using phosphate buffer and sample extract. A control was prepared using phosphate buffer and hydrogen peroxide and the absorbance was

measured. The hydrogen peroxide radical scavenging (HRS) activity percentage was calculated using the equation:

$$HRS\% = \left(\frac{\text{Abs control} - \text{Abs sample}}{\text{Abs control}}\right) * 100$$

3.2.11 Statistical analysis

The results for the antioxidant activity, and mineral and phytochemical content of the *Physalis* accessions obtained were reported as mean ± standard deviation (SD) using SPSS version 20 (IBM SPSS, 2010). All the measurements were performed in triplicate. One-way analysis of variance (ANOVA) was performed using the statistical software SPSS version 20 (IBM SPSS, 2010). The means from all analyses were separated by Tukey's HDS multiple comparisons test at $\alpha = 0.05$. Regression and correlation analysis was also performed to determine the impact of polyphenols on the antioxidant activity. Regression was analyzed using the statistical software SPSS version 20 IBM SPSS (2010) while correlation studies were analyzed using Microsoft excel version 2016.

3.3 Results

3.3.1 Amplification and sequencing success rate

The PCR amplification results in both ITS2 and *rbcL* regions achieved success rates of 100% (Table 3.2 and Appendix 1B and 1C). The lengths of the ITS2 and *rbcL* sequences were in the range of 301–663 bp and 520–733 bp, respectively. Images of the ITS2 and rbcL amplicons are indicated in Figure 3.2. and 3.3. respectively. The average lengths of ITS2 and *rbcL* sequences were 561 bp and 616 bp, respectively. The GC contents of the ITS2 and *rbcL* sequences were in the ranges of 60%–65.2% and 42.7%–43.9%, respectively. The average GC content of ITS2 sequences was 61.1%, which was significantly higher than that of the *rbcL* sequences (43.1%). The sequencing success rate was 99% and 100% for ITS2 and *rbcL* sequences, respectively (Table 3.2). Therefore, there were 9 sequences for the ITS2 genes and 10 sequences for the *rbcL* genes of *Physalis* accessions from Sorget forest studied.

Table 3.2. Efficiency of PCR amplification and sequencing for *Physalis* accessions based on ITS2 and *rbcL* barcode genes

Barcode Region	**Samples Tested (n)**	**Number of Amplicons Produced**	**Number of Sequences Produced**	**Amplification Efficiency (%)**	**Sequencing Efficiency (%)**	**Alignment Length (bp)**	**Mean Sequence Length (bp)**	**GC Content (%)**
ITS2	10	10	9	100	99	663	561	61.1
rbcL	10	10	10	100	100	730	616	43.1

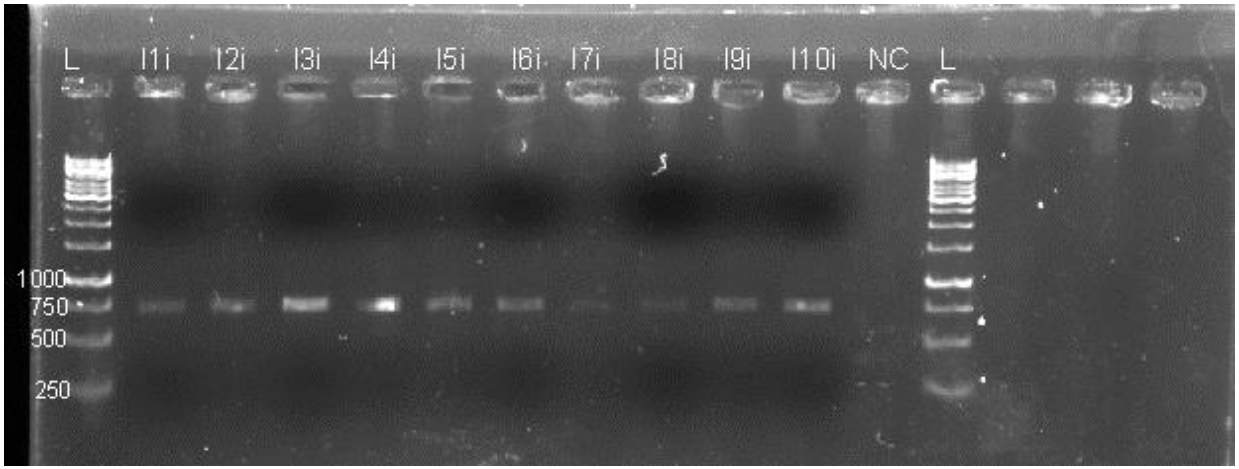


Figure 3.2. Image of PCR amplicons of *Physalis* accessions from Londiani based on the ITS2 marker**.**

The letter L in Figure 3.2 above represents the DNA ladder (1kb DNA ladder, Sigma Aldrich, St. Louis, Missouri, USA), l1i-l10i represents the *Physalis* accessions amplicons based on ITS2 sequence amplification, NC represents the negative control for the experiment.

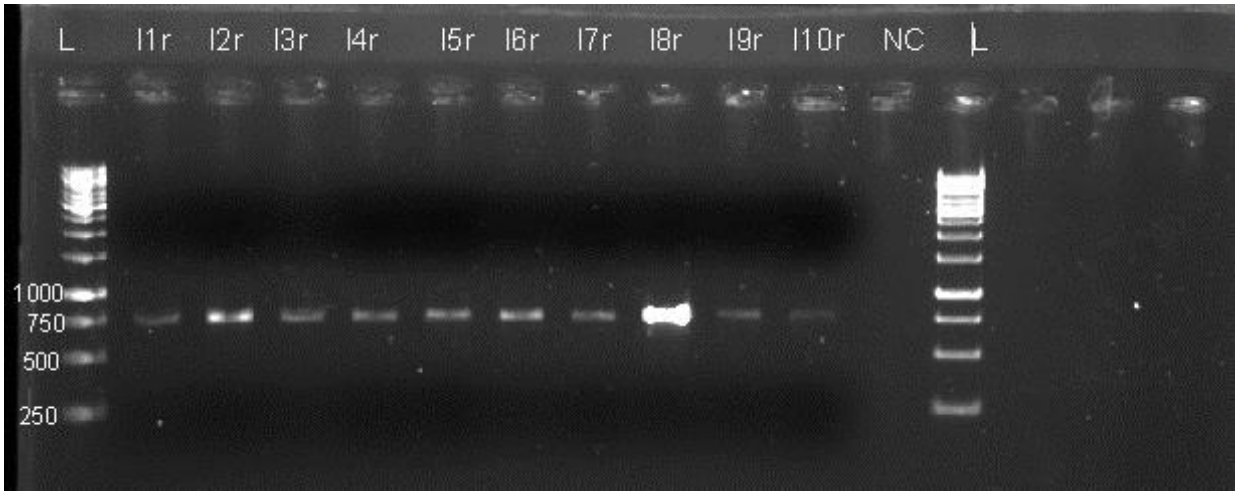


Figure 3.3. Image of PCR amplicons of *Physalis* accessions from Londiani based on the *rbcL* marker.

The letter L in Figure 3.3 above represents the DNA ladder, l1r-l10r represents the *Physalis* accessions amplicons based on *rbcL* sequence amplification, NC represent the negative control for the experiment.

3.3.2 Species discrimination based on BLASTn analysis

According to the BLASTn analysis of ITS-2 sequences, seven of the nine *Physalis* accessions were identified as *P. purpurea*, and one was identified as *P. peruviana* and one as *Physalis aff philadelphica* (Table 3.3). The BLASTn analysis of *rbcL* sequences identified all the *Physalis* accessions as *P. minima* (Table 3.3). The percentage identity for the nine *Physalis* accessions based on ITS2 sequences ranged from 86.00 to 94.4%. The percentage identity for 10 *Physalis* accessions based on *rbcL* sequences ranged from 99.86 to 100%. Eight out of the ten *rbcL* sequences of *Physalis* accessions gave a 100% sequence similarity with *P. minima* (NC_048515.1 from the GenBank). Two accessions (OQ507154.1 and OQ507156.1) were also identified as *P. minima* (NC_048515.1 from the GenBank) with a similarity identity of 99.59 and 99.86%, respectively (Table 3.3).

Table 3.3. BLASTn analysis results for *Physalis* accessions based on ITS2 and *rbcL* markers.

Londiani 1	*P. minima*	NC_048515.1	0.0	100	OQ507152.1	-	-	-	-	-
Londiani 2	*P. mimima*	NC_048515.1	0.0	100	OQ507153.1	*P. purpurea*	MH763740.1	0.0	92.98	OQ372021.1
Londiani 3	*P. minima*	NC_048515.1	0.0	99.59	OQ507154.1	*P. purpurea*	MH763740.1	0.0	94.06	OQ372022.1
Londiani 4	*P. minima*	NC_048515.1	0.0	100	OQ507155.1	*P. peruviana*	AY665914.1	0.0	97.41	OQ372023.1
Londiani 5	*P. minima*	NC_048515.1	0.0	99.86	OQ507156.1	*P. purpurea*	MH763740.1	0.0	94.45	OQ372024.1
Londiani 6	*P. minima*	NC_048515.1	0.0	100	OQ507157.1	*P. aff. philadelphica*	AY665868.1	2e-144	91.35	OQ372025.1
Londiani 7	*P. minima*	NC_048515.1	0.0	100	OQ507158.1	*P. purpurea*	MH763740.1	1e-148	86.00	OQ372026.1
Londiani 8	*P. minima*	NC_048515.1	0.0	100	OQ507159.1	*P. purpurea*	MH763740.1	0.0	93.07	OQ372027.1
Londiani 9	*P. minima*	NC_048515.1	0.0	100	OQ507160.1	*P. purpurea*	MH763740.1	0.0	88.96	OQ372028.1
Londiani 10	*P. minima*	NC_048515.1	0.0	100	OQ507161.1	*P. purpurea*	MH763740.1	0.0	90.16	OQ372029.1

3.3.3 Multiple sequence alignment

The multiple sequence alignment of combined ITS2 and *rbcL* gene sequences and retrieved sequences from the BLASTn analysis based on MUSCLE had a sequence length of 730 bp. The MSA is presented in Appendix 2A. A high rate of substitution mutations was noted for the ITS2 sequences, while very few substitutions, deletions, and insertion mutations were noted for the *rbcL* sequences. The MSA of ITS2 sequences of the nine *Physalis* accessions had a sequence length of 399 bp (https://espript.ibcp.fr/ESPript/temp/1397454355/0-0-1688383432-esp.pdf) (Figure 3.4). There was a high rate of substitution and deletion mutation within this MSA. A deletion point mutation was observed at position 9 of the MSA whereby, in *Physalis* accessions OQ372026.1 and OQ372023.1, the nucleotide thymine was deleted while all other accessions contained a thymine at this position (Figure 3.4). Another deletion point mutation was noted at position 46 of the MSA whereby guanine was deleted for the *Physalis* accessions OQ372023.1, OQ372026.1, OQ372027.1, and OQ372029.1 (Figure 3.4). These *Physalis* accessions also had a deletion macrolesion mutation of eight nucleotides from position 179 to 187. Both transition and transversion point mutations were also identified in the alignment. At position 107 of this MSA, there was a transition point mutation for the *Physalis* accession OQ372023.1, whereby guanine replaced adenine. A transversion point mutation was observed at position 119 for the *Physalis* accession OQ372023.1 whereby guanine replaced the thymine found on all other sequences. Insertion mutations were not identified on this MSA (Figure 3.4). The multiple alignment of *rbcL* sequences of 10 *Physalis* accessions had a sequence length of 614 bp (https://espript.ibcp.fr/ESPript/temp/1129027434/0-0-1688384286-esp.pdf) (Figure 3.5). There were no deletions, insertions or substitution mutations noted in this MSA (Figure 3.5). All *Physalis* accessions had similar sequences with the key difference being the length of the sequences.

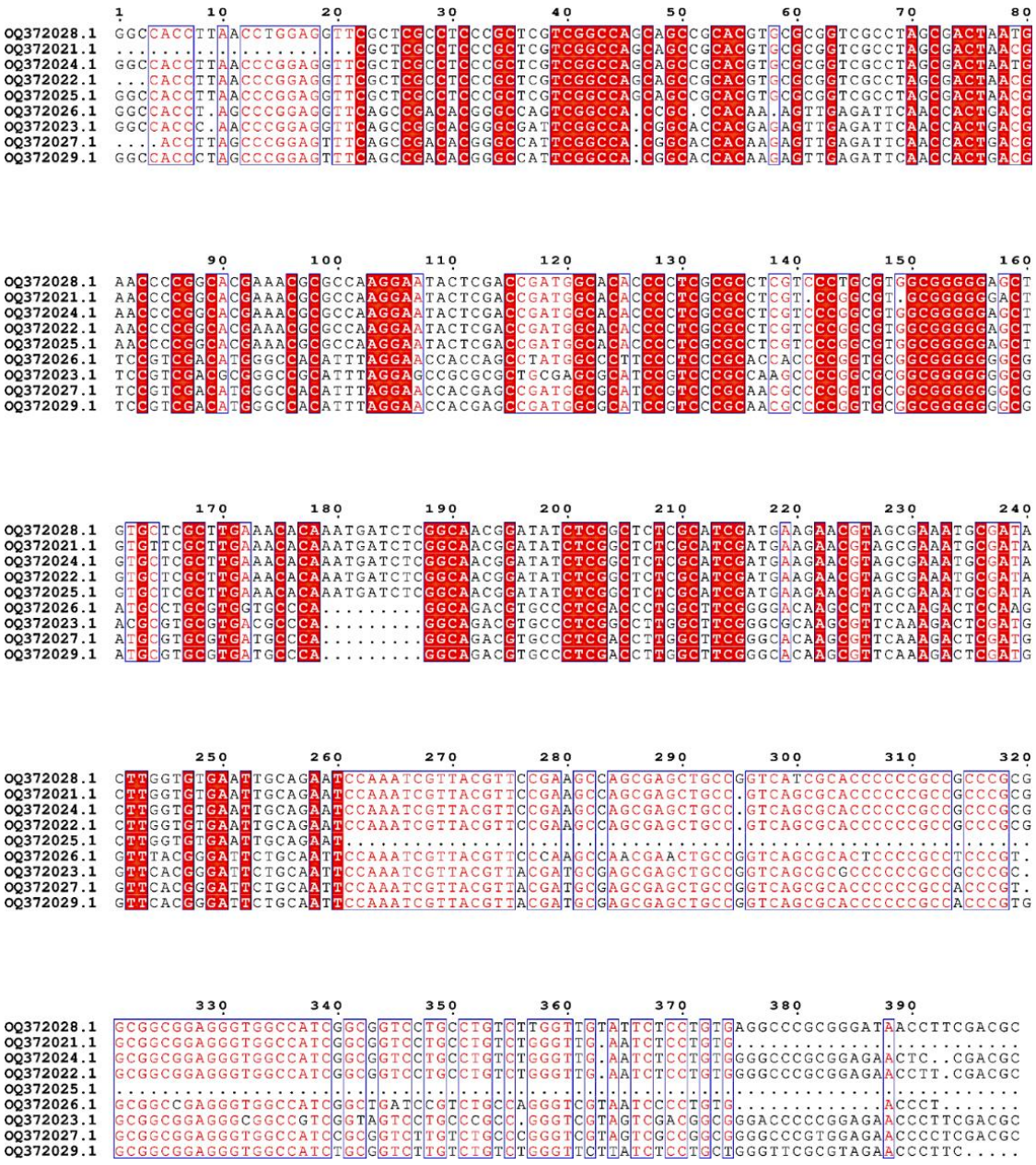


Figure 2.4. Multiple sequence alignment for *Physalis* accessions based on ITS2 gene only

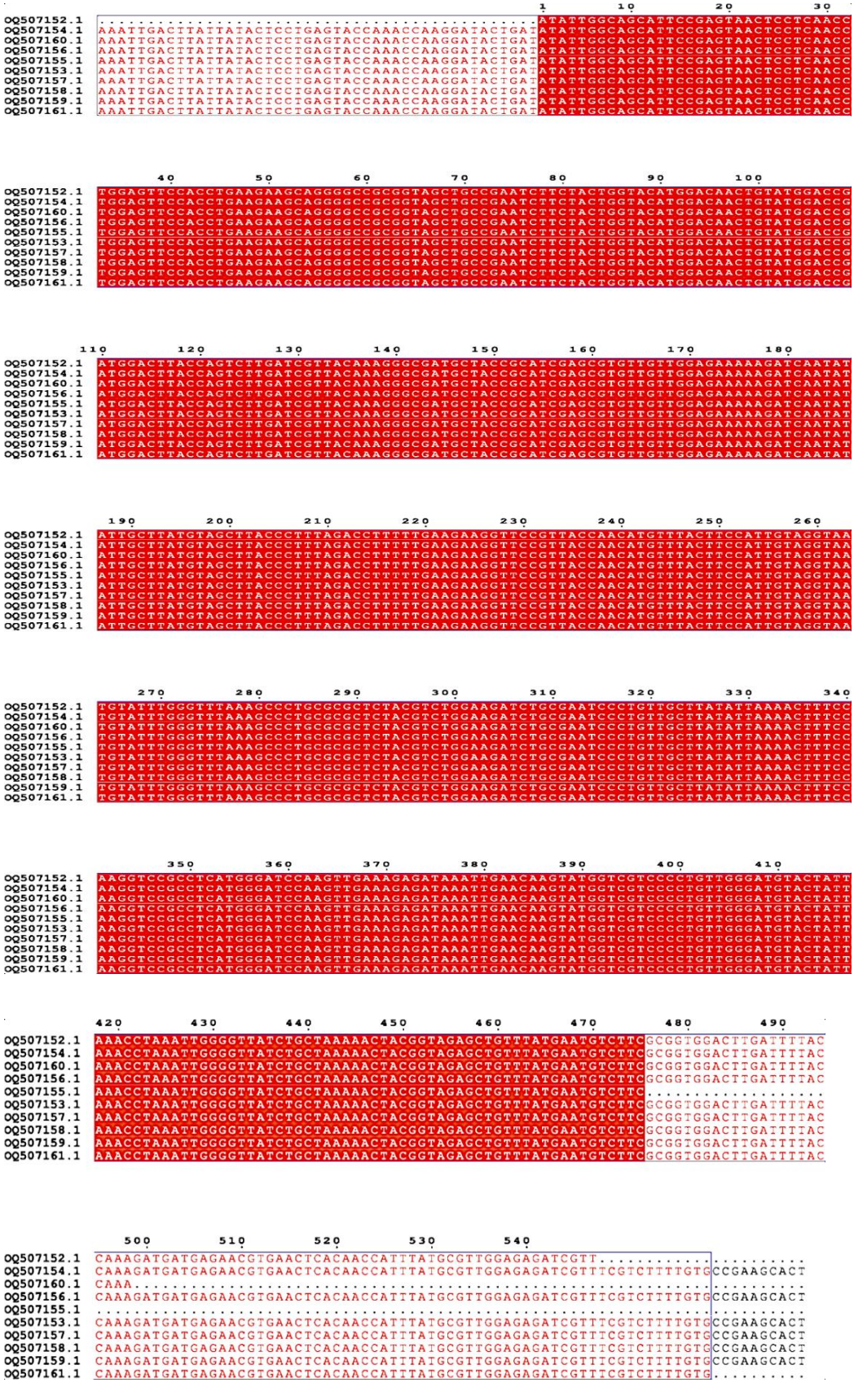


Figure 3.5. Multiple sequence alignment of *Physalis* accessions based on *rbcL* gene only

3.3.4 *Physalis* species identification based on phylogenetic analysis

The phylogram tree was constructed based on Bayesian inference and the combination of ITS2 and *rbcL* sequences. All the *Physalis* accessions were clustered together based on the *rbcL* gene sequences (Figure 3.6). The *Physalis rbcL* genes did not form clades with their reference sequences on the phylogram and there was no species discrimination. Therefore, the *Physalis rbcL* gene sequences species names (*P. minima*) indicated in the phylogram were based on the BLASTn analysis. Based on the ITS2 gene sequences, the *Physalis* accessions were identified with a percentage posterior probability of 100 as *P. purpurea* (Figure 3.6) and the sequences were deposited in GenBank with accession numbers OQ372021.1–OQ372029.1.

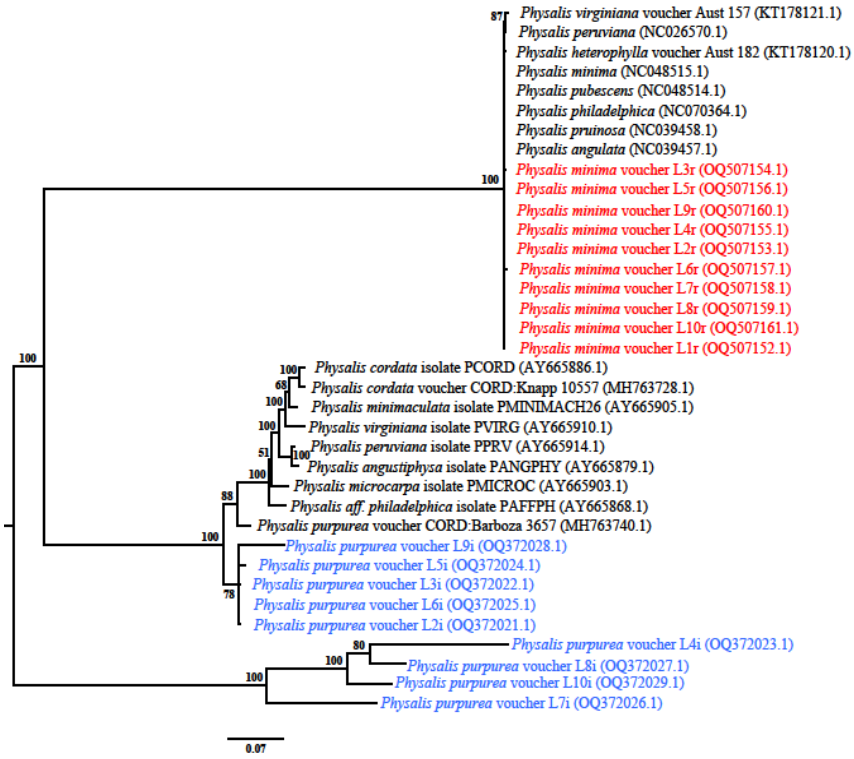


Figure 3.6. Phylogenetic tree generated by MrBayes for the *Physalis* accessions based on ITS2 and *rbcL* gene sequences

Different colors are used to represent *Physalis* accessions (experimental) as the variant species and reference sequences used in this study. Black represents the two clusters and sequences retrieved from GenBank for both ITS2 and *rbcL* genes. Blue represents *P. purpurea* sequences and red represents *P. minima* sequences. The values above the branches represent the percentage posterior probability statistic from the MrBayes phylogram.

3.3.5 Intraspecific divergence of *Physalis* accessions

DNA divergence was determined based on the number of polymorphic (segregating) sites (S), nucleotide diversity, and the total number of substitutions. Based on ITS2 gene sequences, nucleotide

diversity, the total number of nucleotide substitutions, and the number of polymorphic (segregating) sites were 0.27629, 134, and 124, respectively. There was no nucleotide diversity (0), nucleotide substitutions (0.00000), or polymorphic segregating sites (0) among the *Physalis* accessions based on *rbcL* gene sequences.

3.3.6 Genetic distance within *Physalis* accessions

The genetic distance within *Physalis* accessions based on ITS2 and *rbcL* gene sequences was assessed. The overall average genetic distance among the *Physalis* accessions based on ITS2 gene sequences was 1.67 ± 0.77, while there was no genetic distance based on *rbcL* gene sequences. The mean average genetic distance within (intraspecific) *Physalis* species was 1.67 ± 0.84 and 0 based on ITS2 and *rbcL* gene sequences, respectively.

3.3.7 Nucleotide polymorphism and genetic diversity of *Physalis* accessions

There were 124 segregating sites identified with ITS2 gene sequences, while no segregating sites were recorded for the *rbcL* gene sequences (Table 3.4). The lack of segregating polymorphic sites within a population is an indication that all plants within this population are identical. The high number of polymorphic segregating sites for the ITS2 *Physalis* gene sequences is an indication that this gene is highly diverse among the *Physalis* accessions and has undergone differentiation. There were 21 singleton and 103 parsimony sites identified among the segregating sites of *Physalis* accessions based on the ITS2 gene sequences.

Table 3.4. DNA polymorphism of *Physalis* accessions based on ITS2 and *rbcL* markers.

	ITS2			*rbcL*		
Polymorphic sites/Segregation sites (S)	124	Position in the gene	Variants	0	Positions in the gene	Variants
Singleton	21	49,58,90,107,116,117,118,119,120,125,138,139,144,158,162,164,172, 219,225,237,239	2	0		2
Parsimony informative sites	103	23,24,25,30,32,33,37,38,48,51,59,61,62,64,65,66,67,68, 69,71,7377,79, 81,82,84,85,88,91,93,94,95,97,99,100, 101,102,108,109,111,112,114, 126,129,132,136,140, 146,148,157,160,161,166,169,174,176,192, 193, 194,196,197,198,199,204,206,207,209,212,216,217,218,220,221,223, 227,228,229,230,232,234,235,240,241, 245,246,248,251,253,254, 255, 256,257,259 28,36,52,110,113,123,137,165,173,244	2 3	0		2 3
Nucleotide diversity (Pi)	0.27629			0.00000		
Average number of nucleotide differences (k)	61.889			0.000		
Sequence length (base pairs)	399			614		
Number of sequences	9			10		

3.3.8 Tajima's neutrality test

Tajima's neutrality test was performed for both the ITS2 and *rbcL* gene sequences to assess the selection and nucleotide diversity of the *Physalis* accessions. The number of segregating sites (S) for ITS2 and *rbcL* were 180 and 0, respectively. The Tajima values of the *Physalis* accessions based on the ITS2 and *rbcL* gene sequences were 0.779171 and 0, respectively. The Tajima D of the ITS2 sequences indicated a negative selection pressure in the *Physalis* population. The nucleotide diversity of the ITS2 and *rbcL* sequences based on Tajima's test was 0.190894 and 0, respectively, indicating a variation in the ITS2 barcode region of *Physalis* accessions.

3.3.9 Genetic differences and barcoding gap analysis

Automatic barcode gap discovery (ABGD) results generated by the K80 Kimura measure of distance (K2P) based on the ITS2 gene sequences for *Physalis* accessions were used to assess the presence of a barcoding gap. The *rbcL* gene sequences for *Physalis* accessions were not able to provide results on barcoding as there was no variation in the sequences at an intraspecific and interspecific level. Based on the ITS2 gene sequences, all pairwise distances were ranked by increasing the distance values from 0.02 to 0.70 and three barcoding gaps were detected (Figure 3.7). The first and smallest barcode gap was observed between distances of 0.15 (15%) and 0.18 (18%) (Figure 3.7). The second and largest barcode gap was evident between the distances of 0.22 (22%) and 0.44 (44%) (Figure 3.7). The third barcode gap was found between the distances of 0.59 (59%) and 0.70 (70%).

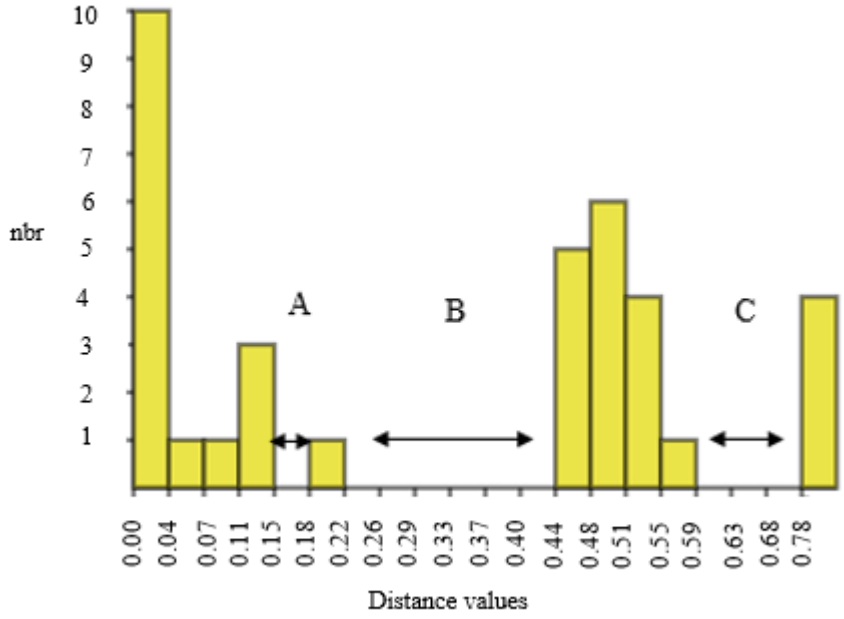


Figure 3.7. A histogram indicating the hypothetical distribution of pairwise differences of ITS2 gene sequences for nine *Physalis* accessions.

Low divergence is presumably intraspecific divergence, whereas higher divergence indicates interspecific divergence. The abbreviation nbr on the *y*-axis of the histogram stands for the number of pairwise comparisons. A, B and C represent the three barcode gaps identified.

3.3.10. Mineral analysis

The mineral content of *Physalis* accessions was determined for macro- and microminerals (Tables 3.5, 3.6, Appendix 3A and 3B). The highest macromineral content among the *Physalis* accessions was noted for potassium at a mean of 527.778 ± 260.526, while the lowest was the magnesium content at a mean of 33.911 ± 29.942 (Table 3.5). Sodium and calcium contents were at moderate levels in *Physalis* accessions with means of 377.46 ± 147.193 and 128.121 ± 20.976, respectively (Table 3.5).

Table 3.5. Macro-mineral content of fruits of *Physalis* accessions.

Data are expressed as mean ± SD of three independent accessions. CV: coefficient of variation; CV: coefficient of variation; n = 3; means followed by single letters in a column differ significantly at a 5% level of significance.

Sample ID	**ITS Accession Number in the GenBank**	**Ca (ppm)**	**Na (ppm)**	**K (ppm)**	**Mg (ppm)**
L1	OQ507152.1	145.493 ± 6.087 ^{aa}	445.378 ± 51.116 ^{aa}	352.941 ± 24.758 ^{aa}	61.056 ± 93.957 ^{aa}
L2	OQ372021.1	58.700 ± 7.451 ^{a}	275.910 ± 18.080 ^{aa}	247.549 ± 8.947 ^{aa}	24.247 ± 20.009 ^{aa}
L3	OQ372022.1	121.803 ± 26.283 ^{aa}	175.070 ± 49.239 ^{aa}	497.549 ± 14.036 ^{aa}	13.604 ± 6.519 ^{aa}
L4	OQ372023.1	77.778 ± 2.618 ^{aa}	208.687 ± 81.942 ^{aa}	470.588 ± 11.725 ^{aa}	8.342 ± 6.455 ^{aa}
L5	OQ372024.1	81.132 ± 9.804 ^{aa}	263.306 ± 14.761 ^{aa}	811.275 ± 77.914 ^{aa}	71.532 ± 43.883 ^{aa}
L6	OQ372025.1	140.042 ± 26.804 ^{aa}	441.176 ± 21.091 ^{aa}	681.372 ± 37.395 ^{aa}	24.346 ± 22.820 ^{aa}
L7	OQ372026.1	131.447 ± 3.328 ^{aa}	456.583 ± 47.782 ^{aa}	450.981 ± 12.484 ^{aa}	19.381 ± 9.648 ^{aa}
L9	OQ372028.1	147.170 ± 23.966 ^{aa}	410.364 ± 25.673 ^{aa}	823.530 ± 24.961 ^{aa}	61.520 ± 48.887 ^{aa}
L10	OQ372029.1	133.962 ± 22.440 ^{aa}	380.952 ± 10.052 ^{aa}	414.216 ± 22.517 ^{aa}	21.169 ± 17.301 ^{aa}
	Mean	128.121 ± 20.976	377.46 ± 14.193	527.778 ± 26.526	33.911 ± 29.942
	CV	16.372%	39.000%	49.363%	88.296%

The micro-mineral content of *Physalis* accessions was also determined (Table 3.6). The highest micro-mineral content was noted for zinc at a mean of 24.364 ± 12.572 (Table 3.6). The lowest micro-mineral content was noted for lithium at a mean of 0.047 ± 0.024 (Table 3.6). A moderate micro-mineral content was noted for iron, manganese, copper, and nickel (Table 3.6).

Table 3.6. Trace element content in fruits of *Physalis* accessions.

Sample ID	Accession Number	Fe (ppm)	Zn (ppm)	Ni (ppm)	Cu (ppm)	Li (ppm)	Mn (ppm)
L1	OQ507151.1	4.597 ± 3.081 ^{aa}	17.534 ± 3.369 ^{aa}	0.214 ± 0.000 ^{aa}	0.015 ± 0.006 ^{aa}	0.035 ± 0.022 ^{aa}	0.565 ± 0.258 ^{aa}
L2	OQ372021.1	6.398 ± 2.543 ^{aa}	7.618 ± 2.702 ^{aa}	0.166 ± 0.083 ^{aa}	0.158 ± 0.050 ^{aa}	0.079 ± 0.060 ^{aa}	0.491 ± 0.205 ^{aa}
L3	OQ372022.1	5.806 ± 2.864 ^{aa}	8.538 ± 4.987 ^{aa}	0.357 ± 0.124 ^{aa}	0.270 ± 0.079 ^{aa}	0.019 ± 0.004 ^{aa}	1.139 ± 0.467 ^{aa}
L4	OQ372023.1	5.780 ± 2.215 ^{aa}	84.663 ± 37.191 ^{a}	1.048 ± 0.527 ^{a}	0.427 ± 0.413 ^{aa}	0.203 ± 0.091 ^{a}	0.954 ± 0.423 ^{aa}
L5	OQ372024.1	6.317 ± 1.391 ^{aa}	7.771 ± 0.176 ^{aa}	0.167 ± 0.109 ^{aa}	1.322 ± 0.468 ^{aa}	0.021 ± 0.013 ^{aa}	2.102 ± 0.135 ^{aa}
L6	OQ372025.1	6.129 ± 1.268 ^{aa}	35.276 ± 24.020 ^{aa}	0.428 ± 0.189 ^{aa}	1.809 ± 1.523 ^{a}	0.010 ± 0.008 ^{aa}	1.250 ± 0.074 ^{aa}
L7	OQ372026.1	6.640 ± 1.341 ^{aa}	7.311 ± 1.240 ^{aa}	0.476 ± 0.289 ^{aa}	1.089 ± 0.161 ^{aa}	0.022 ± 0.010 ^{aa}	1.454 ± 0.158 ^{aa}
L9	OQ372028.1	8.145 ± 1.218 ^{aa}	37.270 ± 36.851 ^{aa}	0.929 ± 0.500 ^{aa}	1.072 ± 0.116 ^{aa}	0.014 ± 0.005 ^{aa}	1.954 ± 0.434 ^{aa}
L10	OQ372029.1	6.989 ± 0.492 ^{aa}	13.293 ± 2.609 ^{aa}	0.357 ± 0.124 ^{aa}	0.402 ± 0.522 ^{aa}	0.018 ± 0.000 ^{aa}	1.870 ± 0.181 ^{aa}
	Mean	6.311 ± 1.824	24.364 ± 12.572	0.460 ± 0.216	0.810 ± 0.371	0.047 ± 0.024	1.309 ± 0.482
	CV	28.902%	51.601%	46.957%	45.802%	51.064%	36.822%

Data are expressed as mean ± SD of three independent accessions. CV: coefficient of variation; CV: coefficient of variation; n = 3; means followed by single letters in a column differ significantly at a 5% level of significance.

3.3.11 Phytochemical content and *in vitro* antioxidant activity

Ripe fruits of *Physalis* accessions were assessed for phytochemicals (total phenolic acid, tannic acid, and flavonoid) and antioxidant activities (Table 3.7 and Appendix 3C). The phytochemical contents and antioxidant activities did not change significantly ($p > 0.05$) among the *Physalis* accessions (Table 3.7). The TPC of different *Physalis* accessions did not change considerably, ranging from 0.024 ± 0.025 to 0.092 ± 0.053 mg GAE/g DW (Table 3.7). *Physalis* accessions L1 and L5 recorded significantly higher TTC compared with the other eight accessions (Table 3.7). The TFC of the different *Physalis* accessions did not change significantly and ranged from 0.058 ± 0.034 to 0.152 ± 0.089 mg Rutin/g DW. Two different chemical assays (DPPH and HRS) were performed to assess the antioxidant activity of *Physalis* accessions (Table 3.7 and Appendix 3 D). The DPPH and HRS values obtained for the antioxidant property did not show significant differences ($p > 0.05$) among the different *Physalis* accessions (Table 3.7). The DPPH and HRS among the different *Physalis* accessions ranged from 29.846 ± 13.537 to 97.344 ± 2.263% and 8.696 ± 7.609 to 64.131 ± 9.962%, respectively (Table 3.7).

Table 3.7. Phytochemical content and radical scavenging activities of *Physalis* accessions.

Sample ID	Accession Number	TPC (mg GAE/g DW)	TTC (mg Tannic acid/g DW)	TFC (mg Rutin/g DW)	DPPH RSA %	HRS activity %
L1	OQ507152.1	0.092 ± 0.053 ^{aa}	0.158 ± 0.004 ^{a}	0.145 ± 0.073 ^{aa}	29.846 ± 13.537 ^{a}	64.131 ± 9.962 ^{aa}
L2	OQ372021.1	0.059 ± 0.040 ^{aa}	0.126 ± 0.045 ^{aa}	0.072 ± 0.020 ^{aa}	94.095 ± 0.182 ^{aa}	52.174 ± 9.962 ^{aa}
L3	OQ372022.1	0.024 ± 0.025 ^{aa}	0.099 ± 0.039 ^{aa}	0.070 ± 0.017 ^{aa}	75.862 ± 2.970 ^{aa}	30.435 ± 18.827 ^{aa}
L4	OQ372023.1	0.035 ± 0.026 ^{aa}	0.115 ± 0.067 ^{aa}	0.063 ± 0.040 ^{aa}	44.868 ± 6.556 ^{a}	6.159 ± 7.863 ^{aa}
L5	OQ372024.1	0.034 ± 0.008 ^{aa}	0.184 ± 0.015 ^{a}	0.096 ± 0.026 ^{aa}	67.539 ± 17.427 ^{a}	41.667 ± 12.120 ^{aa}
L6	OQ372025.1	0.081 ± 0.051 ^{aa}	0.047 ± 0.022 ^{aa}	0.071 ± 0.051 ^{aa}	73.088 ± 9.318 ^{aa}	59.420 ± 44.952 ^{aa}
L7	OQ372026.1	0.060 ± 0.016 ^{aa}	0.049 ± 0.039 ^{aa}	0.058 ± 0.034 ^{aa}	95.045 ± 7.149 ^{aa}	53.623 ± 26.721 ^{aa}
L8	OQ372027.1	0.082 ± 0.011 ^{aa}	0.061 ± 0.023 ^{aa}	0.097 ± 0.069 ^{aa}	96.156 ± 3.924 ^{aa}	8.696 ± 7.609 ^{aa}
L9	OQ372028.1	0.060 ± 0.027 ^{aa}	0.041 ± 0.015 ^{aa}	0.128 ± 0.038 ^{aa}	97.344 ± 2.263 ^{aa}	62.319 ± 8.786 ^{aa}
L10	OQ372029.1	0.080 ± 0.071 ^{aa}	0.072 ± 0.016 ^{aa}	0.152 ± 0.089 ^{aa}	89.666 ± 16.692 ^{aa}	39.131 ± 20.738 ^{aa}
	Mean	0.061 ± 0.033	0.095 ± 0.029	0.095 ± 0.046	76.351 ± 8.002	41.776 ± 16.754
	CV	54.098%	30.526%	48.421%	10.481%	40.104%

Data are expressed as the mean ± SD of three independent accessions. CV: coefficient of variation; n = 3; TPC: total phenol content; TTC: total tannin content; TFC: total flavonoid content; DPPH RSA: 2, 2-diphenyl-2-picrylhydrazyl radical scavenging activity; HRSA: hydrogen peroxide radical scavenging activity. Means followed by a single letter in a column differ significantly at a 5% level of significance.

3.3.12 Correlation analysis between phytochemical contents and antioxidant activities

Correlation and regression studies were performed to assess the effect of each phytochemical content (phenolic acid, tannins, and flavonoids) on the radical scavenging activity of *Physalis* accessions (Table 3.8). DPPH radical scavenging was largely facilitated by the phenolic acid content, with a correction of r = 0.327, while flavonoids and tannins had a negative correlation (Table 3.8). Hydrogen peroxide radical scavenging was largely facilitated by phenolic acids and flavonoids at a correlation value of r = 0.3599 and 0.2877, while tannins had a negative correlation (Table 3.8). The regression analysis showed that all phytochemicals (phenolic acids, tannins, and flavonoids) had a linear relationship to the radical scavenging of DPPH and hydrogen peroxide radicals (Table 3.8).

Table 3.8. Correlation and regression analysis of phytochemical content and radical scavenging activity of *Physalis* accessions.

Polyphenol Content	**Correlation Coefficient (r) for DPPH RSA**	**Correlation Coefficient (r) for HRSA**	**ANOVA (*p* Value) for Hypothesis Testing of Slope of Regression Line for DPPH RSA**	**ANOVA (*p* Value) for Hypothesis Testing of Slope of Regression Line for HRSA**
Phenolics	0.327	0.3599	0.928	0.307
Tannins	-0.6316	-0.0374	0.050	0.918
Flavonoids	-0.1150	0.2877	0.752	0.420

3.4. Discussion

DNA barcoding is relatively fast in terms of species identification and discrimination (Chen *et al*., 2010). DNA barcodes such as *rbcL*, *mat*K, *psbA-trnH,* and ITS2 are very efficient in identifying unknown plant species, with *ITS2* being among the best DNA barcodes for species identification and discrimination (Zhao *et al*., 2018; Chen *et al*., 2010). Based on BLASTn and the phylogenetic analysis of ITS2 and *rbcL* gene sequences, the *Physalis* accessions used in the current study were all confirmed to belong to the genus *Physalis*.

The results of this study showed that the used barcode regions have different abilities of species discrimination and identification. ITS2 was proposed as a core barcode for seed plants by the Consortium for the Barcode of Life (CBOL) Plant Working Group (China Plant BOL Group, 2011). It was clear that the *Physalis* accessions used in the current study were identified as *P. purpurea* based on ITS2 barcode. The phylogenetic tree was able to discriminate the *Physalis* accessions and had reliable clades with a posterior probability of 80%. Species discrimination based on the *rbcL* gene was not possible as there was no formation of clades on the Bayesian inference phylogenetic tree. The high conservation of the *rbcL* barcode gene in *Physalis* accessions makes it a less ideal candidate for DNA barcoding when compared to other barcode genes such as ITS2 (Ralte and Singh, 2021). The ability of ITS2 to emerge as a better barcode than *rbcL* is clearly supported in other studies on *Physalis* and other plants (Feng *et al*., 2016; Zhao *et al*., 2018; Chen *et al*., 2010).

The *rbcL* barcode gene of *Physalis* accessions under study did not display genetic distance, diversity, or polymorphism. This is an indication that there were no genetic differences between the *Physalis* accessions used. Therefore, the *rbcL* barcode is highly conserved in some species of *Physalis,* which has also been reported in other studies (Xu *et al*., 2021; Lee *et al*., 2013). The ITS2 barcode gene exhibited genetic variation among the *Physalis* accessions with a notable nucleotide diversity of 0.27629. This has also been reported in the DNA barcoding and identification of Solanaceae plants (Ralte and Singh, 2021). The genetic distance observed among the *Physalis* accessions in this study based on ITS2 concurs with previous studies where the ITS2 barcode region showed high genetic variation due to its high mutation rate (Tan *et al*.,

2020). High intraspecies variation was observed based on the ITS2 sequences, supporting the successful identification of *Physalis* accessions using the ITS2 barcode region. The Tajima D value of *Physalis* accessions based on the ITS2 gene was 0.779171, indicating low-frequency mutations and balancing selection within the population (Korneliussen *et al*., 2013). Three barcode gaps were identified for the ITS2 gene sequences among *Physalis* accessions using the ABGD method. The presence of barcode gaps is crucial for species delimitation and serves as the basis for plant species identification and discrimination (Collins and Cruickshank, 2013).

Commercial interest in plants of the genus *Physalis* has been rising worldwide due to its nutritional value, edible fruits, and the current and potential medicinal uses (Feng *et al*., 2016). The *Physalis* accessions were investigated for mineral content and the analysis revealed that *P. purpurea* fruits are rich in potassium, sodium, calcium, and magnesium, which concurs with reports from other studies on the wild edible fruits of *Physalis* (Musinguzi *et al*., 2007). Based on these findings, this fruit can be used as an alternative to the daily intake of minerals, which are essential for human health. Potassium is important for the function of the cardiovascular system in humans (Sica, 2002). Potassium/sodium balance is fundamental for the transmission of electrical impulses in the heart (Kowey, 2002). Magnesium is an important mineral in protein synthesis, oxidative phosphorylation, the regulation of body temperature and muscle contractions, and it is a cofactor for many enzymes (Jahnen-Dechent and Ketteler, 2012). Calcium is also another important element required for blood clotting, growth, bone formation, cell metabolism, and heart function (Berchtold *et al*., 2000; Ross *et al*., 2012). The *Physalis* fruits also contain trace elements including zinc, iron, copper, nickel, lithium, and manganese. Low levels of iron and nickel in *Physalis* have also been reported in other studies (Erkaya *et al*., 2012; Musinguzi *et al*., 2007). Zinc is important in the catalytic activity of enzymes, cellular signaling, and facilitates the modification of the structures of DNA and RNA, proteins, and cellular membranes (Brown *et al*., 2001). Iron is an essential vitamin that is required in small amounts for DNA synthesis, oxygen transportation, and the electron transport chain (Abbaspour *et al*., 2014). Copper is also an essential element that is required in small amounts for facilitating lung elasticity, neurovascularization, the metabolism of iron, adequate growth, energy metabolism, reactive oxygen species detoxification, and cardiovascular integrity (Ruiz *et al*., 2021; National Research Council, 2000). The functions of nickel in animals are not well known

(Genchi *et al*., 2020). However, in plants and bacteria it is required as a cofactor for the enzymes involved in growth and germination in plants (Genchi *et al*., 2020). Lithium is an essential element in the physiological regulation of mood (Demling *et al*., 2001). Manganese is required as a cofactor for enzymes that perform cholesterol, carbohydrate, and protein metabolism (National Research Council, 2001). The presence of minerals in plant accessions can be linked to their medicinal properties as well (Okwu, 2005). Our analyses show the potential of *P. purpurea* as an excellent mineral supplement in nutraceuticals. Though little attention is given to this wild fruit, our findings indicate a richness of nutrients and its potential application as a nutritional supplement.

The *Physalis* accessions were rich in phytochemicals such as phenols, tannins, and flavonoids, which have also been identified in the genus in other previous studies (Sathyadevi and Subramanian, 2015). The levels of these different types of phytochemicals were different in the *Physalis* accessions, an indication of the existence of different cultivars within *P. purpurea*. Variation in phytochemical secondary metabolites in plants, such as phenolic acid and flavonoids, have been linked to environmental stress during growth and development (Pant *et al*., 2021). Ecologically limiting factors like lighting, carbon dioxide, soil salinity, temperature, and soil fertility can affect the biochemical and physiological responses of plants and their secondary metabolite production (Li *et al*., 2020). Abiotic stressors lead to fluctuations in the chemical constituents of plants, selectively altering the content of secondary metabolites such as phytochemicals (Pant *et al*., 2021). Soil salinity, an abiotic stressor for plants, has been shown to cause the accumulation of secondary metabolites such as flavonoids in plants as a response to nutritional imbalance, decreased photosynthesis, and the uptake of nutrients (Banerjee and Roychoudhury, 2017). Flavonoid accumulation when plants are under stress due to the increased salinity of soils provides a curative effect for affected plants (Gengmao *et al*., 2015). The tannin and flavonoid contents were the highest among the phytochemicals identified in the *Physalis* accessions studied. Previous studies have found that phenols have the highest phytochemical content, followed by tannins, with flavonoids having the lowest content (Kasali *et al*., 2021; El-Beltagi *et al*., 2019). The concentration of phenols might have varied in this study as compared to others due to the geographical variations, environmental/abiotic stressors, the method of extraction of phenols, the sugars present, and the carotenoid and ascorbic acid contents (Pant *et*

al., 2021; Aryal *et al*., 2019; Oszmiański *et al*., 2018). The presence of phenols has been associated with antioxidant properties and, therefore, the ability to scavenge for free reactive oxygen species is facilitated by phytochemicals (El-Beltagi *et al*., 2019; Sahoo *et al*., 2013). According to the radical scavenging activity assays conducted, the fruit extracts of *Physalis* accessions scavenged free radicals such as DPPH and hydrogen peroxide due to the presence of phytochemicals. Phenolic and flavonoid contents in plants are important for their antioxidant properties, which allow them to scavenge reactive free radicals by donating hydrogen atoms to the free radicals (Amarowicz *et al*., 2004).

Correlation studies of the polyphenol content and the ability of polyphenols to promote radical scavenging have shown a linear relationship between phenolic and flavonoid contents in relation to radical scavenging capacity (Shrestha and Dhillion, 2006). In this study, the correlation analysis showed a positive correlation between the phenolic content and the DPPH radical scavenging activity. However, a negative correlation was observed for the tannin and flavonoid contents ($r = -0.1150$) and the DPPH radical scavenging activity. This is an indication that the phenolic content of *P*. *purpurea* contributed more towards the DPPH radical scavenging activity than tannins and flavonoids. This concurs with similar studies on the role of phenolic content from the genus *Physalis* in DPPH radical scavenging (Karpagasundari and Kulothungan, 2014). Phenol and flavonoid contents showed a moderate correlation to the hydrogen peroxide scavenging capability. This concurs with other studies showing that phenol and flavonoid contents contribute towards hydrogen peroxide scavenging activity (Aryal *et al*., 2019). The presence of tannins in the *Physalis* accessions did not show any DPPH radical and hydrogen peroxide scavenging activities. However, tannic acid in other plant studies has been shown to have radical scavenging activity against DPPH radicals and hydrogen peroxide (Karpagasundari and Kulothungan, 2014; Gülçin *et al*., 2010).

3.5 Conclusions

The *Physalis* accessions used in the current study were identified as *P. purpurea* based on the ITS2 barcode region. The high genetic variation among the *Physalis* accessions based on the ITS2 sequences allowed for the clear identification of *Physalis* species as *P. purpurea*. There was no genetic variation among the *rbcL* sequences of *Physalis,* an indication that the gene is

relatively conserved. The study confirmed that the fruits of *P. purpurea* contained a high content of minerals, including calcium, sodium, magnesium, and potassium. The fruits were also rich in phenolic acids, tannins, and flavonoids, and exhibited antioxidant properties. The phenolic compounds and flavonoids were the major contributors to the radical scavenging activity of the *P. purpurea* fruits. Therefore, the underutilized *P. purpurea* can be used as an excellent source of antioxidants for the management of oxidative stress-induced human diseases.

CHAPTER FOUR

4.0 Molecular identification, genetic diversity and secondary structure predictions of *Physalis* species using ITS2 DNA barcoding

4.1 Introduction

The genus *Physalis* are well-known herbaceous plants which belong to the *Solanaceae* family. *Physalis* is native to the Andes of South America which includes countries such as Colombia, Peru and Ecuador (Simbaqueba *et al*., 2011). The genus has several species that grow in the wild with only a few species under cultivation as food crops and ornamentals (Khan and Bakht, 2015). *Physalis* plant was initially consumed largely by the Inca people but later introduced to Africa and India after the entry of Christopher Colombus into the Americas (Popenoe *et al*., 1990). *Physalis* species including *P. peruviana*, *P. alkengi*, *P. purpurea*, *P. pubescens* and *P. philadelphica* are widely cultivated worldwide for their nutritional, medicinal and economic value (Shenstone *et al*., 2020; Yen *et al*., 2010). These plants are rich in both nutrients (provitamin A, vitamins C, vitamin B complex, phosphorus and fiber) and phytochemicals such as phytosterols, polyphenols, saponins, peruvioses, fisalins and withanolides (Petkova *et al*. 2021; Golubkina *et al*., 2018). Economic interest in *Physalis* fruits has been steadily increasing over the years with countries in South East Asia and others such as Colombia, Egypt, Kenya, Zimbabwe and South Africa being major exporters (Novoa *et al*., 2006).

Precise identification of *Physalis* species in order to ensure safe utilization especially for medicinal applications - like in any other medicinal plants - is paramount to prevent improper use (Shinwari *et al*., 2018). It is also important for the maintenance and conservation of *Physalis* genetic resources since most natural *Physalis* species are being destroyed and cleared to pave the way for alternative land use as well as urbanization (Feng *et al*., 2016). This could lead to the loss/extinction of important species. The use of morphological identification of *Physalis* species is unreliable due to phenotypic similarities among the different species (Menzel, 1951). For example, *P. minima* are morphologically confused for *P. angulata* or *P. pubescens* due to similar phenotypic characters (Feng *et al*., 2016). Moreover, morphological characters are affected by environmental and developmental factors resulting in inaccurate identification of species

(Buckley *et al*., 1997). Molecular characterization using DNA markers such as simple sequence repeats (SSR) and random amplified microsatellites (RAM) have been used to comprehensively show genetic variability among genotypes (Muraguri *et al*., 2021; Chacón *et al*., 2016; Simbaqueba *et al*., 2011). However, these approaches do not provide sufficient discriminating capacity for classifying the *Physalis* genotypes into different species. A simple and accurate method is necessary for foolproof identification and determination of genetic relationships between the different cultivars of *Physalis* species.

DNA barcoding is a reliable tool of choice for studying genetic relationships between plants for species identification and delineation. It uses short and standardized DNA sequences that exhibit a sufficient level of variation to discriminate among species (Barcaccia *et al*., 2015; Hebert *et al*., 2003). The chloroplast-plastid region DNA sequences (such as *matK, rbcL, psbA-trnH, ycf*) and a nuclear internal transcriber spacer (ITS) region with less interspecific barcode gaps have been proposed as potential plant barcodes (Elansary *et al*., 2017). The ITS2 is a DNA spacer localized between the ribosomal 5.8S and 28S, in the chromosomal or corresponding polycistronic transcript region most commonly used for species discrimination studies (Lahaye *et al*., 2008). DNA barcode, ITS2 has been proposed as an efficient barcode in the identification of medicinal plants (Chen *et al*., 2010). It has been utilized in *Physalis* identification where it showed efficient species discrimination in comparison to chloroplast DNA barcodes due to its advantages such as small fragment length, good universality, high interspecific divergence and small intraspecific variation (Yao *et al*., 2010). A study done on the identification of *Physalis* species using ITS2 in China showed that the barcode is effective in species identification (Feng *et al*., 2016). ITS2 has also been used to identify *P. angulata* among *Solanaceae* plants (Ralte and Singh, 2021).

To date, there have been no reports on the identification of *Physalis* species both in the wild and those cultivated in Kenya. Molecular characterization using simple sequence repeat (SSR) markers made the assumption that only *P. peruviana* is present in Kenya (Muraguri *et al*., 2021). The present study was therefore aimed at identifying and clarifying the phylogenetic relationships of *Physalis* species in Kenya using the sequence and structural information of ITS2 barcode gene. The ITS2 barcode-anchored species delimitation would be useful for genetic resource conservation as well as augmenting future breeding programs.

4.2 Materials and methods

4.2.1 Plant material and sample collection

Leaves of *Physalis* plants were randomly collected from eight Counties in Kenya namely Nyamira, Kiambu, Nakuru, Kajiado, Nyeri, Homa Bay, Elgeyo-Marakwet and Kericho (Figure 4.1 and Table 4.1). The samples were collected from April to June, 2019. The eight counties were selected based on the presence of *Physalis* germplasm. Geographical coordinates of locations and the number of *Physalis* plants sampled are presented in Table 4.1. Leaves were sampled in triplicates from 68 *Physalis* accessions (Appendix 1 A) based on availability as all plants were collected from the wild except for the samples from Elgeyo-Marakwet County which were collected from a *Physalis* plant farmer. The collection of *Physalis* leaf samples and storage in preparation for genomic DNA extraction was done as described in section 3.2.1.

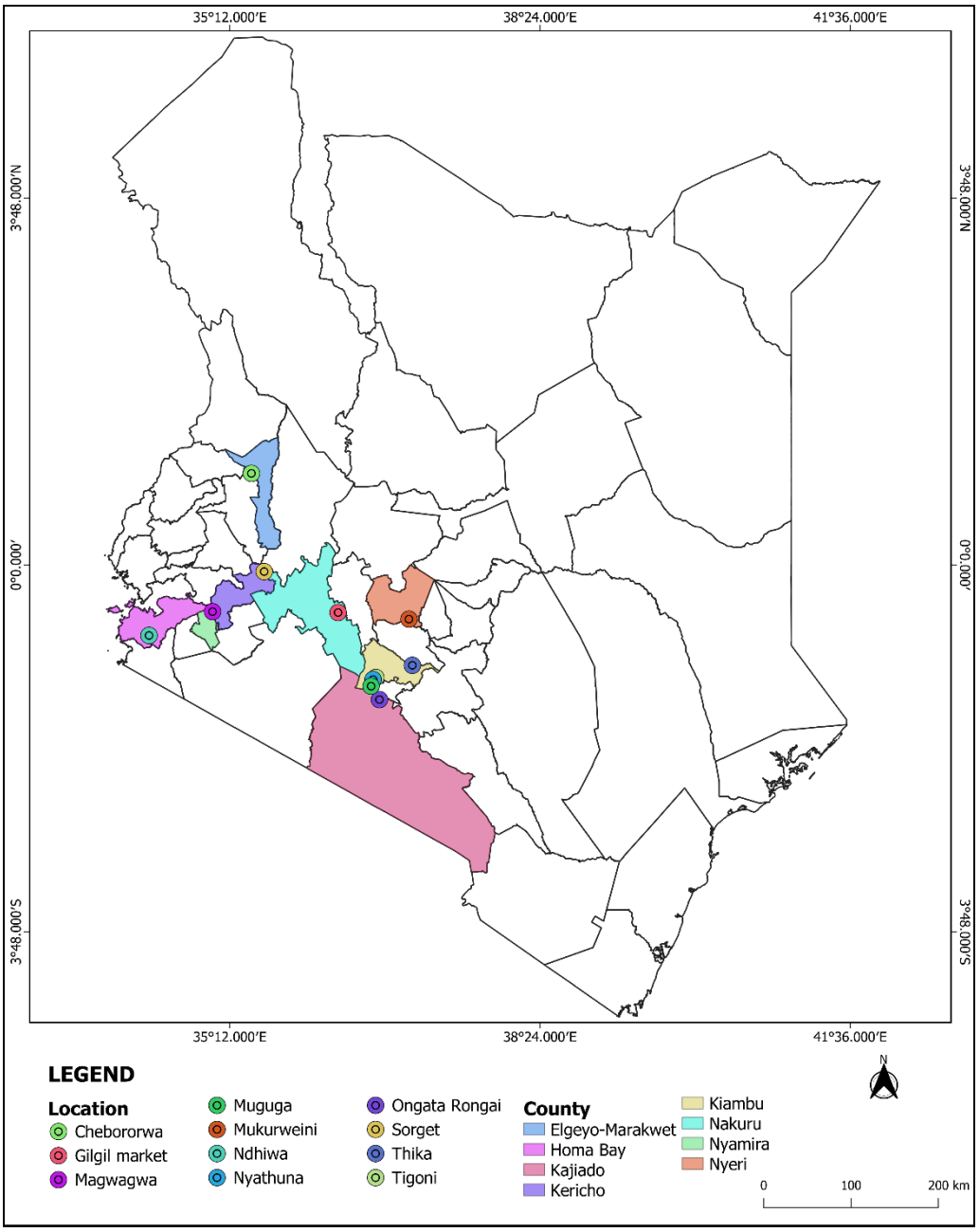


Figure 4.1. Locations from the eight counties in Kenya where *Physalis* plants were sampled.

Table 4.1. Geographical coordinates and number of *Physalis* samples collected from each of the selected eight Counties in Kenya

No.	Main Location (County)	Specific Location	Latitude	Longitudes	Number of leaf samples collected
1.	Kericho	Londiani, Sorget	0.0684° S	35.5548° E	10
2.	Elgeyo-Marakwet	Chebororwa	0.9487° N	35.4234° E	13
3.	Homa Bay	Ndhiwa	0.7299° S	34.3671° E	3
4.	Nyeri	Mukurweini	0.5609° S	37.0488° E	5
5.	Kajiado	Ongata Rongai	1.3939° S	36.7442° E	5
6.	Nakuru	Gilgil market	0.4923° S	36.3173° E	1
7.	Kiambu	Tigoni	1.1651° S	36.7065° E	17
		Thika	1.0388° S	37.0834° E	9
		Nyathuna	1.1859° S	36.6782° E	1
		Muguga	1.2551° S	36.6580° E	1
8.	Nyamira	Magwagwa	0.4830° S	35.0222° E	3

4.2.2 Molecular authentication of *Physalis* plants

4.2.2.1 Genomic DNA extraction

Physalis accessions genomic DNA extraction, purification, viewing and storage was carried out as described in section 3.2.2.1.

4.2.2.2 PCR amplification and sequencing

Polymerase chain reaction (PCR) amplification for internal transcribed spacer (ITS) of nuclear ribosomal DNA of *Physalis* accessions was performed using ITS2 primer. PCR amplification, purification of PCR products and sequencing was performed as described in section 3.2.2.2.

4.2.2.3 Sequence and Phylogenetic analysis

The ITS2 sequences achieved were manually curated using BioEdit version 7.0.5.3 software (Hall, 1999). Analysis of sequences in relation to curation of sequences, BLASTn analysis and preparation of multiple sequence alignments was performed as described in section 3.2.2.3. The two MSA prepared were for phylogenetic analysis and genetic diversity studies of *Physalis* based on ITS2 marker only. The first MSA was performed using the ITS2 sequences of 34 *Physalis* accessions from this study and seven *Physalis* ITS2 reference sequences (AY665886.1, MH763728.1, AY665905.1, AY665914.1, AY665903.1, AY665868.1 and MH763740.1) from NCBI database while the second was prepared using the ITS2 sequences of the 34 *Physalis* accessions only. The first MSA was used for phylogenetic analysis of *Physalis* accessions based on ITS2 marker while the second was utilized for the determination of genetic diversity, distance, polymorphism, neutrality and barcoding gap analysis of *Physalis* accessions based on ITS2 marker. A phylogenetic tree was prepared using the Bayesian inference method by MrBayes software version 3.2.7a (https://nbisweden.github.io/MrBayes/) as described in section 3.2.2.3 on 13th February 2023.

4.2.3 ITS2 secondary structure predictions

The RNA secondary structure predictions of the identified *Physalis* species were performed using the nucleotide sequences based on three species identified from the MrBayes phylogenetic tree, using rRNA database of *RNAfold* WebServer v2.4.18 (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) (Lorenz *et al*., 2011; Gruber *et al*., 2008; Mathews *et al*., 2004). The secondary structures of the representative *Physalis* accessions used in the current study were generated. RNA secondary structures were predicated based on the minimum free energy prediction method.

4.2.4 DNA divergence between and within *Physalis* accession populations

DNA divergence between and within *Physalis* species identified from the MrBayes phylogenetic tree based on ITS2 sequences was determined using the DnaSP software version 6.12.03 (Kartavtsev *et al*., 2011). The 34 ITS2 *Physalis* sequences were used to create a multiple sequence alignment (MSA). This MSA was uploaded and trimmed using Jalview software

version 2.11.2.0 to ensure uniform sequence lengths. The trimmed MSA was then used for DNA polymorphism analysis as described by Troshin *et al*. (2018, 2011). The edited MSA was then fed into the DnaSP software to assess divergence within and between *Physalis* species accessions.

DNA divergence between the three *Physalis* species identified was based on the analysis of nucleotide diversity (Pi), average nucleotide substitution per site between populations (Dxy) and number of nucleotide substitutions per site between populations (Da) as outlined by Jukes and Cantor algorithm on DnaSP. DNA divergence within each of the three *Physalis* species based on ITS2 gene was performed as described in section 3.2.3.

4.2.5 Determination of genetic distance between and within *Physalis* accessions

The interspecific, intraspecific and overall mean genetic distance of 34 *Physalis* accessions was calculated based on the ITS2 sequences using Kimura 2 parameter (K2P) with gamma distribution and a gamma parameter of 0.27 on the MEGA version 11 software (Kumar *et al*., 2018). The MSA used in the genetic diversity analysis was uploaded into the MEGA 11.0 software and grouping of ITS2 *Physalis* sequences into the 3 identified species was performed. Genetic distance was then calculated between and within grouped species based on K2P model.

4.2.6 Determination of DNA polymorphism

Sequence polymorphism of the ITS2 sequences was determined for all the 34 *Physalis* accessions through the use of DNA Sequence Polymorphism (DnaSP) software and analysis performed according to the parameters described in section 3.2.5.

4.2.7 Tajima's neutrality test

Tajima's neutrality test was performed using MEGA 11.0 software (Tamura *et al*., 2021). This test helped to determine the frequency of mutations and selection among the 34 *Physalis* accessions based on ITS2 sequences studied as described by Tajima *et al*., (1989). All ambiguous positions were removed for each sequence pair (pairwise deletion option).

4.2.8 DNA barcoding gap analysis

The ITS2 MSA for the 34 *Physalis* sequences were uploaded to ABGD website (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html accessed on 17 February 2023) and distance analysis performed based on K80 Kimura measure of distance described in section 3.2.7.

4.3 Results

4.3.1 Amplification and sequencing success rate

The success rate of PCR amplification calculated as the percentage of samples expected band size in three attempts was 75% (Table 4.2 and Appendix 1 B). Some of the images of the *Physalis* accessions amplicons are represented in Figure 3.2., 4.2. and 4.3. Sequence recovery success rate of the ITS region was 67%. The lengths of the ITS2 sequences generated from *Physalis* accessions were in the range of 237 - 707 bp, with an average of 523 bp. The mean GC content was 61%, with a range of 55.1 - 66.9% (Table 4.2). All the sequences generated from the amplification of ITS2 barcode were successfully deposited into the GenBank database (Table 4.3).

Table 4.2. Efficiency of PCR amplification and sequencing for *Physalis* accessions based on ITS2 DNA barcode region.

Barcode Region	Samples Tested (n)	Number of Amplicons Produced	Number of Sequences Produced	Amplification Efficiency (%)	Sequencing Efficiency (%)	Alignment Length (bp)	Mean Sequence Length (bp)	GC Content (%)
ITS2	68	51	34	75	67	707	523	61

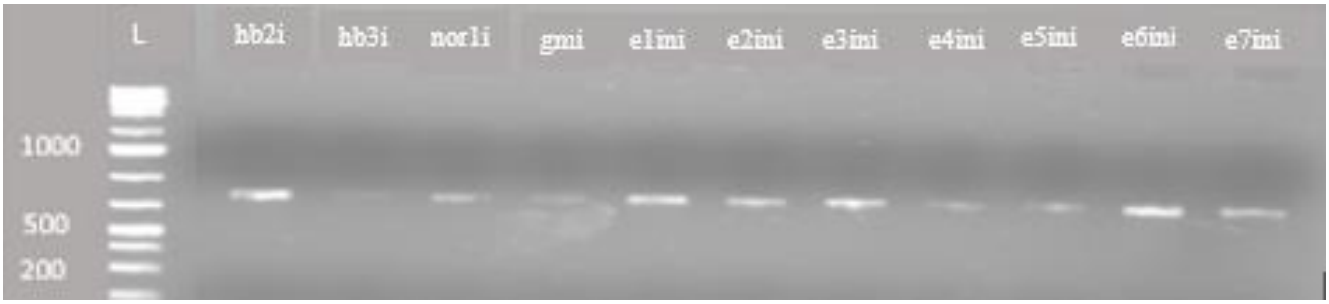


Figure 4.2. PCR amplicons for ITS2 sequences of *Physalis* accessions from Homabay (hb2i and hb3i) Kajiado (nor1i), Nakuru (gmi) and Elgeyo-Marakwet counties (e1-e7ini).

The letter L in Figure 4.2 above represents the DNA ladder, hb2i-e7ini represents the *Physalis* accessions amplicons based on ITS2 sequence amplification.

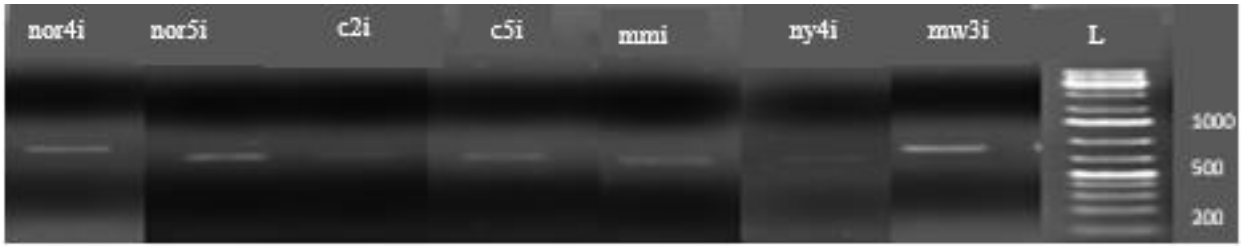


Figure 4.3. PCR amplicons for ITS2 sequences of *Physalis* accessions from Kajiado (nor4i and nor5i), Elgeyo-Marakwet (c2i and c5i), Baringo (mmi), Nyeri (ny4i) and Nyamira (mw3i) counties.

The letter L in Figure 4.3 above represents the DNA ladder, nor4i-mw3i represents the *Physalis* accessions amplicons based on ITS2 sequence amplification.

4.3.2 Identification using BLASTn analysis

The ITS2 sequence of each sample was used to perform BLASTn analysis independently in order to retrieve top hits available in the database and filtered by pairwise identity. The BLASTn analysis results for the 34 *Physalis* accessions revealed that all sequences generated were of the targeted loci. BLASTn analysis of ITS2 sequences from this study identified all the 34 (100%) *Physalis* accession as *Physalis* species. The highest similarity recorded for the 34 ITS2 sequences was 99.37% for *P. peruviana* (AY665914.1) (Table 4.3). Based on ITS2 sequences, 23, 4, 3 and 2 accessions were found to be related to *P. purpurea*, *P. cordata*, *P. peruviana* and *P. aff. Philadelphica*, respectively. Additionally, one sample each was found to be related to *P. minimaculata* and *P. microcarpa* (Table 4.3).

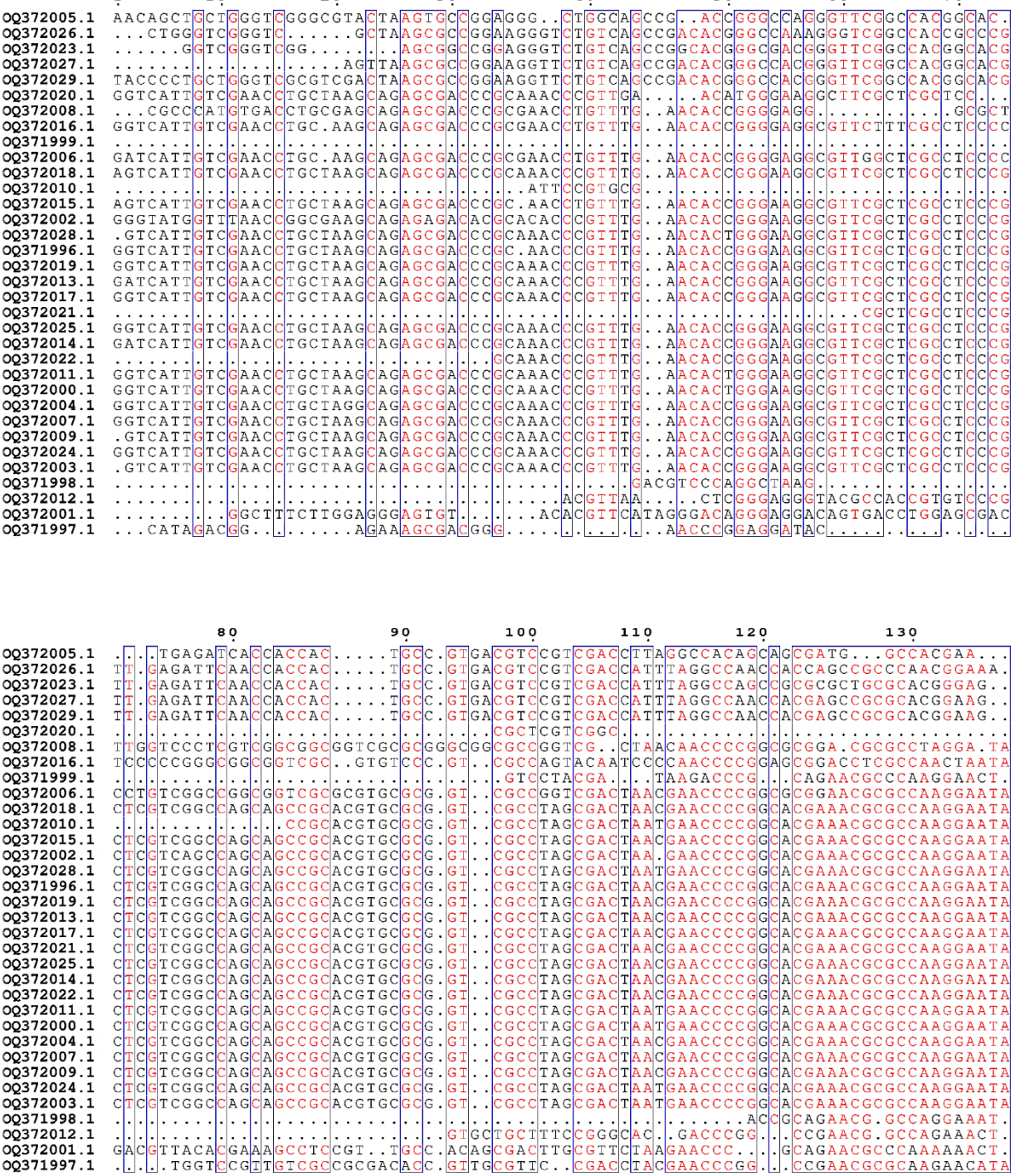
Table 4.3. BLASTn analysis results for the *Physalis* accessions based on ITS2 barcode gene.

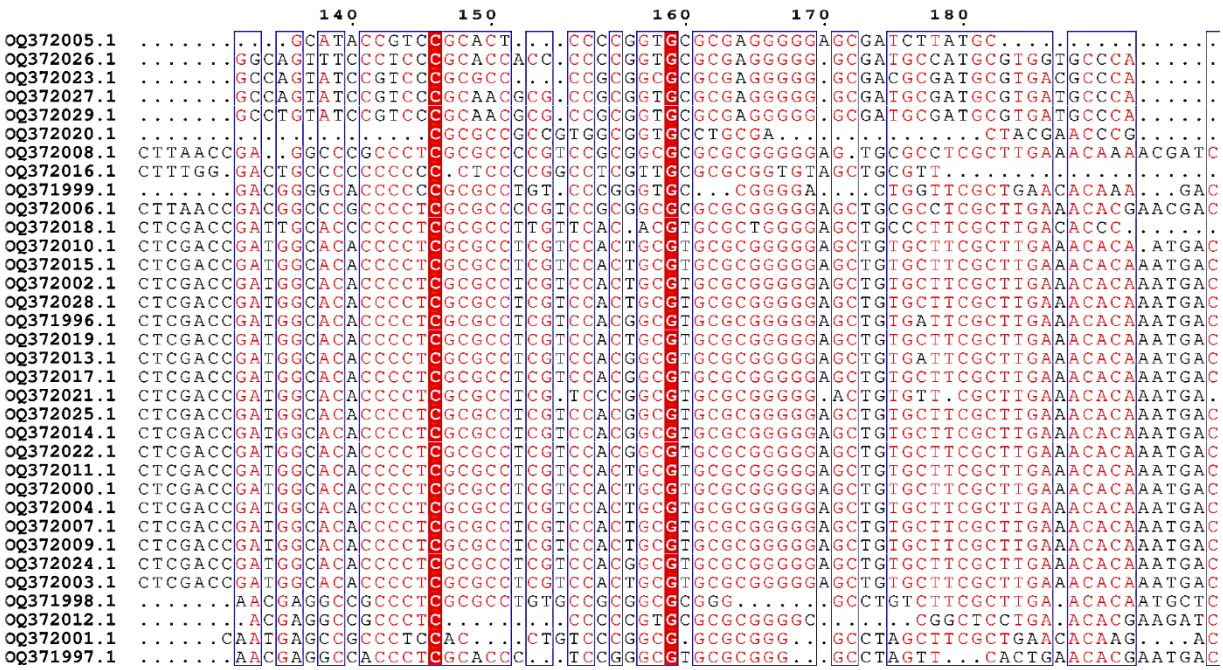
	Sample ID	Species of Best BLAST match	GenBank accession number (of database)	Max score	E value	Percent identity (%)	Percent coverage (%)	GenBank accession number
1	Chebororwa C2iF	*P. purpurea*	MH763740.1	972	0.0	94.70	91	OQ371996.1
2	Chebororwa e2iniF	*P. purpurea*	MH763740.1	418	4e-112	89.86	82	OQ371999.1
3	Chebororwa e3iniF	*P. purpurea*	MH763740.1	963	0.0	94.26	91	OQ372000.1
4	Chebororwa e5iniF	*P. purpurea*	MH763740.1	760	0.0	89.35	87	OQ372002.1
5	Chebororwa e6iniF	*P. purpurea*	MH763740.1	937	0.0	93.75	93	OQ372003.1
6	Chebororwa e7iniF	*P. purpurea*	MH763740.1	965	0.0	94.27	90	OQ372004.1
7	Gilgil gmi	*P. purpurea*	MH763740.1	161	5e-35	81.98	70	OQ372005.1
8	Muguga DMiF	*P. purpurea*	MH763740.1	880	0.0	93.83	97	OQ372007.1
9	Mukurweini ny5iF	*P. peruviana*	AY665914.1	355	4e-93	88.82	54	OQ372008.1
10	Ndhiwa hb2iF	*P. purpurea*	MH763740.1	965	0.0	94.28	92	OQ372009.1
11	Ongata Rongai nor1iF	*P. purpurea*	MH763740.1	966	0.0	94.28	92	OQ372011.1
12	Ongata Rongai nor4iF	*P. purpurea*	MH763740.1	955	0.0	94.09	92	OQ372013.1
13	Ongata Rongai nor5iF	*P. purpurea*	MH763740.1	959	0.0	94.51	90	OQ372014.1
14	Tigoni T2iF	*P. purpurea*	MH763740.1	577	7e-160	91.79	95	OQ372017.1
15	Tigoni T9iF	*P. purpurea*	MH763740.1	750	0.0	92.22	96	OQ372019.1
16	Tigoni T11iF	*P. purpurea*	MH763740.1	150	9e-32	84.85	62	OQ372020.1
17	Londiani 2i	*P. purpurea*	MH763740.1	686	0.0	92.98	100	OQ372021.1
18	Londiani 3i	*P. purpurea*	MH763740.1	896	0.0	94.06	97	OQ372022.1

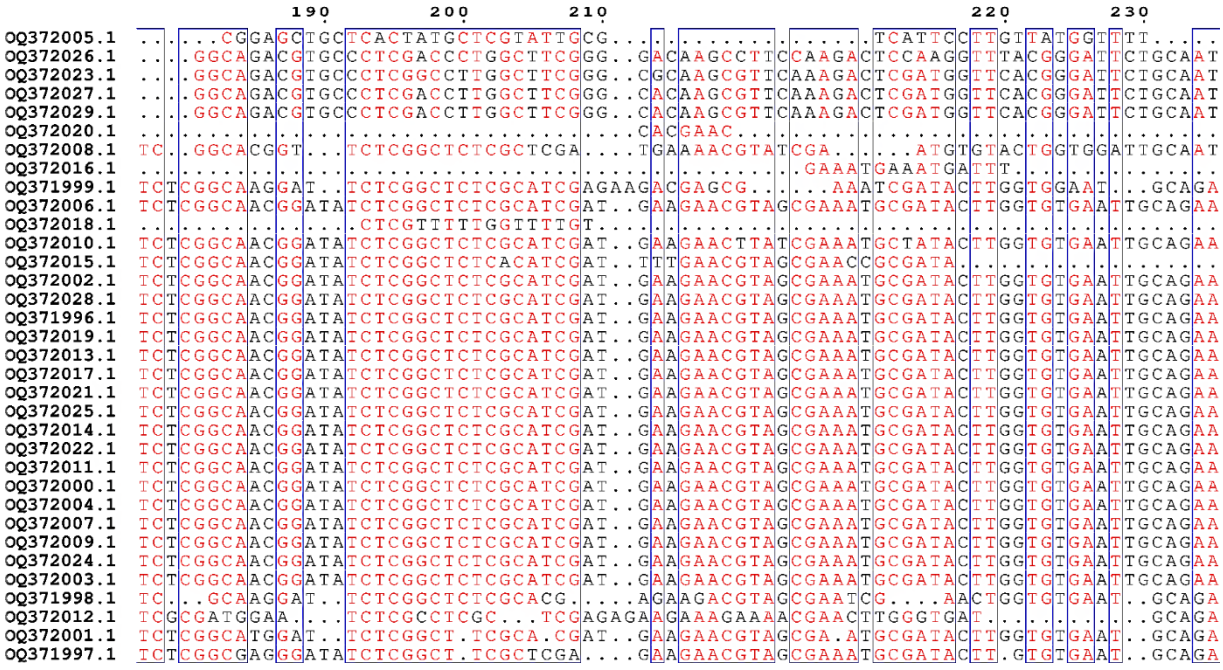
19	Londiani 5i	*P. purpurea*	MH763740.1	948	0.0	94.45	93	OQ372024.1
20	Londiani 7i	*P. purpurea*	MH763740.1	540	1e-148	86.00	98	OQ372026.1
21	Londiani 8i	*P. purpurea*	MH763740.1	885	0.0	93.07	94	OQ372027.1
22	Londiani 9i	*P. purpurea*	MH763740.1	758	0.0	88.96	100	OQ372028.1
23	Londiani 10i	*P. purpurea*	MH763740.1	747	0.0	90.16	99	OQ372029.1
24	Chebororwa C5iF	*P. cordata*	AY665886.1	451	5e-122	87.78	81	OQ371997.1
25	Chebororwa e1iniF	*P. cordata*	MH763728.1	316	1e-81	82.24	95	OQ371998.1
26	Chebororwa e4iniF	*P. cordata*	AY665886.1	407	1e-108	86.06	73	OQ372001.1
27	Ongata Rongai nor3iF	*P. cordata*	AY665886.1	300	2e-76	82.00	86	OQ372012.1
28	Thika TK8iF	*P. minimaculata*	AY665905.1	355	2e-93	88.51	94	OQ372015.1
29	Thika TK9iF	*P. peruviana*	AY665914.1	163	1e-35	80.36	80	OQ372016.1
30	Magwagwa mw3iF	*P. peruviana*	AY665914.1	1147	0.0	99.37	92	OQ372006.1
31	Londiani 4i	*P. peruviana*	AY665914.1	1044	0.0	97.41	91	OQ372023.1
32	Tigoni T4iF	*P.microcarpa*	AY665903.1	250	1e-61	86.78	81	OQ372018.1
33	Londiani 6i	*P. aff. philadelphica*	AY665868.1	425	2e-144	91.35	100	OQ372025.1
34	Nyathuna NKiF	*P. aff. philadelphica*	AY665868.1	374	7e-99	90.00	91	OQ372010.1

4.3.3 Multiple sequence alignment

Delimited ITS2 sequences were used for MSA preparation (Appendix 2B and Figure 4.4). The MSA of the curated 34 ITS2 *Physalis* sequences and their seven reference sequences retrieved from BLASTn analysis had a sequence length of 707 bp. The MSA was compressed using ESPript 3 (http://espript.ibcp.fr) and is indicated in Appendix 2B (https://espript.ibcp.fr/ESPript/temp/1101891838/0-0-1680467018-esp.pdf accessed on 17th February 2023). This alignment had a high rate of nucleotide substitutions among and between *Physalis* species studied. The substitution mutations entailed both transition and transversion point mutations. The MSA of only the 34 ITS2 *Physalis* sequences, prepared by MUSCLE, trimmed and viewed by Jalview had a sequence length of 533 bp. It was compressed using ESPript and is indicated in Figure 4.4 (https://espript.ibcp.fr/ESPript/temp/1035513530/0-0-1688384112-esp.pdf accessed on 17th February 2023). The latter MSA also showed substitution mutations of the transition type between species. For example, at position 116 of the alignment, most *P. purpurea* have an adenine nucleotide while the *P. cordata* and *P. peruviana* have a guanine nucleotide. The substitution at position 116 of the MSA indicates a transition mutation between species. On the other hand, at positions 130 and 138 of the MSA, all *P. cordata* have an adenine nucleotide while *P. purpurea* and *P. peruviana* species have a guanine at this position (Figure 4.4). This is an indication of transition mutations between species. An insertion mutation of 6 nucleotides is noted between nucleotide 89 and 90.







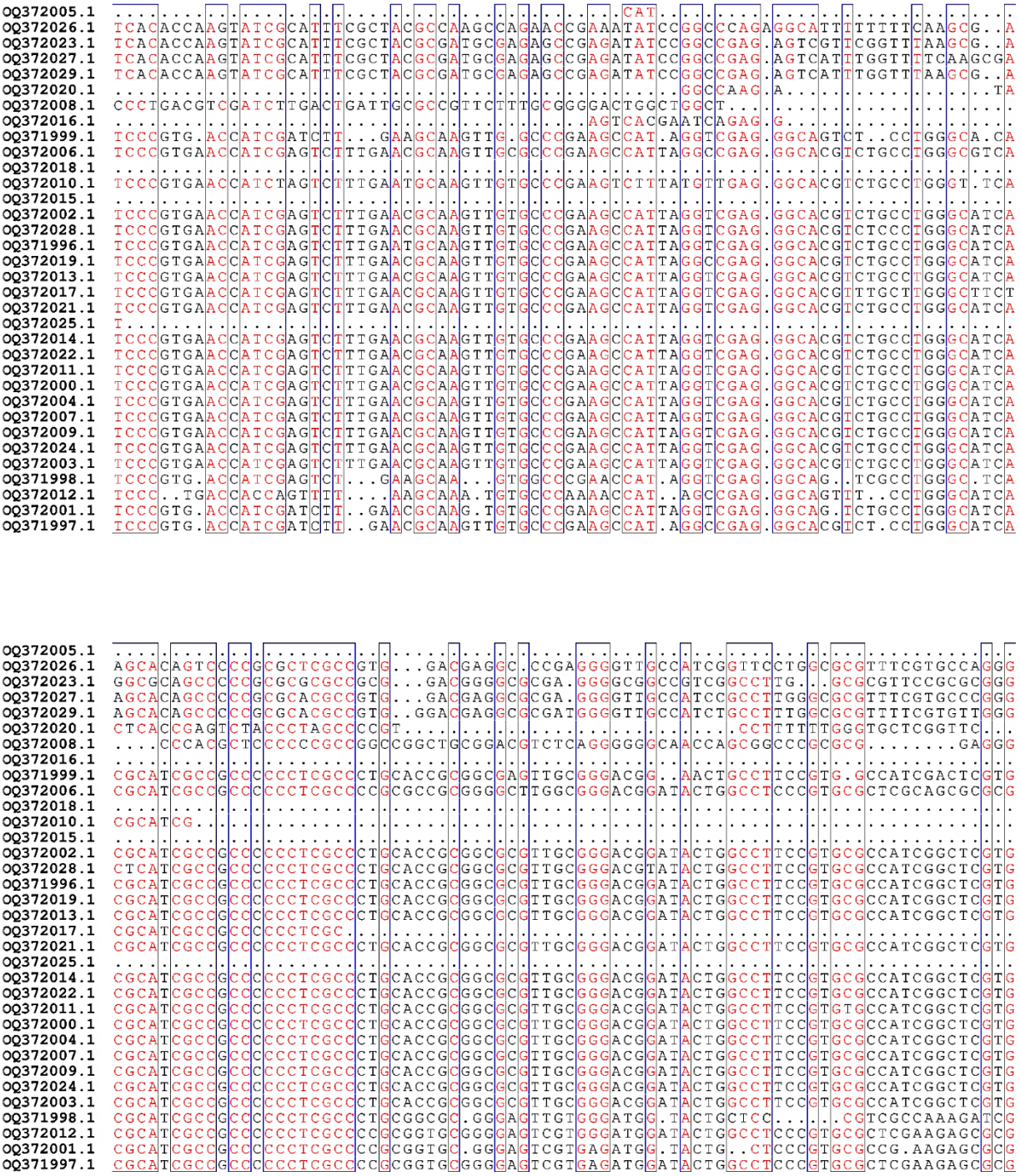


Figure 4.4. Multiple sequence alignment (MSA) for 34 *Physalis* accessions based on ITS2 sequences.

4.3.4 Phylogenetic-based identification

The phylogenetic tree based on ITS2 sequences for the different *Physalis* species assigned each of the *Physalis* accessions to its related species (Figure 4.5). The 3 *Physalis* species namely *P. cordata* (OQ5372012.1, OQ371998.1, OQ372001.1 and OQ371997.1), *P. peruviana* (OQ372016.1, OQ372008.1 and OQ372006.1) and *P. purpurea* (OQ371996.1, OQ371999.1, OQ372000.1, OQ372002.1 – OQ372005.1, OQ372007.1, OQ372009.1 - OQ2011.1, OQ372013.1 - OQ372015.1, OQ372017.1, OQ372018.1 - OQ372029) formed independent clades with > 80% branch support (Figure 4.5), indicating that *Physalis* species could be successfully discriminated using ITS2 sequences. The phylogenetic analysis indicates that the *Physalis* accessions showed species variation with different percentage indices (Figure 4.5).

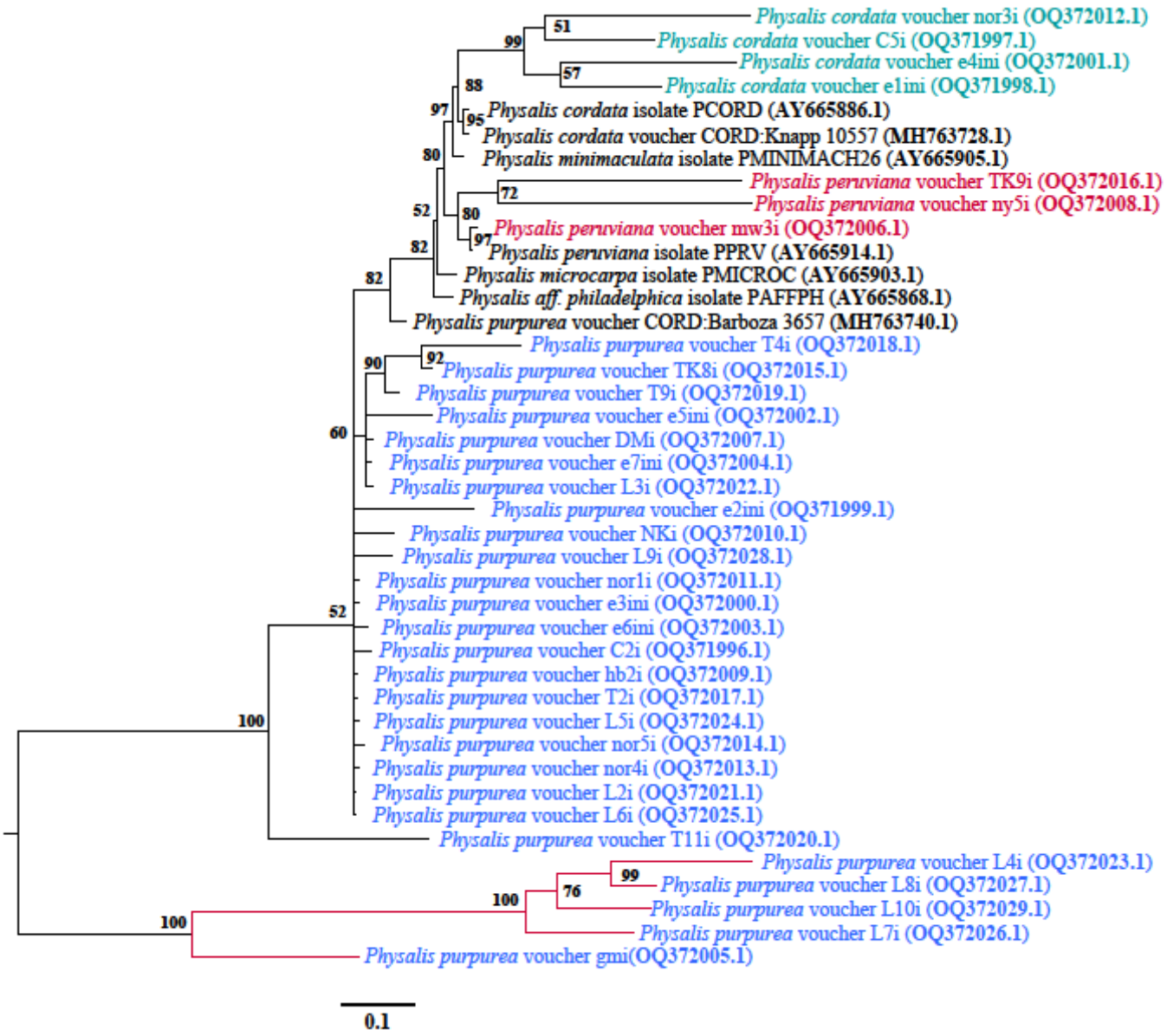


Figure 4.5. Phylogenetic tree derived from Bayesian inference analysis of the ITS2 gene of 34 *Physalis* sequences.

Plants from this study are presented in blue, red and green colors. Black color represents reference sequences for different *Physalis* species retrieved from GenBank after BLASTn analysis, red represents *P. peruviana*, blue represents *P. purpurea* and green represents *P. cordata*. Numbers above branches represent the percentage posterior probability statistic for the MrBayes phylogenetic tree.

4.3.5 ITS2 RNA secondary structures predictions

Besides the use of divergence of primary sequences of ITS2, variations in ITS2 secondary structures were also used to identify *Physalis* species. The ITS2 secondary structure predictions based on minimum free energy (MFE) is shown in Figure 4.6. The optimal secondary structure for *P. cordata* (OQ372001.1), *P. peruviana* (OQ372006.1) and *P. purpurea* (OQ371996.1) had minimum free energies of -204.90 kcal/mol, -312.90 kcal/mol and 266.90 kcal/mol with free energy of thermodynamic ensemble of -210.51 kcal/mol, -322.47 kcal/mol and -275.94 kcal/mol and the frequency of the MFE structure in the ensemble of 0.01%, 0.00% and 0.00%, respectively. The ensemble diversity for *P. cordata* (OQ372001.1), *P. peruviana* (OQ372006.1) and *P. purpurea* (OQ371996.1) was 136.00, 146.99 and 149.89, respectively.

The secondary structure of *P. cordata* (OQ372001.1 used as a representative) had 26 double helices, 26 loops and 2 single helices. The secondary structure of *P. peruviana* (OQ372006.1 used as a representative) had 39 double helices, 36 loops and 5 single helices. *Physalis* (OQ371996.1) was chosen as the representative accession for *P. purpurea* and the secondary structure had 43 double helices, 41 loops and 4 single helices. The secondary structure predictions showed variations among the 3 *Physalis* species. The predicted ITS2 secondary structures of the 3 *Physalis* species represented 3 different structures with a central ring and different helical orientations (Figure 4.6). The studied *Physalis* species showed a unique secondary structure that differed with the reference structure in two respects, the length of helices and the number of loops on their helices (Figure 4.6). The variation in helices length in the secondary structure of ITS2 was observed in different *Physalis* species (Figure 4.6). The loop number, position, size and angle from the centroid were distinguishable in all the three *Physalis* species. Besides the differences in the number of stems and rings, the shape and distribution of stem-loops in the secondary structure of the 3 *Physalis* species were notably different (Figure 4.6).

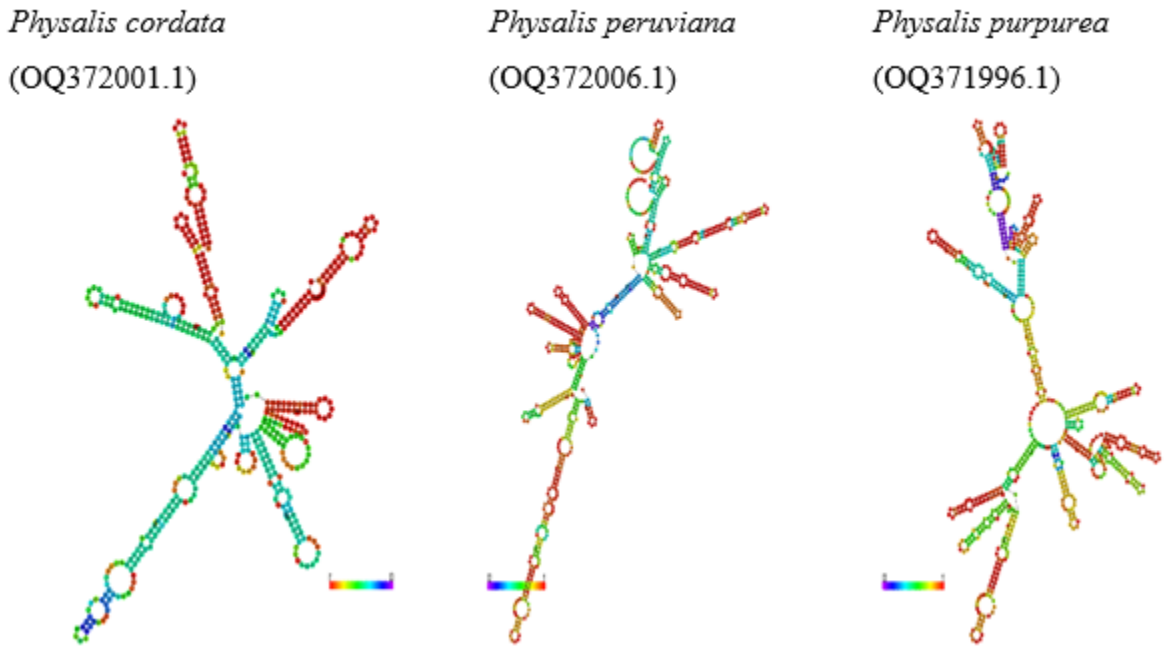


Figure 4.6. The predicted secondary structures of the ITS2 region of three *Physalis* species.

4.3.6 Genetic divergence analysis

4.3.6.1 DNA divergence between populations based on ITS2 sequences

Divergence between *Physalis* accessions in this study was determined by calculating the nucleotide diversity, average nucleotide substitutions per site between populations and number of net nucleotide substitutions per site between populations (Table 4.4). There were varying shared mutations among the *Physalis* accessions. The highest number of shared mutations at 11 was observed between *P. peruviana* and *P. cordata* populations. The lowest shared mutations at 3 were observed between *P. peruviana* and *P. purpurea*. The highest nucleotide diversity (0.36923) was obtained between *P. peruviana* and *P. cordata* while the lowest nucleotide diversity (0.15062) was recorded between *P. peruviana* and *P. purpurea*. There were no fixed differences between *P. peruviana* and *P. purpurea* populations as well as between *P. purpurea* and *P. cordata*. Two fixed differences were recorded between *P. peruviana* and *P. cordata* populations, which had the highest nucleotide diversity.

Table 4.4. DNA divergence between (interspecific) *Physalis* species populations based on ITS2 sequences

Population	*P. peruviana* (P1)	*P. cordata* (P2)	*P. peruviana* (P1)	*P. purpurea* (P2)	*P. purpurea* (P1)	*P. cordata* (P2)
Polymorphic sites in each population	37	33	7	16	17	8
Total number of polymorphic sites	49		17		18	
Average number of nucleotide differences	24.000		3.464		3.796	
Nucleotide diversity Pi (t)	0.36923		0.15062		0.16503	
Number of fixed differences	2		0		0	
Mutations polymorphic in population 1 (P1) but monomorphic in population 2 (P2)	32		4		17	
Mutations polymorphic in P2 but monomorphic in P1	28		18		3	
Shared mutations	11		3		6	
Average number of nucleotide differences between populations	25.750		4.284		4.185	
Average nucleotide substitution per site between populations (Dxy)	0.39615		0.18626		0.18196	
Number of net nucleotide substitutions per site between populations (Da)	0.04359		0.01383		0.00449	

4.3.6.2 DNA divergence within populations based on ITS2 sequences

DNA divergence within each population of the identified *Physalis* species was determined based on ITS2 sequences (Table 4.5). The highest nucleotide diversity was recorded within the *P. peruviana* population at 0.26324 despite this population having moderate number of polymorphic segregating sites at 80 and moderate number of nucleotide substitutions at 89 (Table 4.5). The lowest nucleotide diversity was recorded for *P. purpurea* population at 0.15883 and this corresponded to the lowest number of polymorphic segregating sites at 37 and lowest number of substitutions at 37. Moderate nucleotide diversity was recorded for *P. cordata* population at 0.17167 and this population had the highest number of polymorphic segregating sites (89) and number of nucleotide substitutions (102).

Table 4.5. DNA divergence within (intraspecific) *Physalis* species populations based on ITS2 sequences

Physalis species	*P. peruviana*	*P. cordata*	*P. purpurea*
Total number of sequences	3	4	27
Number of polymorphic (segregating) sites (S)	80	89	37
Nucleotide diversity Pi (Total)	0.26324	0.17167	0.15883
Nucleotide diversity Pi (JC-Total)	0.32423	0.19494	0.17848
Theta (Total)	0.27726	0.18545	0.26665
Total number of substitutions	89	102	37

4.3.7 Genetic distance between and within *Physalis* species based on ITS2 sequences

The overall average genetic distance among all *Physalis* accessions studied was determined to be 0.51 ± 0.04. The highest genetic distance (0.703) was obtained between *P. cordata* and *P. peruviana* and the lowest genetic distance (0.050) was between *P. purpurea* and *P. peruviana* (Table 4.6).

Table 4.6. Mean genetic distance between (interspecific) *Physalis* species based on ITS2 sequences

Groups	*P. purpurea*	*P. peruviana*	*P. cordata*
P. purpurea		0.050	0.057
P. peruviana	0.571		0.071
P. cordata	0.633	0.703	

The average intra-specific distance within each *Physalis* population was determined based on ITS2 sequences. The highest mean intraspecific distance (0.46±0.05) was recorded within the *P. peruviana* with no significant difference with the intra-specific distance of *P. purpurea* (0.43±0.05). The lowest mean intra-specific distance was within the *P. cordata* population at 0.28±0.03.

4.3.8 Nucleotide polymorphism and neutrality tests

Eight polymorphic sites were identified for the ITS2 gene sequences of the 34 *Physalis* accessions used in this study (Table 4.7). The nucleotide diversity (Pi) of the gene sequences was 0.14810 (Table 4.7). The eight polymorphic sites had one singleton and seven parsimony informative site mutations (Table 4.7).

Table 4.7. DNA polymorphism of *Physalis* accessions based on ITS2 marker

Polymorphic sites/ Segregation sites (S)	8	Position in the gene	Variants
Singleton	1	188	4
Parsimony informative sites	7	169	2
		171	2
		172	2
		183	2
		186	2
		189	2
		170	3
Nucleotide diversity (Pi)	0.14810		
Average number of nucleotide differences (k)	1.777		
Sequence length (base pairs)	533		
Number of sequences	34		

Tajima's neutrality test was performed on all 34 ITS2 sequences of *Physalis* accessions to establish selection of the species based on the Tajima D value and the nucleotide diversity. The number of segregating sites (S) and nucleotide diversity (π) was 464 and 0.155388, respectively. The Tajima D value obtained was -1.034267.

4.3.9 Genetic differences and barcoding gap analysis

Automatic barcode gap discovery (ABGD) results generated by the K80 Kimura measure of distance (K2P) based on the ITS2 gene sequences for *Physalis* accessions were used to assess the presence of a barcoding gap. Based on the ITS2 gene sequences, all pairwise distances were ranked by increasing the distance values from 0.00 to 1.36 and absence of barcoding gaps was detected (Figure 4.7).

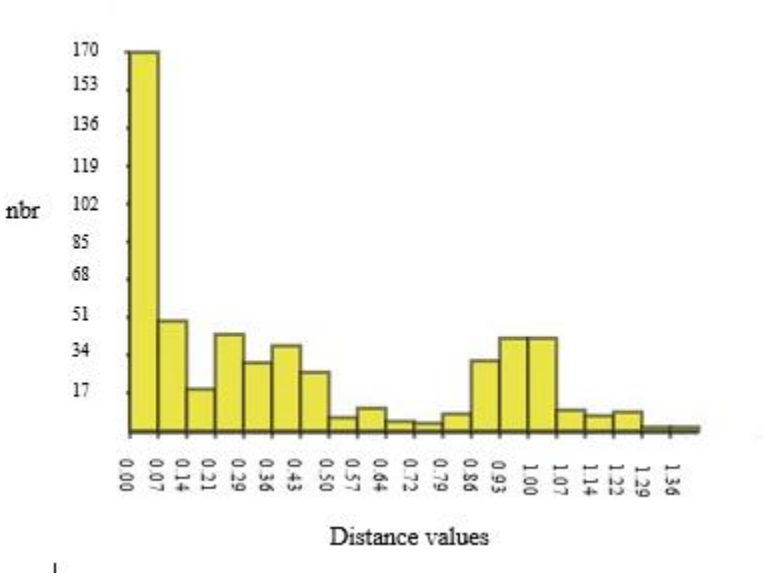


Figure 4.7. A histogram indicating the hypothetical distribution of pairwise differences of ITS2 gene sequences for 34 *Physalis* accessions.

Low divergence is presumably intraspecific divergence, whereas higher divergence indicates interspecific divergence. The abbreviation nbr on the *y*-axis of the histogram stands for the number of pairwise comparisons. No barcode gaps were detected.

4.4 Discussion

Currently, there is a growing demand for *Physalis* species worldwide due to their nutritional and medicinal value (Shenstone *et al*., 2020). *Physalis* species are widely diverse as there are more than 85 species of *Physalis* distributed throughout the world (Vargas *et al*., 2001). Different *Physalis* species have different applications and therefore accurately identifying the *Physalis* plants using molecular genetics will enhance the quick and precise identification of species for utilization, genetic resource conservation and development of genetic breeding programs (Shenstone *et al*., 2020; Feng *et al*., 2016). Morphological identification is not reliable due to high phenotypic similarities among *Physalis* species (Whitson and Manos, 2005). It is therefore important to use molecular identification tools such as DNA barcoding in the identification of *Physalis* (Feng *et al*., 2014). This study focused on the use of the DNA ITS2 region as a recognition tag to identify and discriminate *Physalis* species in Kenya.

The results from sequence characteristics, genetic distance, phylogenetic relationships and secondary structure analyses showed the remarkable potential of ITS sequences for distinguishing *Physalis* species. The potential to discriminate at the species level and ease of amplification makes ITS a favourable locus for the barcoding of plants (Pang *et al*., 2011; Li *et al*., 2010). Several researchers have demonstrated the usefulness of the nuclear ITS region as a DNA barcode for determining high levels of inter-specific genetic divergence among land plants (Mishra *et al*., 2017; Zhang *et al*., 2016). One of the most important criteria required for a successful DNA barcode is the presence of a high rate of PCR amplification and good quality sequences among the different target species. In this study, the use of the ITS2 barcode locus was successful as the PCR amplification and direct sequencing success rates were 75% and 67%, respectively. This compares to other reported studies whereby the amplification of the ITS2 gene gives a higher success rate than the sequencing (Kang *et al*., 2017; Zhang *et al*., 2015). Contrary to the present study, Feng *et al*. (2016) reported a 100% PCR amplification and sequencing success rate of ITS region in *Physalis* species. The amplification and sequencing success rate of ITS2 has been shown in other studies to be much lower compared to other barcoding genes such as *rbcL*, *matK* and *psbA-trnH* (Kang *et al*., 2017; Huang *et al*., 2015).

The ITS2 phylogenetic tree based on Bayesian inference revealed that the *Physalis* accessions were closely clustered together with the sequences of *P. cordata*, *P. peruviana* and *P. purpurea* chosen from the BLASTn search. The resultant phylogenetic tree identified and discriminated the *Physalis* accessions into three species namely *P. cordata*, *P. peruviana* and *P. purpurea*. Studies have shown that plant species identification based on BLASTn and phylogenetic tree analysis are reliable as long as the species under study have a reference dataset in the GenBank (Ross *et al*., 2008). This is the first report indicating the identification of *P. purpurea* and *P. cordata* from *Physalis* accessions in Kenya. This is an indication that the ITS2 region can accurately discriminate *Physalis* species and also identify new species that had not been previously reported in a particular region. ITS2 was able to effectively distinguish *Physalis* species and facilitate the identification of three *Physalis* species among the accessions studied. This supports previous studies that the ITS2 barcode has a higher discriminatory ability among species than other barcodes such as *rbcL* which tend to be highly conserved in different plant species (Li *et al*., 2017). In other studies, ITS2 has been identified as a promising DNA barcode due to its fast substitution rate and hence variability that provides satisfactory resolving power for closely related species (Zhang *et al*., 2015).

Although ITS2 nucleotide sequences are substituted at a fast rate, their secondary structures are largely conserved and can also be utilized in the identification of species (Feng *et al*., 2016, Zhang *et al*., 2015). The ITS2 sequence usually functions as a secondary structure in cells. In addition, the secondary structure contains genetic information that can be used for the classification and identification of plant species (Chen *et al*., 2012). In this study, we explored the differences in ITS2 secondary structures between three species of the genus *Physalis*. The ITS2 secondary structure inherently gave a visual distinctiveness between the three *Physalis* species. In our study, the prediction of the secondary structures in the three *Physalis* species revealed diverse secondary structures with distinguishable loop numbers, positions and elevations from the centroid. The ITS2 secondary structures revealed the uniqueness of the generated DNA barcode sequences. These unique genetic structures at the conserved nuclear region of *Physalis* species can be used to develop species-specific markers for the identification of *Physalis* accessions.

The genetic diversity of *Physalis* accessions studied largely concurred with the phylogenetic and genetic distance analysis. Both genetic diversity and genetic distance indicated the relatedness of species and in this study, *P. purpurea* and *P. peruviana* seem to be closely related while *P. peruviana* and *P. cordata* seem to be more diverse. Genetic diversity within (intraspecific divergence) species is assessed to determine the diversity of a group of organisms within a species. The high nucleotide diversity within *P. peruviana* is an indication that genetic change is much higher in this species than other species identified in this study. Intraspecific variation is important for evolution, adaptation and competition seen within the same species that can over time lead to genetically distinct members of a population adapted to specified environmental conditions such as cultivars (O'Dell and Rajakaruna, 2011). This information concurs with a previous study that showed high genetic divergence within *P. peruviana* populations and even when compared to other related taxa (Garzón-Martínez *et al*., 2015). The ABDG analysis was able to show the general intraspecific divergence within the *Physalis* accessions in this study as well. The maximum intraspecific distance P_{max} based on the ITS2 gene was 0.1 an indication that there is divergence within *Physalis* species. Intraspecific divergence is important as it creates variation among a species and allows for better conservation of the species in different environments.

The nucleotide polymorphism showed that the 34 *Physalis* accessions had eight polymorphic sites, with one singleton and seven parsimony informative sites. Higher polymorphic sites are consistent with high genetic diversity, an indication that the gene analyzed was effective in discrimination of *Physalis* species. The Tajima D value for the 34 *Physalis* accessions in this study gave a negative value of -1.034267, an indication that the population had a negative selection and higher low frequency mutations that can help in the differentiation of the *Physalis* species (Carlson *et al*., 2005). This study shows that ITS2 is an ideal candidate barcode for discrimination of *Physalis* species.

4.5 Conclusion

In the present study, ITS2 was found to possess a sufficient variable region between the different species and accessions for the determination of genetic divergence with high discriminatory ability. DNA barcode ITS2 was highly efficient in the identification and discrimination of

Physalis species. ITS2 was found to possess a sufficiently variable region between the different accessions and species for the determination of genetic divergence with high discriminatory ability. The *Physalis* accessions were identified and discriminated into three species namely *P. peruviana*, *P. purpurea* and *P. cordata*. Our study revealed significant variations in ITS2 secondary structure predictions that enhanced discrimination among the three identified *Physalis* species. This study provides insight into the scope of the development of species-specific primers for the discrimination of the three *Physalis* species. Understanding the RNA molecules would be useful for the selection of parents in trait-specific breeding strategies for *Physalis* improvement.

CHAPTER FIVE

5.0 Molecular identification and genetic diversity assessment of *Physalis* accessions using ribulose-1,5-bisphosphate carboxylase large (*rbcL*) DNA barcode

5.1 Introduction

Physalis genus is one of the key genera in the plant family of *Solanaceae* (Nightshade family) (Martínez, 1998). Various species of Physalis have been identified both morphologically and using molecular genetic analysis. *Physalis* species are found all around the world in tropical and temperate regions growing mostly as wild plants with a few countries such as Colombia taking up farming of *Physalis* for economical use (Chacón *et al*., 2016). Few species of *Physalis* are native to countries in the regions of Americas and others have been introduced in Eurasia and Southeast Asia (Vdovenko *et al*., 2021; Feng *et al*., 2020). In China five species of *Physalis* have been identified and these include; *P. minima, P. alkekengi, P. peruviana, P. pubescens* and *P. angulata* (Chinese Academy of Sciences, 1978). Two variants of *Physalis* have also been identified in China and they include; *Physalis angulata var. villosa* and *P. alkekengi var. franchetii* (Chinese Academy of Sciences, 1978). *P. peruviana* L. is native to the Andes of South America (Cailes, 1952). In the United States the *Physalis* species *Physalis grisea* is usually cultivated while in Mexico the two *Physalis* species *P. angulata* L. and *P. philadelphica* Lam are cultivated (Vargas-Ponce at al., 2016). *Physalis* species grow in areas from sea level to elevations of about 3,000 metres, they can also be found in a wide array of different climates such as high humidity areas to deserts as well (Vargas-Ponce *et al*., 2016; Toledo, 2013). The three species of *Physalis* namely *P. philadelphica, P. pubescens* and *P. peruviana* are widely cultivated in different parts of the world, mostly for their edible fruits (Sang-Ngern *et al*., 2016).

Physalis is a plant with high content of antioxidants, minerals and vitamins hence possessing medicinal value. The fruits of some *Physalis* species are edible and have high nutritive and antioxidant value (Shenstone *et al*., 2020; Puente *et al*., 2011). *Physalis* species of pharmacological interest such as *P. cordata, P. minima, P. pubescens* and *Physalis pruinosa* have been shown to produce compounds such as flavonoids, physalins, withanolides and saponin with medicinal value (Shah and Bora, 2019; Lima *et al*., 2014). The high number of diverse

metabolites produced by *Physalis* species could be an indication of their variability at the genetic level among the different species (Sandoval-Padilla *et al*., 2022). Morphologically *Physalis* species appear similar and can be easily misidentified (Feng *et al*., 2020). *Physalis* species are also morphologically similar to other plants of the Solanaceae family such as *N. physalodes* (Feng *et al*., 2016). Traditional morphological identification has greatly been hampered by heritable variations in *Physalis* species as well as differences in growth environments (Vargas-Ponce *et al*., 2011). The biologically active compounds with medicinal properties differ among the various species of *Physalis* that morphological look similar (Feng *et al*., 2018). In order to prevent improper use of *Physalis* in medicinal application it is crucial that reliable identification of *Physalis* species is done.

DNA barcoding has proven reliable in the rapid and accurate identification of plant species using short nuclear and chloroplast genes of plants to assess sequence diversity (Feng *et al*., 2015). Widely used DNA barcodes in the identification of plant species are the chloroplast DNA sequences including ribulose bisphosphate carboxylase large (*rbcL*), maturase K (*matK*), and *psbA-trnH* (Feng *et al*., 2018). Specifically, *rbcL* has been shown to have a higher distinguishing potential of plant species than *matK* or a combination of the two genes in previous studies (Tran *et al*., 2021). Among the several plant barcodes that have been identified, *rbcL* is proposed as one of the ideal plant barcoding loci by the Consortium for Barcode of life (CBOL) (Tran *et al*., 2021). It is however important to note that the ideal loci for plant barcoding will differ depending on whether they are used to identify taxonomic groups or for species discrimination (Tran *et al*., 2021). It has been noted that there is more species variation of the chloroplast DNA barcode *matK* as compared to *rbcL* which seems to be more conserved (Thitikornpong *et al*., 2018). Another study showed that DNA barcodes *matK* and *psbA-trnH* provided more rates of variable sites than *rbcL* which possessed more conservation sites (Li *et al*., 2017). Other studies, however, *rbcL* showed better species resolution of various plants studied as compared to *matK* (Fatima *et al*., 2019). This is a clear indication that the efficiency of a DNA barcode in species resolution can vary in different species of plants.

The objective of this study was to evaluate the efficiency of *rbcL* DNA barcode in the species discrimination of *Physalis* plants collected in various regions of Kenya. This is important as it

will promote genetic conservation and increase in breeding programs of *Physalis* which is considered an orphan plant and grows mostly in the wild. This could also open up a channel for the value addition, classification, protection and conservation of *Physalis* species and accessions in Kenya.

5.2 Materials and methods

5.2.1 Plant material

Leaves and fruits of *Physalis* plants were randomly picked from Baringo, Kericho, Nairobi, Elgeyo Marakwet, Nyeri, Kajiado, Homabay, Kiambu, Nakuru and Machakos counties of Kenya from April 2019 to January 2021 as indicated in Figure 5.1. The locations from where samples were collected including their geographical coordinates are indicated in Table 5.1 and Appendix 1 A. The collected *Physalis* plant samples were placed in zip lock bags and stored in a cool box, transported to the laboratory and identified by the taxonomist Mr. Patrick Mutiso and part of the samples were preserved in the University of Nairobi herbarium in the Department of Biology (Code of Voucher Specimen: KP/UON2019/001). Seventy-eight plants as indicated in Appendix 1 A were sampled for their leaves and fruits based on phenotypic characteristic and availability from the ten counties in Kenya. All *Physalis* plants collected were wild self-propagating plants except those from Elgeyo-Marakwet that were collected from a gooseberry farmer. The leaves were utilized in the extraction of DNA for molecular identification and *Physalis* species discrimination.

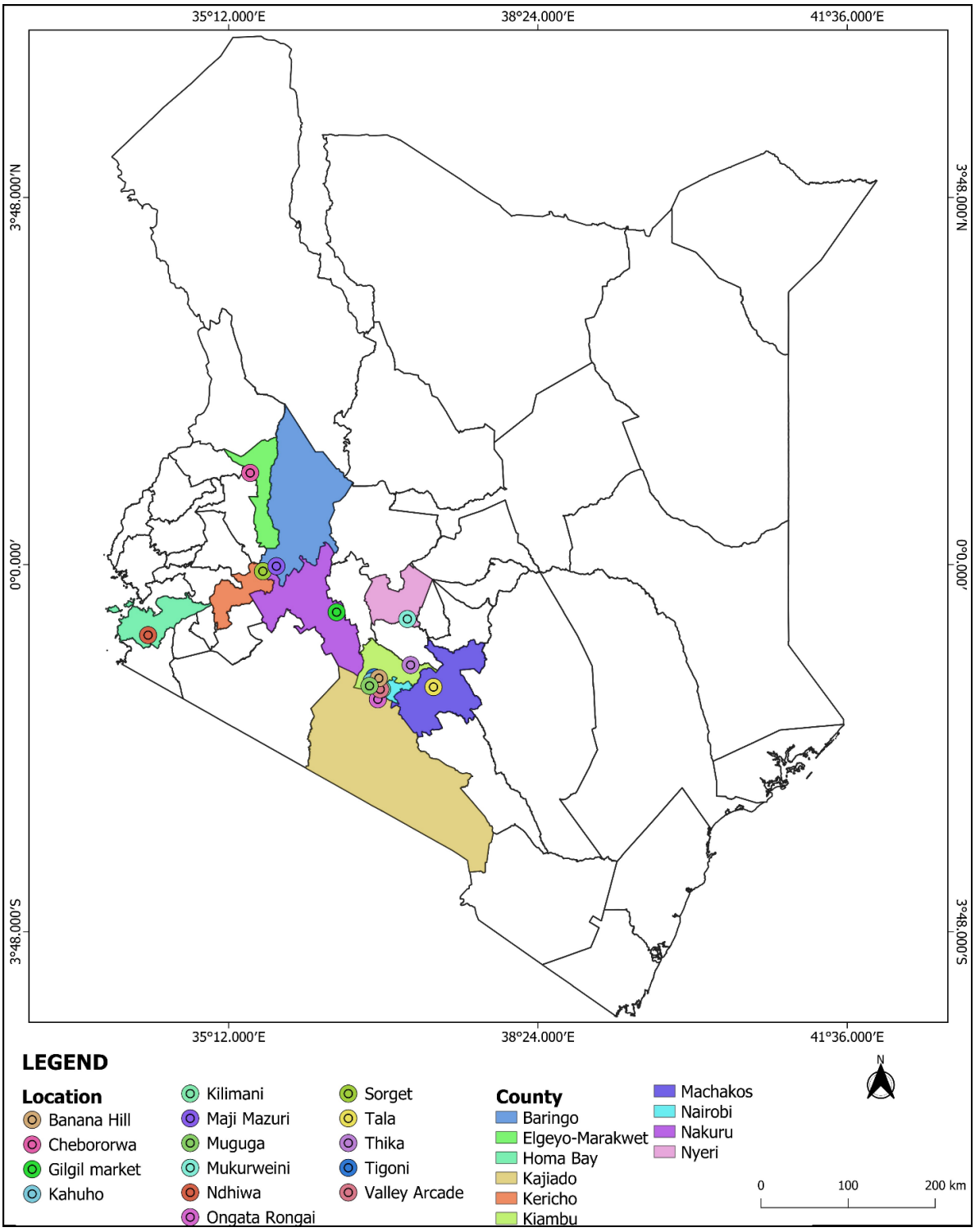


Figure 5.1. Map showing the location of *Physalis* sampling in 10 counties within Kenya.

Table 5.1. Geographical coordinates and number of samples picked from the various regions of *Physalis* sampling within 10 different counties in Kenya

No.	Main Location	Specific Location	Latitude	Longitude	Number of samples collected
1.	Londiani	Sorget	0.0532° S	35.5548° E	10
2.	Elgeyo Marakwet	Cheborорwa	0.9487° N	35.4234° E	13
3.	Homabay	Ndhiwa	0.7299° S	34.3671° E	3
4.	Baringo	Maji Mazuri	0.0076° S	35.6861° E	1
5.	Kajiado	Ongata Rongai	1.3939° S	36.7442° E	5
6.	Nyeri	Mukurweini	0.5609° S	36.0488° E	5
7.	Nakuru	Gilgil market	0.4923° S	36.3173° E	1
8.	Nairobi	Kilimani	1.2893° S	36.7869° E	4
		Valley Arcade	1.2907° S	36.7692° E	2
9.	Kangundo	Tala	1.2670° S	37.3201° E	4
10.	Kiambu	Tigoni	1.1651° S	36.7065° E	17
		Thika	1.0388° S	37.0834° E	9
		Kahuho	1.2082° S	36.6795° E	2
		Banana Hill	1.1760° S	36.7550° E	1
		Muguga	1.2551° S	36.6580° E	1

5.2.2 Molecular authentication of *Physalis* plants

5.2.2.1 Genomic DNA extraction

Physalis accessions genomic DNA was isolated, purified, viewed and stored as described in section 3.2.2.1.

5.2.2.2 Polymerase chain reaction and DNA sequencing

Polymerase chain reaction (PCR) amplification of *Physalis* accessions was done using DNA barcoding primer *rbcL* (Lledo *et al*., 1998) presented in Table 3.1. PCR amplification, purification of PCR products and sequencing was performed as described in section 3.2.2.2.

5.2.2.3 Sequence and phylogenetic analysis

The sequences of *rbcL* gene for *Physalis* accessions attained were curated, subjected to BLASTn analysis, used to prepare MSAs and phylogenetic trees as described in section 3.2.2.3. The MSA for the construction of MrBayes phylogenetic tree was prepared using the 56 *Physalis rbcL* gene sequences that had successfully been sequenced and BLASTn generated reference sequences for *Physalis rbcL* accession sequences. The MSA for the genetic diversity, nucleotide polymorphism, neutrality test and Automatic Barcode Gap Discovery (ABGD) analysis was generated using the 56 *Physalis rbcL* gene sequences only.

5.2.3 Genetic divergence within *Physalis* population based on *rbcL* sequences

DNA divergence within *Physalis* accession was assessed based on *rbcL* marker using DnaSP software version 6.12.03 (Kartavstev, 2011). DNA divergence within *Physalis* population based on *rbcL* sequences was done as described in section 3.2.3.

5.2.4 Genetic distance assessment within *Physalis* population based on *rbcL* sequences

Intraspecific and overall genetic distance of *Physalis* accessions based on the *rbcL* marker only was determined as described in section 3.2.4 using the 56 *Physalis* accessions *rbcL* sequences.

5.2.5 Determination of nucleotide polymorphism

Sequence polymorphism of the barcoding region *rbcL* was screened in all the sequenced *Physalis* accessions using the multiple sequence alignment for the 56 *Physalis* accessions *rbcL* sequences only as described in section 3.2.5.

5.2.6 Tajima's neutrality test

A Tajima's neutrality test for *rbcL* sequences of *Physalis* accessions under study was done to determine the frequency of mutations among *Physalis* population and to determine the selection of the population (Tajima, 1989). The test assessed the Tajima D value among *rbcL* sequences and nucleotide diversity using MEGA 11.0 (Nei and Kumar, 2000; Tamura *et al*., 2021). This analysis involved 56 *Physalis* accessions *rbcL* sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were 730 positions in the final dataset. Multiple sequence alignment used for this analysis was based on the 56 *rbcL* sequences only

5.2.7 Barcoding gap determination within *Physalis* accessions based on *rbcL* sequences

The multiple sequence alignment for the 56 *rbcL* sequences only was uploaded to ABGD website (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) and distance analysis performed based on K80 Kimura measure of distance as described in section 3.2.7.

5.3 Results

5.3.1 Amplification and sequencing success rate of *Physalis* accessions based on the *rbcL* marker

A total of 64 DNA samples out of 78 *Physalis* accessions were successfully amplified (Table 5.2 and Appendix 1 C). Some of the *rbcL* amplicons are indicated in gel images (Figure 3.3. and 5.2). The amplification success rate was 82% (Table 5.2). The amplicons from the 64 accessions were sequenced of which sequences were obtained from 56 *Physalis* accessions. The sequencing success rate was 88%. The *rbcL* sequences of the 56 *Physalis* accessions had a length range of 463 - 854 bp and an average length of 690 bp. The range of GC content was 42.1 - 45.5 % with an average of 43.4% (Table 5.2).

Table 5.2. Efficiency of PCR amplification and sequencing for *Physalis* accessions *rbcL* DNA barcode regions.

Barcode region	Samples tested (n)	Number of amplicons produced	Number of sequences produced	Percentage of amplification efficiency (%)	Percentage of sequencing efficiency (%)	Alignment length (bp)	Mean sequence length (bp)	Mean GC content (%)
rbcL	78	64	56	82	88	854	690	43.4

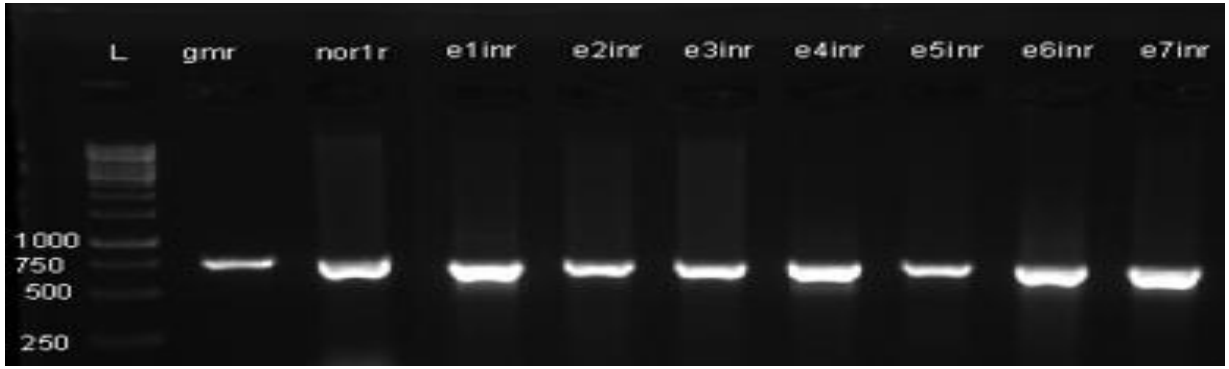


Figure 5.2. PCR amplicons for *rbcL* sequences of *Physalis* accessions from Nakuru, Kajiado and Elgeyo-Marakwet counties.

The letter L in Figure 5.2 above represents the DNA ladder, gmr-e7inr represents the *Physalis* accessions amplicons based on *rbcL* sequence amplification.

The nucleotide base frequencies of the four standard deoxyribonucleotides at various coding positions in the *Physalis* accessions for *rbcL* sequences is presented in Table 5.3. The most abundant deoxyribonucleotide was thymine while cytosine appeared the least in the *rbcL* sequences. It was noted that the GC content of the *rbcL* sequences was significantly lower than the AT content (Table 5.4).

Table 5.3. The nucleotide base frequencies of candidate nucleotide sequences at different coding positions in *Physalis* accessions.

Barcode region	**Base contents (%)**					
	A	**T**	**G**	**C**	**AT**	**GC**
rbcL	28.17	28.44	23.13	20.26	56.56	43.44

5.3.2 Identification of *Physalis* species based using BLASTn analysis based on *rbcL* sequences

The BLASTn analysis results for the *rbcL* sequences of 56 *Physalis* accessions that were successfully sequenced are presented in Table 5.4. All 56 (100%) of the *Physalis* accessions *rbcL* sequences were identified as *Physalis* species after BLASTn analysis. Eight out of the 56 *rbcL* sequences of *Physalis* accessions gave 100% similarity with GenBank database sequences. The eight *Physalis* accessions with 100% identity match in the GenBank include OQ507152.1, OQ507153.1, OQ507155.1, OQ507157.1, OQ507158.1, OQ507159.1, OQ507160.1 and OQ507161.1. All eight sequences of the *Physalis* accessions were identified as *P. minima* based on BLASTn analysis.

Table 5.4. BLASTn analysis results for the *Physalis* accessions based on *rbcL* barcode gene.

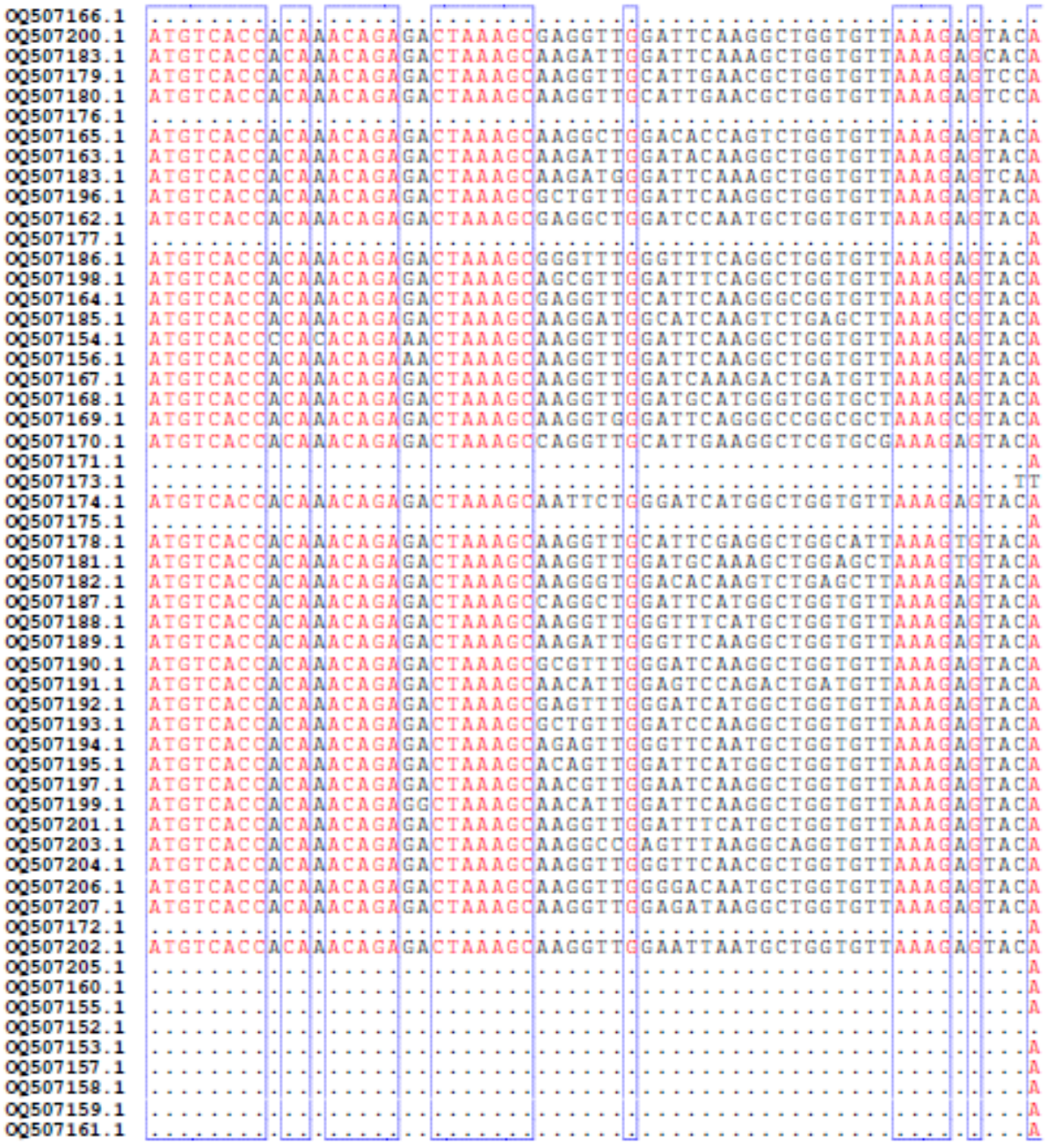
Sample ID	Species of Best BLAST match	GenBank accession number (of database)	Max score	E value	Percent identity (%)	Percent coverage (%)	GenBank accession number
Chebororwa C2r	*P. peruviana*	NC_026570.1	1136	0.0	94.90	99	OQ507163.1
Chebororwa C4r	*P. peruviana*	NC_026570.1	1347	0.0	99.33	99	OQ507164.1
Chebororwa C5r	*P. virginiana*	KT178121.1	739	0.0	95.46	100	OQ507165.1
Chebororwa e1inr	*P. minima*	NC_048515.1	754	0.0	93.18	100	OQ507166.1
Chebororwa e4inr	*P. minima*	NC_048515.1	1351	0.0	99.46	100	OQ507167.1
Chebororwa e6inr	*P. minima*	NC_048515.1	1424	0.0	98.28	99	OQ507168.1
Chebororwa e7inr	*P. minima*	NC_048515.1	1343	0.0	99.19	100	OQ507169.1
Chebororwa e8inr	*P. minima*	NC_048515.1	1253	0.0	98.19	100	OQ507170.1
Gilgil gmi	*P. minima*	NC_048515.1	1144	0.0	99.06	100	OQ507171.1
Maji mazuri mmr	*P. minima*	NC_048515.1	798	0.0	92.20	99	OQ507176.1
Valley Arcade Vanr	*P. minima*	NC_048515.1	1435	0.0	97.28	99	OQ507207.1
Kilimani SHRr	*P. minima*	NC_048515.1	1267	0.0	99.29	99	OQ507175.1
Kilimani KL1	*P. minima*	NC_048515.1	1308	0.0	98.39	100	OQ507174.1
Kilimani ANRr	*P. minima*	NC_048515.1	1280	0.0	98.75	99	OQ507173.1
Mukurweini ny1r	*P. minima*	NC_048515.1	1411	0.0	96.49	100	OQ507178.1
Mukurweini ny4r	*P. angulata*	NC_039457.1	1092	0.0	94.01	99	OQ507179.1
Mukurweini ny5r	*P. angulata*	NC_039457.1	1027	0.0	92.90	98	OQ507180.1
Ndhiwa hb2r	*P. minima*	NC_048515.1	1341	0.0	93.33	99	OQ507181.1
Ndhiwa hb3r	*P. minima*	NC_048515.1	1327	0.0	999.05	99	OQ507182.1
Ongata Rongai nor3r	*P. peruviana*	NC_026570.1	1303	0.0	95.47	99	OQ507183.1
Ongata Rongai nor4r	*P. virginiana*	KT178121.1	1050	0.0	91.10	99	OQ507184.1

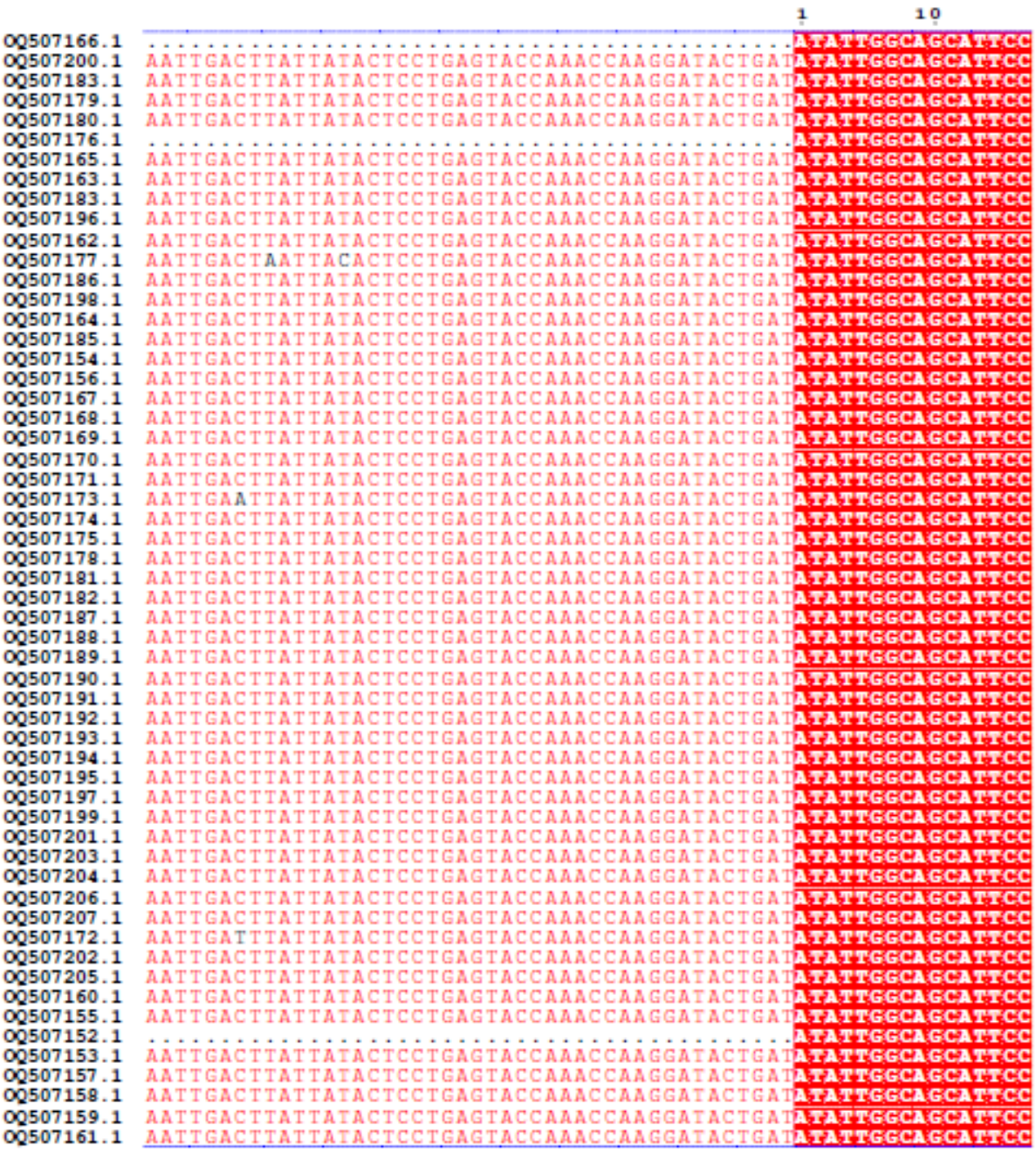
Ongata Rongai nor5r	*P. peruviana*	NC_026570.1	1323	0.0	98.92	99	OQ507185.1
Tala KTr	*P. minima*	NC_048515.1	1123	0.0	97.01	100	OQ507186.1
Banana Hill JCBKr	*P. minima*	NC_048515.1	1192	0.0	96.28	99	OQ507162.1
Kahuho KKr	*P. minima*	NC_048515.1	1016	0.0	99.82	100	OQ507172.1
Muguga DMr	*P. minima*	NC_048515.1	845	0.0	98.74	100	OQ507177.1
Thika TK2r	*P. minima*	NC_048515.1	1310	0.0	99.17	99	OQ507187.1
Thika TK3r	*P. minima*	NC_048515.1	1352	0.0	97.59	99	OQ507188.1
Thika TK4r	*P. minima*	NC_048515.1	1363	0.0	97.04	99	OQ507189.1
Thika TK6r	*P. minima*	NC_048515.1	1336	0.0	97.33	99	OQ507190.1
Thika TK7r	*P. minima*	NC_048515.1	1310	0.0	99.04	99	OQ507191.1
Thika TK8r	*P. minima*	NC_048515.1	1120	0.0	99.04	99	OQ507192.1
Thika TK9r	*P. minima*	NC_048515.1	1411	0.0	98.03	99	OQ507193.1
Tigoni T2r	*P. minima*	NC_048515.1	1053	0.0	98.31	100	OQ507194.1
Tigoni T4r	*P. minima*	NC_048515.1	1384	0.0	96.42	99	OQ507195.1
Tigoni T6r	*P. minima*	NC_048515.1	785	0.0	97.20	100	OQ507196.1
Tigoni T7r	*P. minima*	NC_048515.1	1315	0.0	98.14	93	OQ507197.1
Tigoni T8r	*P. minima*	NC_048515.1	963	0.0	98.89	99	OQ507198.1
Tigoni T9r	*P. minima*	NC_048515.1	1399	0.0	98.13	100	OQ507199.1
Tigoni T10r	*P. minima*	NC_048515.1	824	0.0	94.25	99	OQ50200.1
Tigoni T11r	*P. minima*	NC_048515.1	1330	0.0	99.32	98	OQ507201.1
Tigoni T12r	*P. minima*	NC_048515.1	985	0.0	99.27	100	OQ507202.1
Tigoni T13r	*P. minima*	NC_048515.1	1319	0.0	97.54	99	OQ507203.1
Tigoni T14r	*P. minima*	NC_048515.1	1371	0.0	98.70	99	OQ507204.1
Tigoni T16r	*P. minima*	NC_048515.1	1297	0.0	98.25	99	OQ507205.1
Tigoni T17r	*P. minima*	NC_048515.1	1376	0.0	98.34	97	OQ507206.1
Londiani L1r	*P. minima*	NC_048515.1	1011	0.0	100	100	OQ507152.1
Londiani L2r	*P. minima*	NC_048515.1	1136	0.0	100	100	OQ507153.1
Londiani L3r	*P. minima*	NC_048515.1	1338	0.0	99.59	100	OQ507154.1
Londiani L4r	*P. minima*	NC_048515.1	961	0.0	100	100	OQ507155.1
Londiani L5r	*P. minima*	NC_048515.1	1343	0.0	99.86	100	OQ507156.1
Londiani L6r	*P. minima*	NC_048515.1	1171	0.0	100	100	OQ507157.1
Londiani L7r	*P. minima*	NC_048515.1	1136	0.0	100	100	OQ507158.1
Londiani L8r	*P. minima*	NC_048515.1	1147	0.0	100	100	OQ507159.1
Londiani L9r	*P. minima*	NC_048515.1	1003	0.0	100	100	OQ507160.1
Londiani L10r	*P. minima*	NC_048515.1	1116	0.0	100	100	OQ507161.1

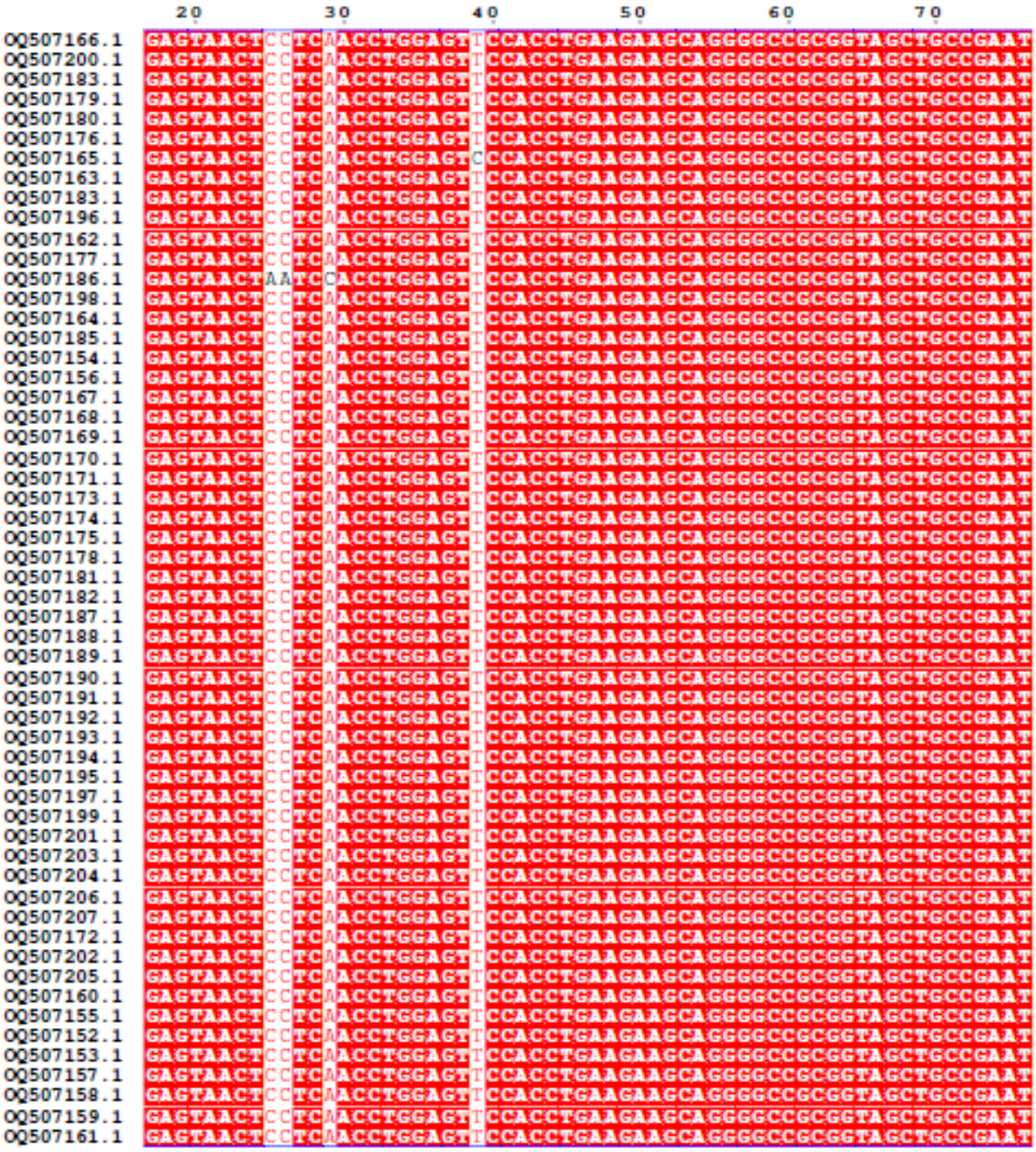
5.3.3 Multiple sequence alignment

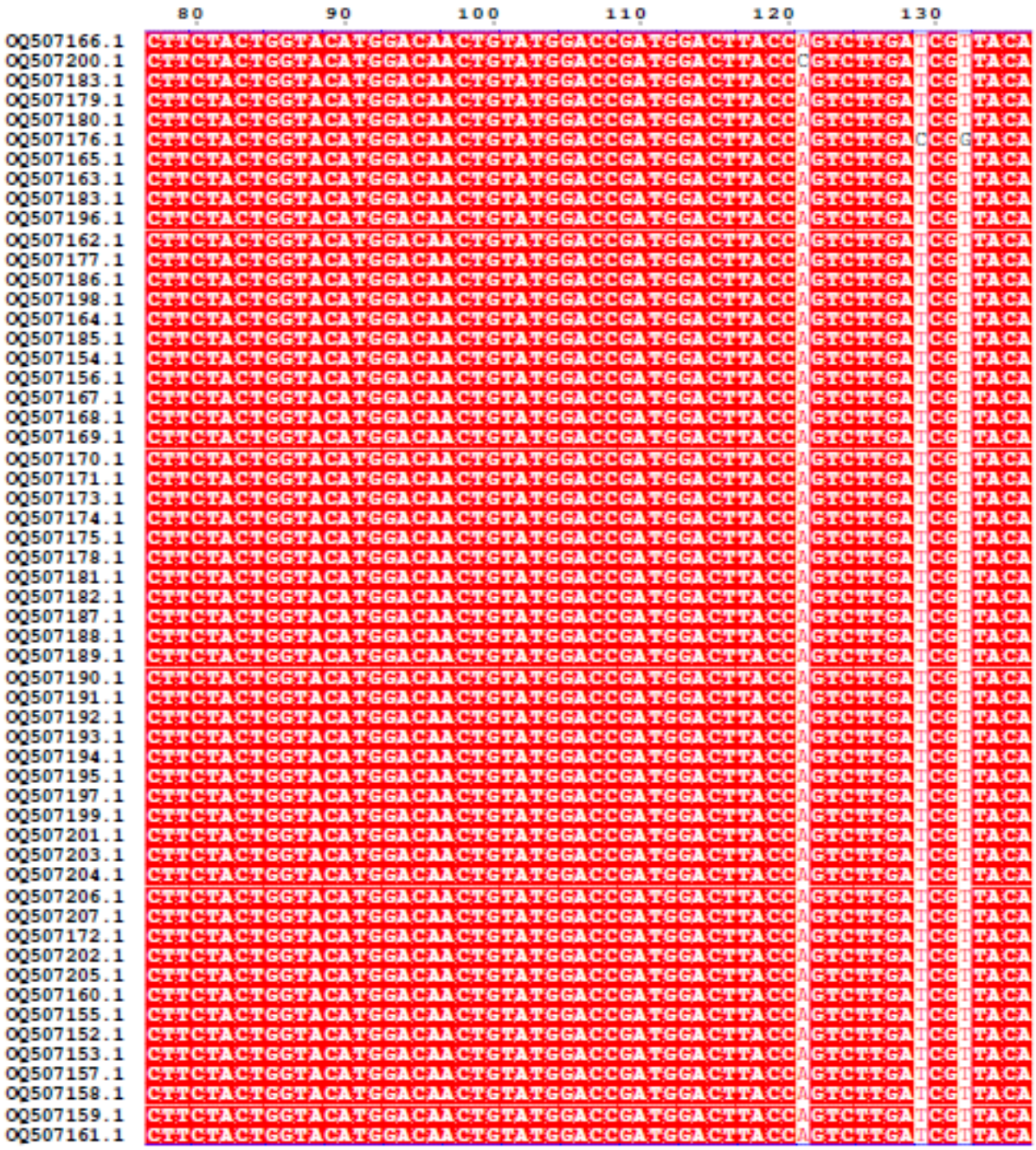
The multiple sequence alignment of curated *rbcL* sequences and their reference sequences from BLASTn analysis was prepared based on MUSCLE and had a sequence length of 854 bp (Appendix 2 C). The alignment was mostly conserved with few substitution point mutations, absence of deletion or addition (frameshift) point mutations and absence of macrolesions (Appendix 2 C). At position 130 and 131 of the MSA a transversion substitution point mutation was identified on *Physalis* accessions OQ507186.1 where cytosine was replaced with adenine (Appendix 2 C). All other accessions had a cytosine base at this position. Another transversion point mutation was noted at position 135 of the same *Physalis* accession where adenine was replaced by cytosine (Appendix 2 C). At position 144 of the MSA a transition point mutation was identified for *Physalis* accession OQ507165.1 where thymine was replaced by cytosine as compared to all other accessions (Appendix 2 C). Another transition point mutation is identified at position 368 of the MSA on *Physalis* accession OQ507166.1 where guanine was replaced by adenine (Appendix 2 C).

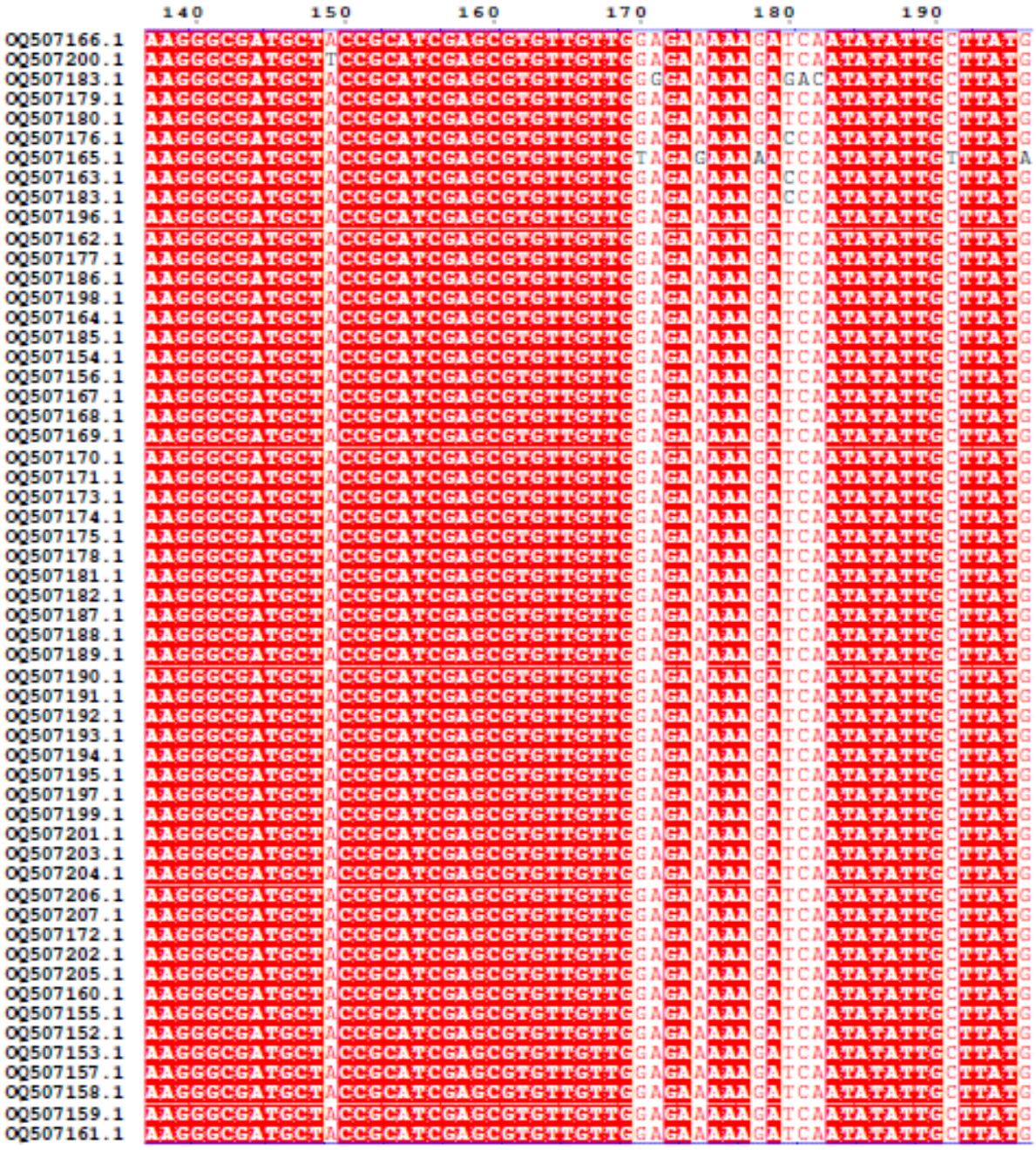
A second MSA was prepared for the *rbcL* sequences only based on MUSCLE, trimmed and edited on Jalview for genetic diversity and polymorphism analysis. This multiple sequence alignment had a sequence length of 730 bp (https://espript.ibcp.fr/ESPript/temp/1947945090/0-0-1683720272-esp.pdf accessed on 17th February 2023) (Figure 5.3). These MSA had no macrolesion or frameshift point mutation (Figure 5.3). It was highly conserved with few substitution point mutations (Figure 5.3). At position 25, 26 and 29 of the MSA, *Physalis* accession OQ507186.1 had transversion point mutations where cytosine replaced adenine at positions 25 and 26 and adenine was replaced by cytosine at position 29 (Figure 5.3). A transition point mutation was seen on the MSA at position 39 where for the *Physalis* accession OQ507166.1 cytosine replaced thymine (Figure 5.3).

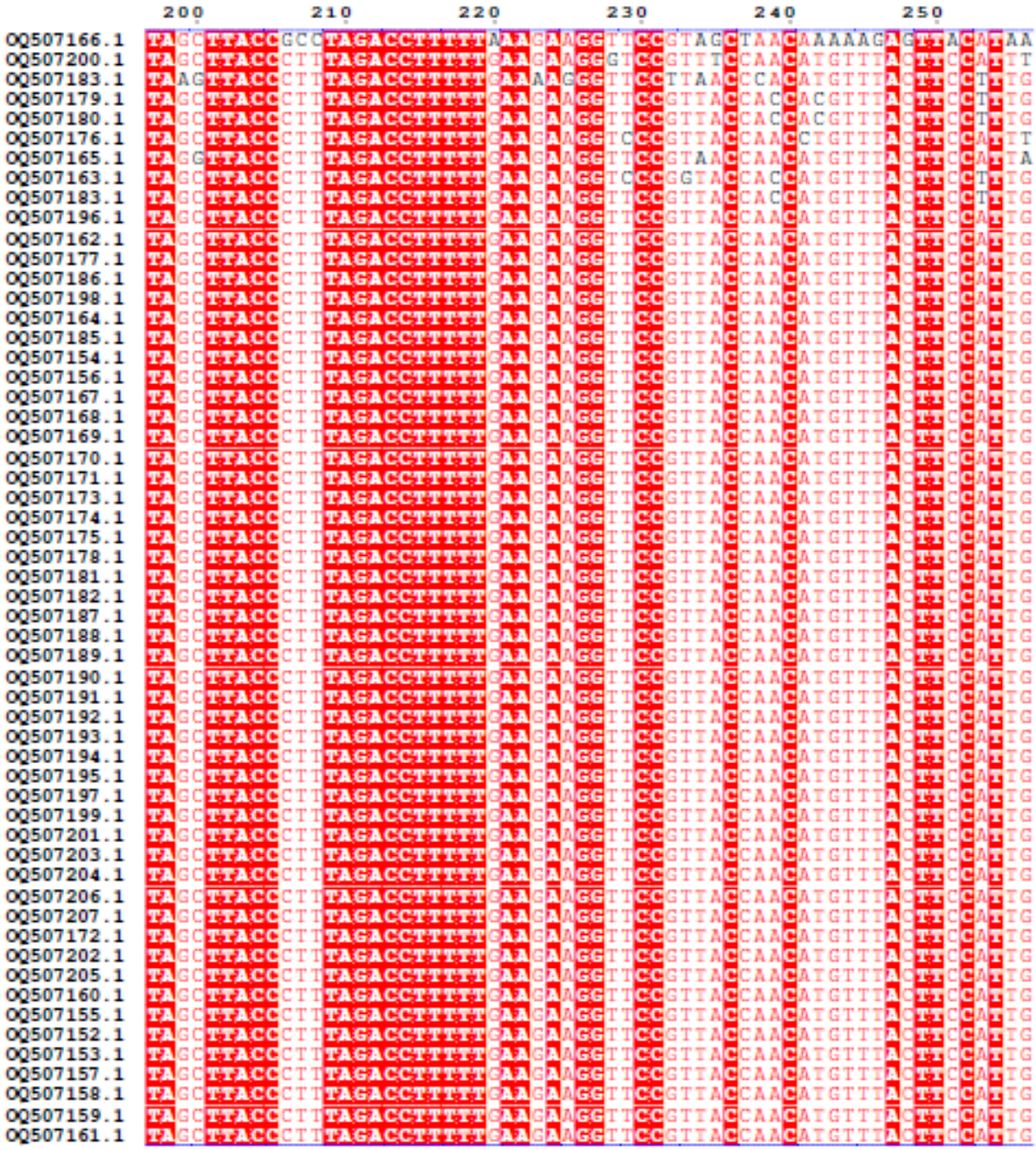

OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1

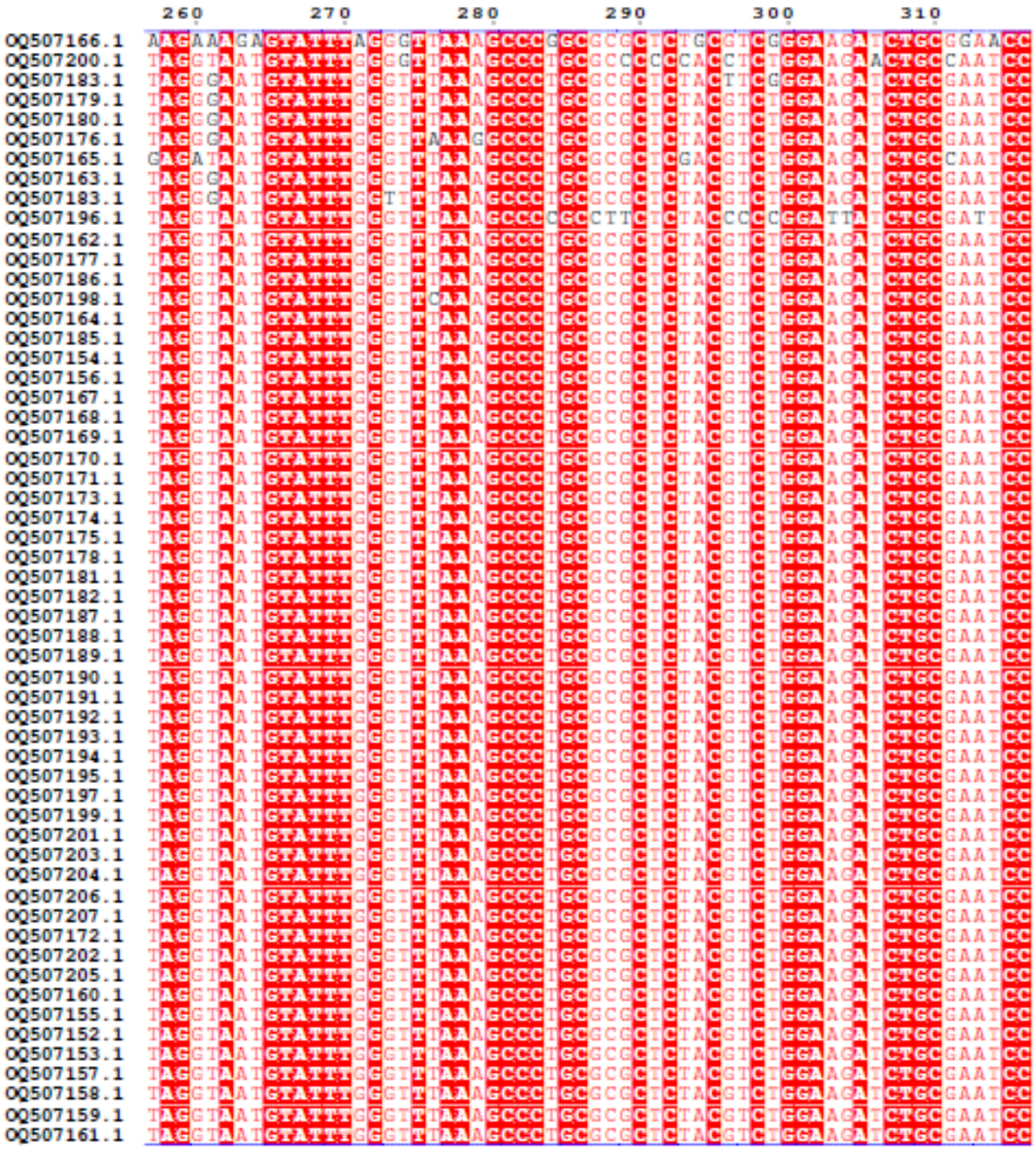

1
10
OQ507166.1 .....ATATTGGCAGCATTCC
OQ507200.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507183.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507179.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507180.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507176.1 .....ATATTGGCAGCATTCC
OQ507165.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507163.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507183.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507196.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507162.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507177.1 AATTGACTAATTACACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507186.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507198.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507164.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507185.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507154.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507156.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507167.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507168.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507169.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507170.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507171.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507173.1 AATTGAATTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507174.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507175.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507178.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507181.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507182.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507187.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507188.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507189.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507190.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507191.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507192.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507193.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507194.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507195.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507197.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507199.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507201.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507203.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507204.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507206.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507207.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507172.1 AATTGATTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507202.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507205.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507160.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507155.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507152.1 .....ATATTGGCAGCATTCC
OQ507153.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507157.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507158.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507159.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC
OQ507161.1 AATTGACTTATTATACTCCTGAGTACCAAACCAAGGATACTGATATATTGGCAGCATTCC

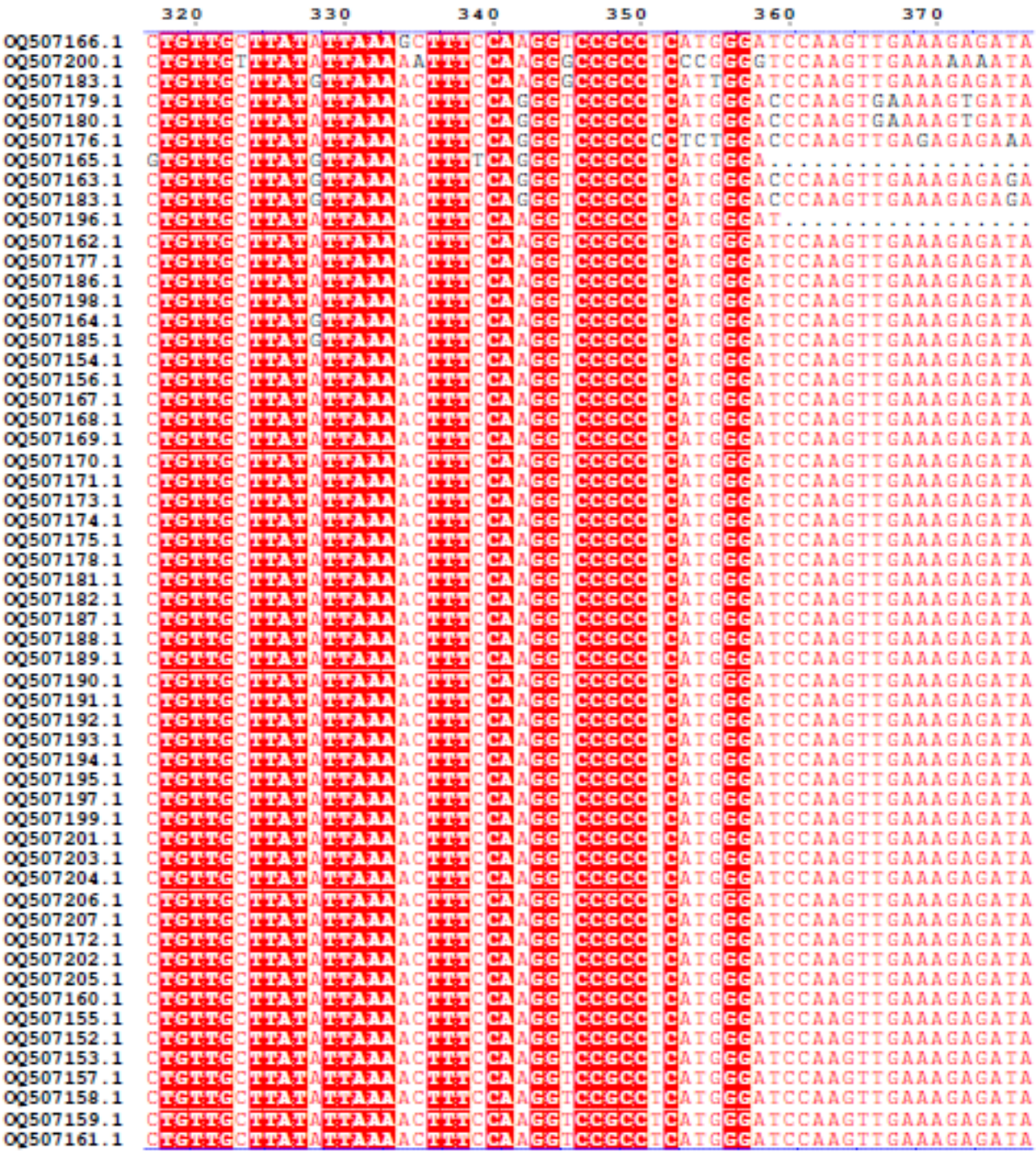

20
30
40
50
60
70
OQ507166.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507200.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507183.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507179.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507180.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507176.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507165.1 GAGTAACTCCTCAACCTGGAGTCCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507163.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507183.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507196.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507162.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507177.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507186.1 GAGTAACTAATCCACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507198.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507164.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507185.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507154.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507156.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507167.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507168.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507169.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507170.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507171.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507173.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507174.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507175.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507178.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507181.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507182.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507187.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507188.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507189.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507190.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507191.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507192.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507193.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507194.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507195.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507197.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507199.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507201.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507203.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507204.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507206.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507207.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507172.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507202.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507205.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507160.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507155.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507152.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507153.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507157.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507158.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507159.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT
OQ507161.1 GAGTAACTCCTCAACCTGGAGTTCCACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAAT

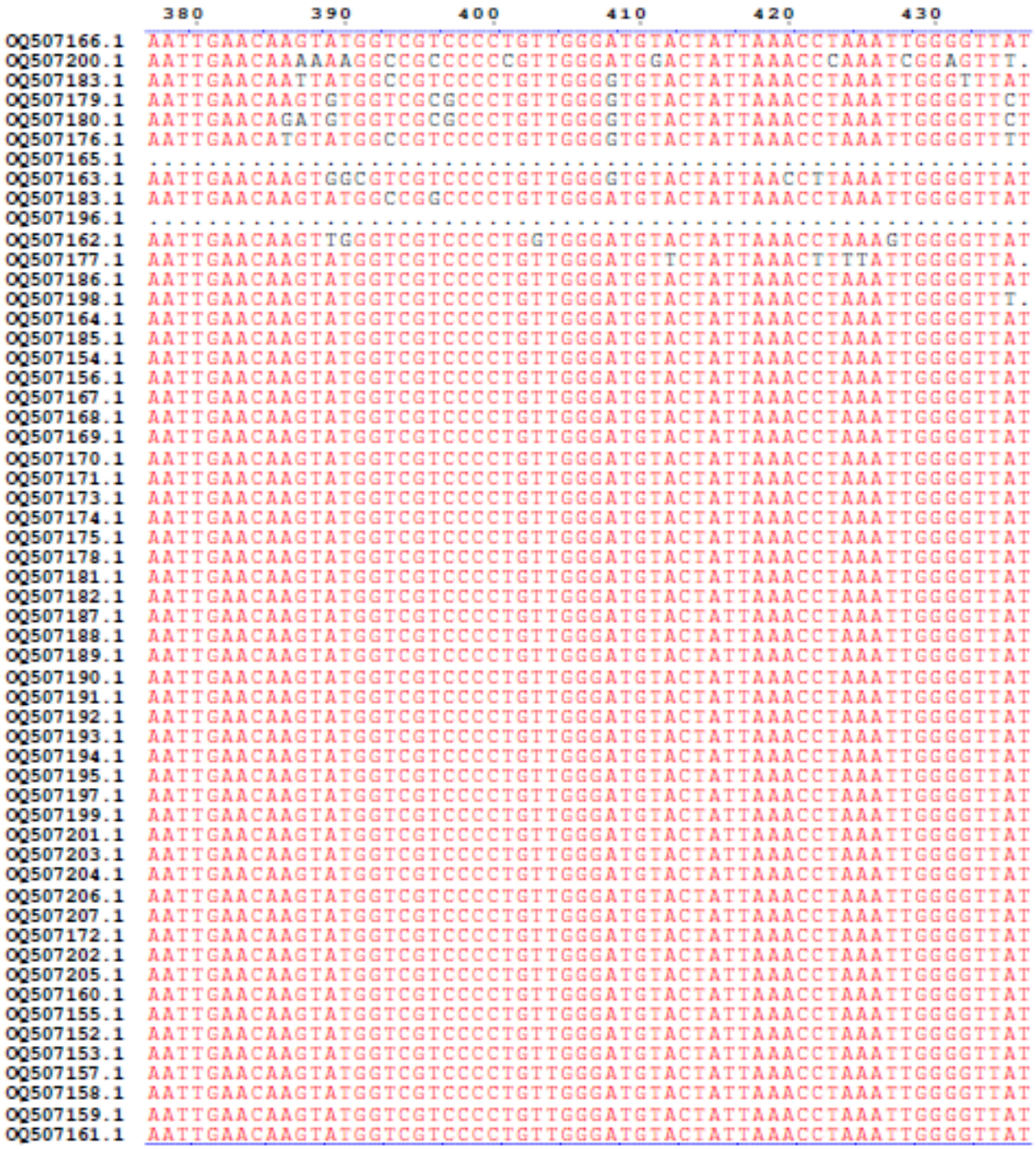

80
90
100
110
120
130
OQ507166.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507200.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCCGTCTTGAICGITACA
OQ507183.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507179.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507180.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507176.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGACCGGTACA
OQ507165.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507163.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507183.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507196.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507162.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507177.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507186.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507198.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507164.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507185.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507154.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507156.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507167.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507168.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507169.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507170.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507171.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507173.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507174.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507175.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507178.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507181.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507182.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507187.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507188.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507189.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507190.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507191.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507192.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507193.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507194.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507195.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507197.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507199.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507201.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507203.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507204.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507206.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507207.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507172.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507202.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507205.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507160.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507155.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507152.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507153.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507157.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507158.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507159.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA
OQ507161.1 CTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAICGITACA

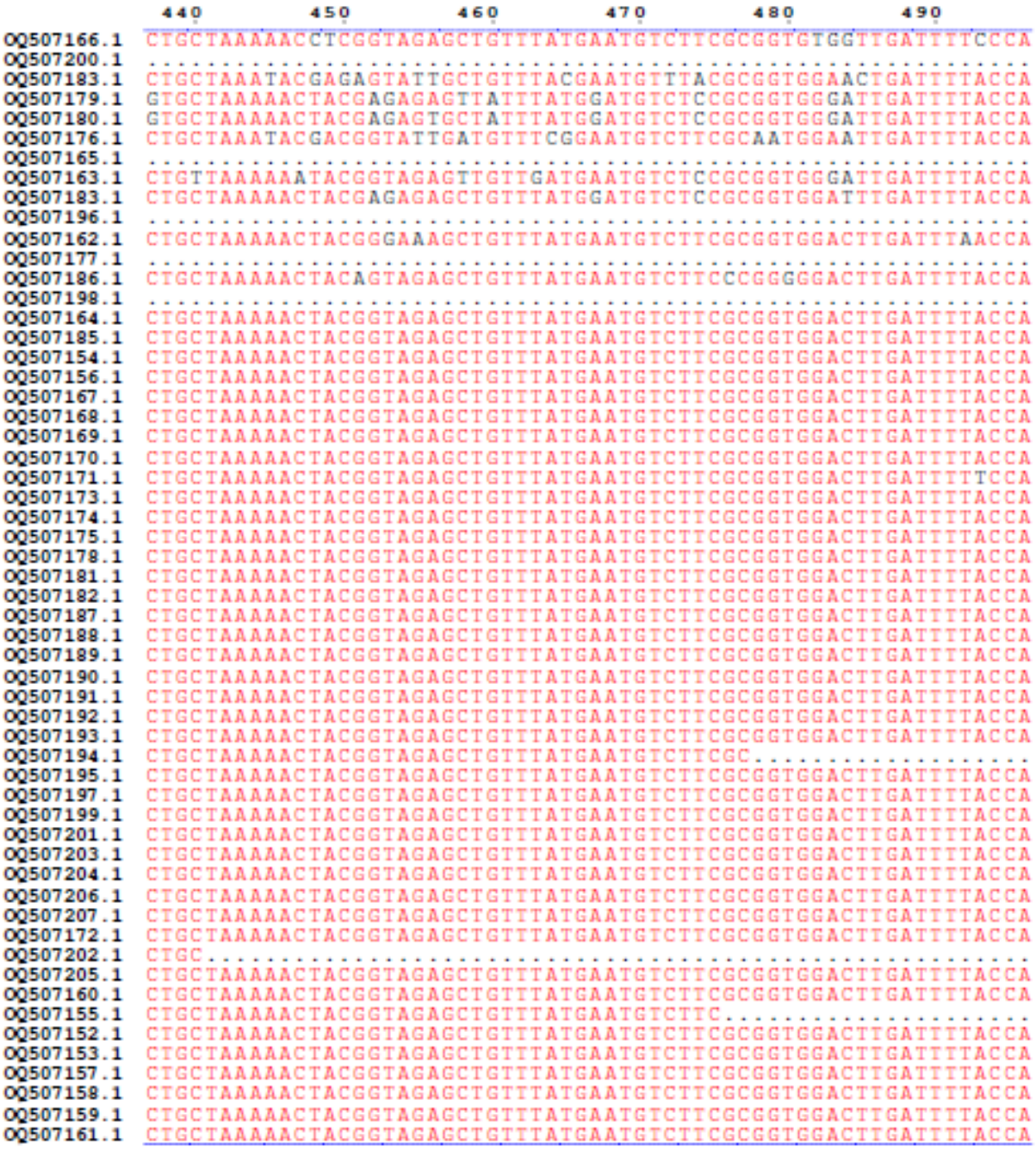

140
150
160
170
180
190
OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1

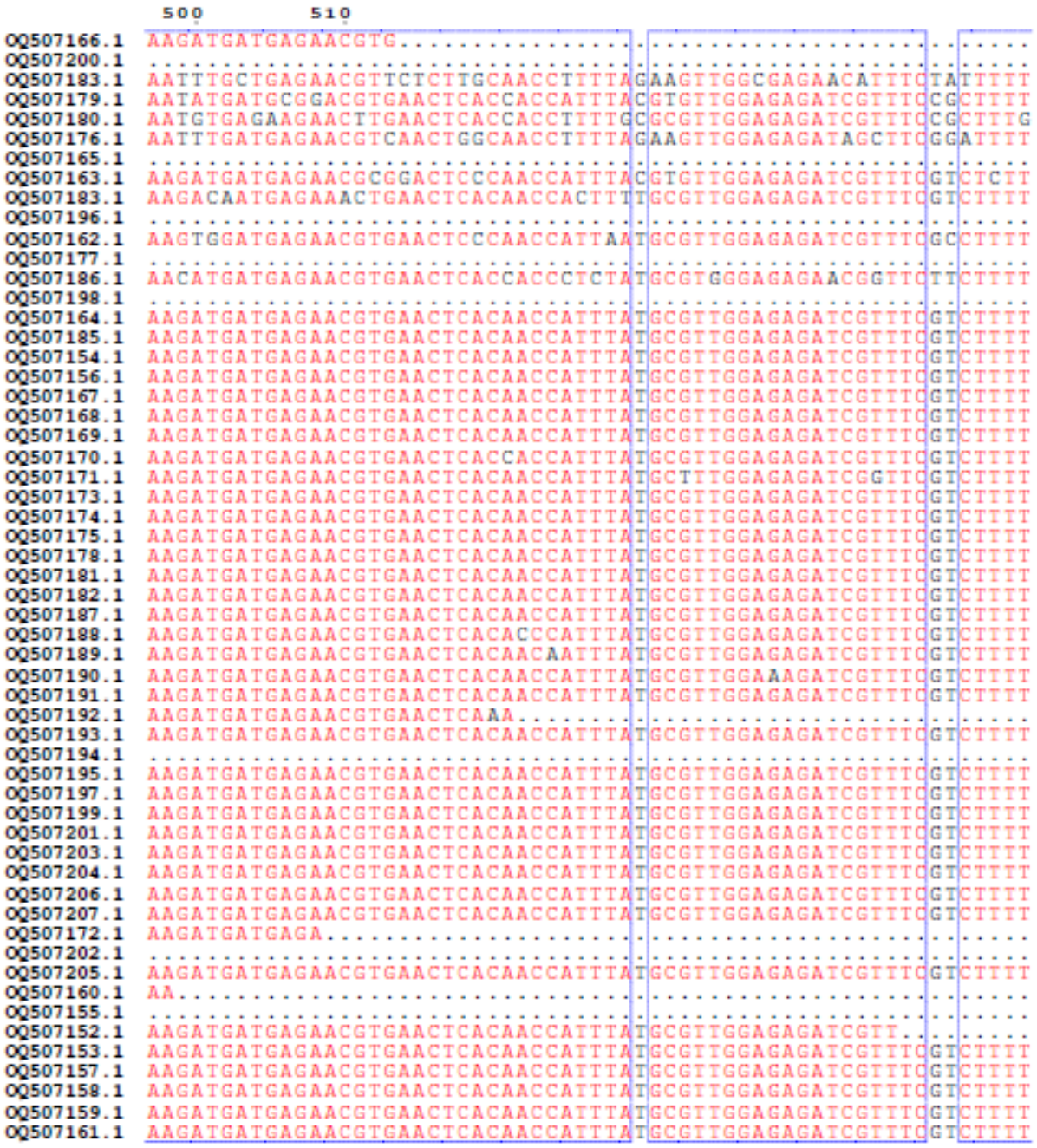

200
210
220
230
240
250
OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1

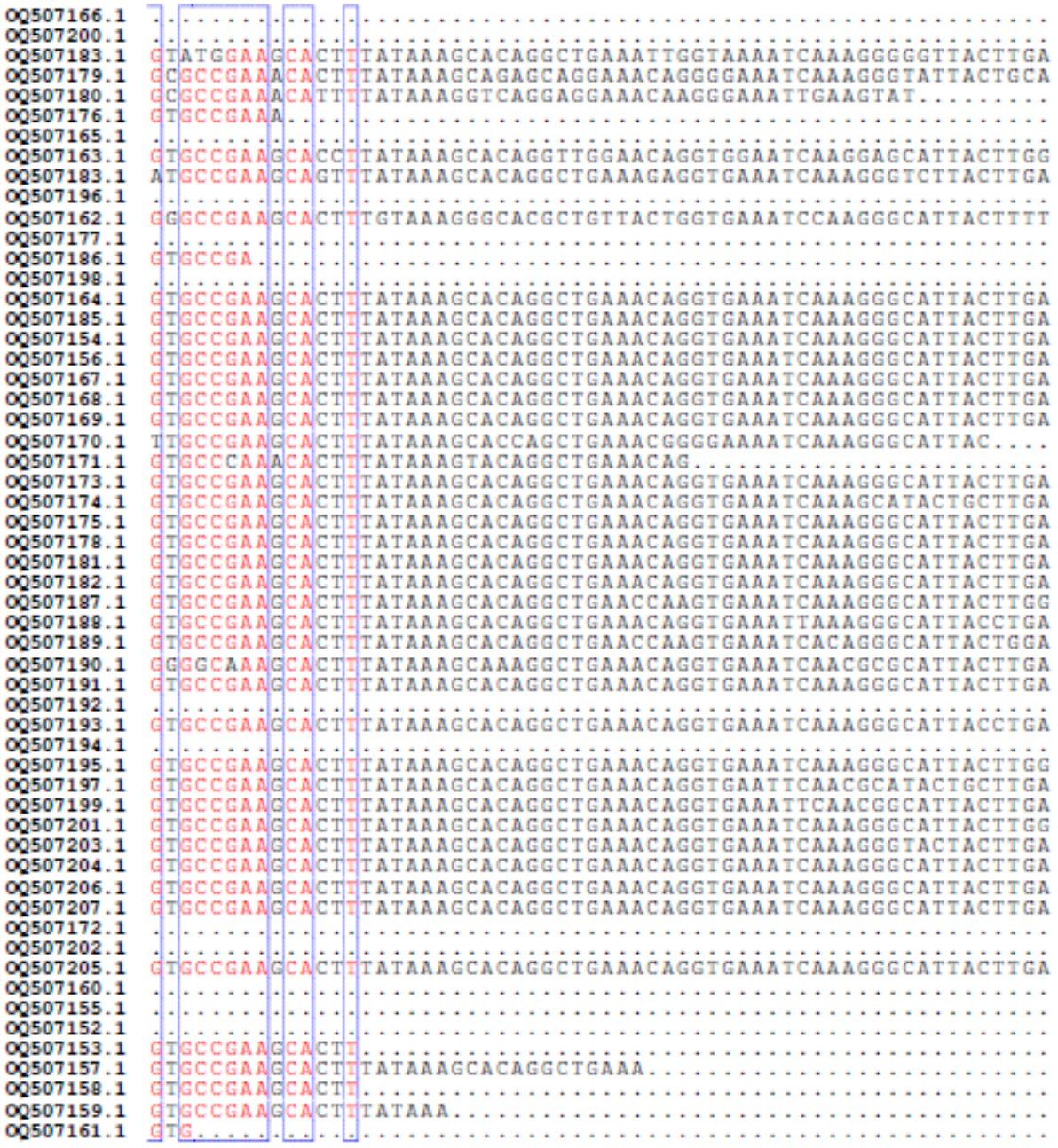

260
270
280
290
300
310
OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1


320
330
340
350
360
370
OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1


380 390 400 410 420 430
OQ507166.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507200.1 AATTGAACAAAAAAGGCCGCCCCCCGTTGGGATGGACTATTAAACCCAAATCGGAGTTT.
OQ507183.1 AATTGAACAATTATGGCCGTCCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGTTTAT
OQ507179.1 AATTGAACAAGTGTGGTCGCGCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTCT
OQ507180.1 AATTGAACAGATGTGGTCGCGCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTCT
OQ507176.1 AATTGAACATGTATGGCCGTCCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTTT
OQ507165.1 ....
OQ507163.1 AATTGAACAAGTGGCGTCGTCCCCTGTTGGGGTGTACTATTAACCTTAAATTGGGGTTAT
OQ507183.1 AATTGAACAAGTATGGCCGGCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507196.1 ....
OQ507162.1 AATTGAACAAGTTGGGTCGTCCCCTGGTGGGATGTACTATTAAACCTAAAGTGGGGTTAT
OQ507177.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTTCTATTAAACTTTTATTGGGGTTA.
OQ507186.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507198.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTT.
OQ507164.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507185.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507154.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507156.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507167.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507168.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507169.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507170.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507171.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507173.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507174.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507175.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507178.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507181.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507182.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507187.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507188.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507189.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507190.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507191.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507192.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507193.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507194.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507195.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507197.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507199.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507201.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507203.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507204.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507206.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507207.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507172.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507202.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507205.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507160.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507155.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507152.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507153.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507157.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507158.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507159.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT
OQ507161.1 AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTAT


440
450
460
470
480
490
OQ507166.1 CTGCTAAAAACCTCGGTAGAGCTGTTTATGAATGTCTTCGCGGTGTGGTTGATTTTCCCA
OQ507200.1 ....
OQ507183.1 CTGCTAAATACGAGAGTATTGCTGTTTACGAATGTTTACGCGGTGGAACTGATTTTACCA
OQ507179.1 GTGCTAAAAACTACGAGAGAGTTATTTATGGATGTCTCCGCGGTGGGATTGATTTTACCA
OQ507180.1 GTGCTAAAAACTACGAGAGTGCTATTTATGGATGTCTCCGCGGTGGGATTGATTTTACCA
OQ507176.1 CTGCTAAATACGACGGTATTGATGTTTCGGAATGTCTTCGCAATGGAATTGATTTTACCA
OQ507165.1 ....
OQ507163.1 CTGTTAAAAAATACGGTAGAGTTGTTGATGAATGTCTCCGCGGTGGGATTGATTTTACCA
OQ507183.1 CTGCTAAAAACTACGAGAGAGCTGTTTATGGATGTCTCCGCGGTGGATTTGATTTTACCA
OQ507196.1 ....
OQ507162.1 CTGCTAAAAACTACGGGAAAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTAACCA
OQ507177.1 ....
OQ507186.1 CTGCTAAAAACTACAGTAGAGCTGTTTATGAATGTCTTCCCGGGGGACTTGATTTTACCA
OQ507198.1 ....
OQ507164.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507185.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507154.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507156.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507167.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507168.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507169.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507170.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507171.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTTCCA
OQ507173.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507174.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507175.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507178.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507181.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507182.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507187.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507188.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507189.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507190.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507191.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507192.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507193.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507194.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGC..................
OQ507195.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507197.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507199.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507201.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507203.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507204.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507206.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507207.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507172.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507202.1 CTGC..
OQ507205.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507160.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507155.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTC......................
OQ507152.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507153.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507157.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507158.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507159.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA
OQ507161.1 CTGCTAAAAACTACGGTAGAGCTGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCA


500
510
OQ507166.1
OQ507200.1
OQ507183.1
OQ507179.1
OQ507180.1
OQ507176.1
OQ507165.1
OQ507163.1
OQ507183.1
OQ507196.1
OQ507162.1
OQ507177.1
OQ507186.1
OQ507198.1
OQ507164.1
OQ507185.1
OQ507154.1
OQ507156.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507178.1
OQ507181.1
OQ507182.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507197.1
OQ507199.1
OQ507201.1
OQ507203.1
OQ507204.1
OQ507206.1
OQ507207.1
OQ507172.1
OQ507202.1
OQ507205.1
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507161.1


OQ507166.1
OQ507200.1
OQ507183.1 GTATGGAAGCACTTTATAAAGCACAGGCTGAAATTGGTAAAATCAAAGGGGGTTACTTGA
OQ507179.1 GCGCCGAAACACTTTATAAAGCAGAGCAGGAAACAGGGGAAATCAAAGGGTATTACTGCA
OQ507180.1 GCGCCGAAACATTTTATAAAGGTCAGGAGGAAACAAGGGAAATTGAAGTAT.....
OQ507176.1 GTGCCGAAA.....
OQ507165.1
OQ507163.1 GTGCCGAAGCACCTTATAAAGCACAGGTTGGAACAGGTGGAATCAAGGAGCATTACTTGG
OQ507183.1 ATGCCGAAGCAGTTTATAAAGCACAGGCTGAAAGAGGTGAAATCAAAGGGTCTTACTTGA
OQ507196.1
OQ507162.1 GGGCCGAAGCACTTTGTAAAGGGCACGCTGTTACTGGTGAAATCCAAGGGCATTACTTTT
OQ507177.1
OQ507186.1 GTGCCGA.....
OQ507198.1
OQ507164.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507185.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507154.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507156.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507167.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507168.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507169.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507170.1 TTGCCGAAGCACTTTATAAAGCACCAGCTGAAACGGGGAAAATCAAAGGGCATTAC....
OQ507171.1 GTGCCCAAACACTTTATAAAGTACAGGCTGAAACAG.....
OQ507173.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507174.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGCATACTGCTTGA
OQ507175.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507178.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507181.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507182.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507187.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAACCAAGTGAAATCAAAGGGCATTACTTGG
OQ507188.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATTAAAGGGCATTACCTGA
OQ507189.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAACCAAGTGAAATCACAGGGCATTACTGGA
OQ507190.1 GGGGCAAAGCACTTTATAAAGCAAAGGCTGAAACAGGTGAAATCAACGCGCATTACTTGA
OQ507191.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507192.1
OQ507193.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACCTGA
OQ507194.1
OQ507195.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGG
OQ507197.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAATTCAACGCATACTGCTTGA
OQ507199.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATTCAACGGCATTACTTGA
OQ507201.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGG
OQ507203.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGTACTACTTGA
OQ507204.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507206.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507207.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507172.1
OQ507202.1
OQ507205.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
OQ507160.1
OQ507155.1
OQ507152.1
OQ507153.1 GTGCCGAAGCACTT.....
OQ507157.1 GTGCCGAAGCACTTTATAAAGCACAGGCTGAAA.....
OQ507158.1 GTGCCGAAGCACTT.....
OQ507159.1 GTGCCGAAGCACTTTATAAA.....
OQ507161.1 GTG.....

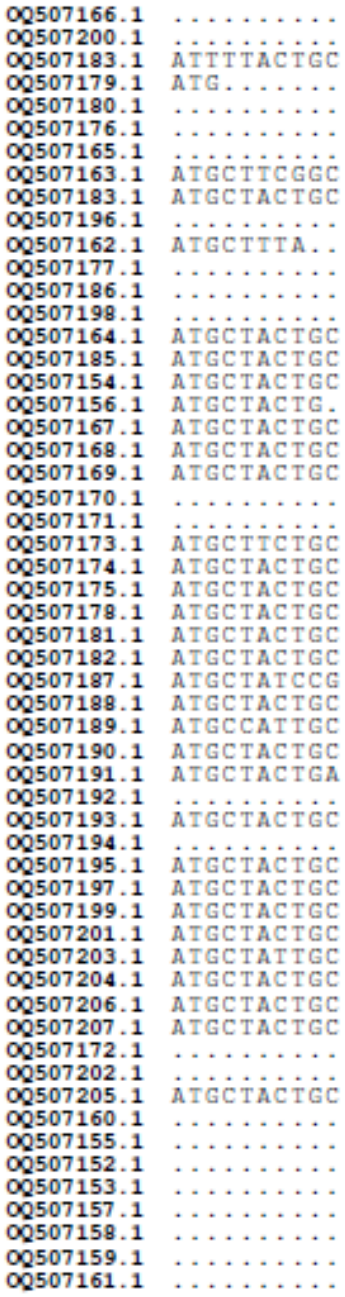


Figure 5.3. Multiple sequence alignment for 56 *Physalis* accessions only based on *rbcL* marker.

5.3.4 Phylogenetic analysis

The *rbcL* gene was not able to facilitate species discrimination based on the phylogenetic tree (Figure 5.4). There was no cladding of *Physalis* accessions with their representative reference

sequences from the GenBank database. It was therefore impossible to discriminate *Physalis* accessions based on their *rbcL* gene sequences and species names assigned were based on BLASTn analysis.

Figure 5.4. Phylogenetic tree generated by MrBayes for *Physalis* accessions based on *rbcL* marker.

Different colors were used to represent *Physalis* accessions and reference sequences. Black represents all reference sequences, orange represents *P. minima*, purple represents *P. peruviana*, pink represents *P. angulata* and blue represents *Physalis virginiana*. The names of species are based on BLASTn analysis as there was no discrimination of species based on the Phylogenetic tree.

There was no cladding of *Physalis* accessions with there reference sequences. The *Physalis* accessions also clustered with different reference sequences other than their counterparts for example the accessions identified as *P. peruviana* clustered with *P. angulata* and *P. minima* accessions and not with the *P. peruviana* reference sequence as was expected. This is an indication that the *rbcL* barcode was unable to effectively identify the *Physalis* accessions studied based on the phylogenetic analysis.

5.3.5 Genetic diversity within *Physalis* population based on *rbcL* marker

DNA divergence within the *Physalis rbcL* sequences was assessed by determining the number of polymorphic (segregating) sites (S), nucleotide diversity and total number of substitutions (Table 5.5). The nucleotide diversity was noted at 0.01333 among the *rbcL* sequences of the *Physalis* accessions.

Table 5.5. Polymorphism and divergence within (intraspecific) Physalis species based on rbcL sequences

Physalis species	*Physalis* species
Total number of sequences	56
Number of polymorphic (segregating) sites (S)	85
Nucleotide diversity Pi (Total)	0.01333
Theta (Total)	0.0.05959
Total number of substitutions	98

5.3.6 Genetic distance within *Physalis* population based on *rbcL* marker

The overall genetic distance was very low at 0.04±0.0. Average intraspecific distance of *Physalis* accessions was also determined as 0.04±0.0. The average intraspecific distance was found to be the same as the overall genetic distance within the *Physalis* accessions.

5.3.7 Nucleotide polymorphism

Eighty-five segregation sites (S) were identified within the *rbcL* genes of 56 *Physalis* accessions under this study (Table 5.6). The nucleotide diversity within these sequences was 0.01333. There were 66 singleton sites and 19 parsimony informative sites among the 85 identified segregating

sites (Table 5.6). The data form this section when compared to data for 10 Physalis accessions from one area as seen on table 3.4 indicated that when more Physalis accessions (56) from various areas are compared in their nucleotide diversity, some slight nucleotide diversity is observed as compared to fewer accessions from the same region where no diversity was observed.

Table 5.6. DNA polymorphism of *Physalis* accessions based on *rbcL* sequences

Polymorphic sites/ Segregation sites (S)	85	Positions in the gene	Variants
Singleton	66	129,130,133,143,225,233,236,253,274, 275,278,282,285,286,295,300,303,310, 311,312,324,327,329,332,336,337,341, 342,345,347,348,349,350,352,355,359, 367,368,375,377,383,391,392,395,398, 401,407,408,410,416,417,418,421,427, 438,439,443,455,462 339,361,380,388,393,397,457	2 3
Parsimony informative sites	19	304,333,338,343,357,364,378,415,432, 446,449,458,459 284,346,360,365,400,403	2 3
Nucleotide diversity (Pi)	0.01333		
Average number of nucleotide differences (k)	4.772		
Sequence length (base pairs)	730		
Number of sequences	56		

5.3.8 Tajima's neutrality test

Tajima's neutrality test was conducted on 56 *Physalis rbcL* sequences to determine the population selection and nucleotide diversity. The Tajima D value of *Physalis rbcL* sequences

was -2.533533. The negative Tajima D value indicates an excess of rare low frequency variations that are associated with population growth and positive selection pressure.

5.3.9 Barcoding gap analysis based on *rbcL* sequences

Automatic Barcode Gap Discovery results generated by distance model Kimura 80 parameter (K80P) based on ITS2 sequences for *Physalis* accessions indicate the presence of two barcoding gaps. The presence of a barcode gap indicates that the interspecific divergence of *rbcL* genes is higher than their intraspecific divergence. Based on the *rbcL* gene sequences, all pairwise distances were ranked by increasing distance values from 0.02 - 0.14 and two barcoding gaps identified (Figure 5.5). The barcode gaps were the same size and the first barcode gap was observed between distances of 0.12 (12%) and 0.13 (13%) (Figure 5.5). The second barcode gap was evident between the distances of 0.13 (13%) and 0.14 (14%) (Figure 5.5).

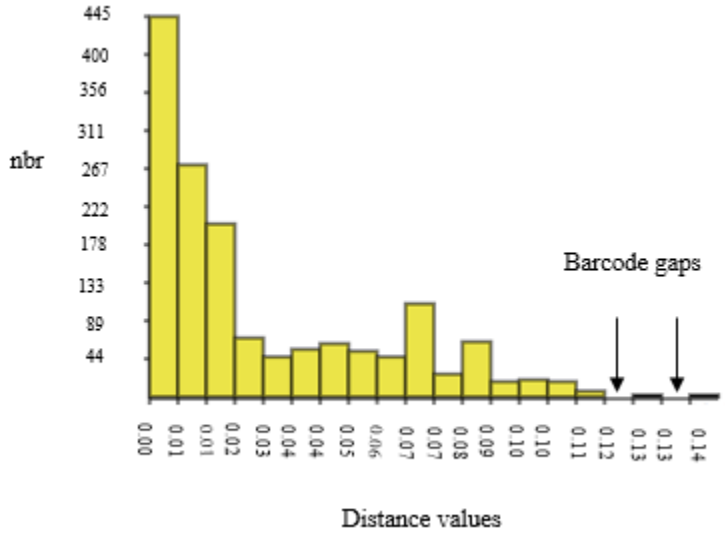


Figure 5.5. A histogram indicating the hypothetical distribution of pairwise differences of *rbcL* gene sequences of 56 *Physalis* accessions

Low divergence is presumably intraspecific divergence, whereas higher divergence indicates interspecific divergence. The abbreviation nbr on the *y*-axis of the histogram stands for the number of pairwise comparisons. A, B and C represent barcode gaps.

5.4 Discussion

Physalis species are relatively similar when it comes to their morphological characteristics. The *Physalis* genus plants also tend to have morphological similarities to the Solanaceae family plants such as *N. physalodes* (Feng *et al*., 2016). Due to these reasons, species identification by morphological characterization is not accurate for the *Physalis* genus. Molecular identification of plant species by DNA barcoding is accurate, reliable and rapid than the use of morphological features (Saddhe and Kumar, 2018). Studies on the identification of *Physalis* acessions in Kenya are very few with one key study focusing on the analysis of genetic diversity of *Physalis* accessions based on SSR markers, which did not discriminate the accessions into different *Physalis* species (Muraguri *et al*., 2021).

In the present study, *rbcL* barcode amplification and sequencing success rates were 82% and 88%, respectively. This high rate of success in amplification and sequencing of the *Physalis* accessions *rbcL* barcode was expected and concurred with results from other studies where *rbcL* barcode was used in identification of species (Kang *et al*., 2017; Huang *et al*., 2015). According to studies done by Kang *et al*. (2017) and Huang *et al*. (2015) the amplification and sequencing success rates for *rbcL* gene were between 75.26% and 63.84 and 97.6% and 90.8%, respectively. The high amplification and sequencing success rate of the *rbcL* gene is proposed to be due to the high conservation of the gene as a result of low frequency mutations (Kang *et al*., 2017).

In the phylogenetic analysis of *Physalis* accessions based on *rbcL* barcode, there was little information on the discrimination of species based on the constructed phylogenetic tree. Species discrimination was limited due to the inability of the *rbcL* barcodes to form clusters and clades with the reference database sequences based on BLASTn analysis. This is expected and has also been demonstrated in many other studies whereby species discrimination power of the *rbcL* barcode is usually the least among DNA barcoding markers (Kang *et al*., 2017; Huang *et al*., 2015; Tripathi *et al*., 2013; Hollingsworth *et al*., 2009). The identities provided for the different *Physalis* accessions on the phylogenetic tree are based on BLASTn analysis results as the phylogenetic tree was not able to efficiently discriminate the *Physalis* accessions into different *Physalis* species. A previous study by Hollingsworth *et al*. (2011) demonstrated that the low efficiency of *rbcL* in inter-specific divergence discrimination shows that the gene is not efficient

in species discrimination but is a good candidate in the identification of plants at the genus level (Hollingsworth *et al*., 2011).

Physalis accessions could not be categorized into various species based on *rbcL* barcode and therefore genetic diversity and distance analysis focused on intra-species variation. There was low nucleotide diversity at 0.0133 and low genetic distance at 0.4 ± 0.0 among the *rbcL* sequences of the *Physalis* accessions used in this study. This data supports the findings on the inability of *rbcL* sequences to discriminate the *Physalis* accessions based on phylogenetic analysis. The sequences were relatively similar with a very low genetic distance and diversity and hence could not cluster effectively into different species due to the high rate of conservation of the barcode. Genetic distance is an estimate of the genetic divergence between species or populations within a species (Beaumont *et al*., 1998). Low genetic distance of zero value observed in this study indicates that there is no genetic difference within species demonstrated that the accessions were of the same species. The very low genetic distance among the *rbcL* barcode sequences approaching zero is an indication that the gene was relatively conserved among these plants and could not be effectively used in their species discrimination. This has also been observed in other studies where *rbcL* barcode could not differentiate wild cinnamon and clove plants (Chandrasekara *et al*., 2021; Nurhasanah *et al*., 2019).

The Automatic Barcode Gap Discovery was able to show two small barcode gaps identified among the *Physalis* accessions based on *rbcL* barcode. This is an indication that the genetic divergence of *rbcL* barcode sequences of *Physalis* accessions was low and the *rbcL* barcode is relatively conserved in *Physalis*. The nucleotide polymorphism of *rbcL* barcode sequences of *Physalis* concurred with the Tajima neutrality test results which indicated a positive selection with a negative Tajima D value of -2.533533. Although 85 segregating sites were identified, the mutations were low frequency mutations associated with a positive selection as indicated by the negative Tajima D value. The *rbcL* barcode sequences of *Physalis* in this study had high singleton and parsimony informative sites among their 85 segregating sites. Sixty-six of the segregating sites were singleton mutations while 19 were parsimony sites. These are low frequency mutations and are associated with low nucleotide diversity based on the Tajima neutrality test (Carlson *et al*, 2005; Tajima, 1989).

5.5 Conclusion

The *rbcL* DNA barcode was effective in genus identification of the *Physalis* accessions but not discrimination of the different species. All 56 *Physalis* accessions used in this study were identified as the genus *Physalis*. The *rbcL* barcode sequences of the 56 *Physalis* accessions were highly conserved and phylogenetic tree analysis could not facilitate *Physalis* species discrimination. Low sequence divergence and low genetic distance among *rbcL* sequences of the *Physalis* accessions were also observed, indicating the *rbcL* barcode conserved among the accessions used in this study.

CHAPTER SIX

6.0 Comparative efficacy of ITS2 and *rbcL* DNA barcodes for identification and genetic diversity assessment of *Physalis* accessions in Kenya

6.1 Introduction

The genus *Physalis* has many species that grow in a wide array of habitats and ecologies, a common feature of the Solanaceae family (Ralte and Singh, 2021). This plant is native to the Andean region of South America with Colombia being the main producer and exporter (Cháves-Gómez *et al*., 2020). The economic value of *Physalis* in Colombia is linked to the high demand for fruits from mainly European countries (Álvarez-Flórez *et al*., 2017; Ordoñez *et al*., 2017). Other exporters of *Physalis* include Australia, New Zealand, Great Britain, Egypt, South Africa, Uganda, Zimbabwe, Kenya, Madagascar, and Southeast Asian countries (Barirega, 2014; Zhang *et al*., 2013; Ramadan and Moersel, 2003). *Physalis* are useful for income generation and have a wide range of nutritional and medicinal applications (Afroz *et al*., 2020; Barirega, 2014). Nutritionally, several *Physalis* species are rich in water- and fat-soluble vitamins (A, E, K, C and B-complex), minerals (magnesium, potassium, calcium and zinc), fatty acids (such as palmitate and linoleic acid), proteins and sugars (Puente *et al*., 2011; Ramadan and Moersel, 2003). The increased consumption of *Physalis* fruits has been associated with a decreased risk of chronic degenerative diseases (Reddy *et al*., 2010). The fruits are also rich in soluble solids, such as sugars like fructose, which are valuable for diabetic sugar control (Barirega, 2014). *P. peruviana* and *P. angulata* are rich in flavonoids, physaloids and other phytochemicals, and have been utilized in ethno-medicine. These phytochemicals have been applied in wound healing and the treatment of various ailments such as jaundice caused by hepatotoxicity, asthma, arthritis and hepatitis (Abdul-Nasir-Deen *et al*., 2020; Zhang and Tong, 2016; Arun and Asha, 2007). Phytochemicals like polyphenols have also contributed to the antioxidant, anti-inflammatory, antidiabetic, antihypertension and anticancer activities of *Physalis* crude extracts (Lan *et al*., 2009; Pinto *et al*., 2009; Franco *et al*., 2007; Wu *et al*., 2006). In addition, *Physalis ixocarpa*, commonly referred to as tomatillo, is a source of nutrients used in the preparation of sauces and salads (Shenstone *et al*., 2020). Due to the wide diversity of *Physalis* species and species-specific applications, there is a need to authenticate and discriminate the different *Physalis* species in

particular regions for efficient utilization, genetic resource conservation, and effective utilization in breeding programs (Feng *et al*., 2016).

The identification of *Physalis* species using morphological properties has resulted in misidentifications due to similarities in the phenotypic characteristics of the different species (Feng *et al*., 2016). For example, *P. minima* and *P. pubescens* are morphologically similar, which presents a challenge in their differentiation using their phenotypic characteristics (Feng *et al*., 2016). Morphological identification is also affected by the environmental/physiological factors, stage of growth and development of plants (Vargas-Ponce *et al*., 2011; Menzel, 1951). The misidentification of *Physalis* species can lead to losses of genetic information due to a lack of genetic conservation (Feng *et al*., 2018). Since the morphological identification of *Physalis* species has proven to be inefficient, there is a need to use robust and accurate means of species identification (Yu *et al*., 2021). Molecular identification is more accurate as it is based on unique nucleotide sequences that are not affected by the morphological characteristics of the species, the development stage (growth phase) or environmental/physiological factors (Schindel and Miller, 2005). To this end, DNA barcoding is one of the molecular techniques that can be used to identify and specify species accurately (Qian *et al*., 2022).

DNA barcoding is a rapid and reliable method of species identification and discrimination using short universal standardized DNA sequences (Saddhe and Kumar, 2018). It has been widely utilized and accepted in the identification of plants and animals as an effective taxonomic tool (Yu *et al*., 2021; Dormontt *et al*., 2018; Kress, 2005). Several DNA barcodes can be utilized in the identification of plants, based on the chloroplast-plastid (ribulose-1,5-bisphosphate carboxylase large (*rbcL*), maturase (*matK*), *psbA-trnH* among others) and nuclear ITS (internal transcriber spacer (ITS1) and (ITS2)) regions. However, factors such as universality, success in amplification and specificity variation need to be considered. These factors influence the efficiency of particular DNA barcodes in the identification and discrimination of plant species, and need to be taken into consideration in the selection of a DNA barcode (Li *et al*., 2021). *rbcL* is one of the universal barcode genes that is ideal for plant species discrimination studies, due to its high amplification and low mutation rate (Kang *et al*., 2017). The low level of mutation in the *rbcL* gene implies that it can be used in detailed studies on intra-species genetic and phylogenetic

variation (Nurhasanah *et al*., 2019). In addition, it is also a commonly used DNA barcode because it is conserved across a wide range of plant species (Manzara and Gruissem, 1988). Conversely, the nuclear DNA barcode, ITS2 gene is considered the best marker for DNA barcoding due to its high species discrimination power, inter- and intra-species level diversity, and high success rate in amplification and sequencing in plants (Kang *et al*., 2017). Therefore, this suggests the combination of chloroplast–plastids and nuclear regions as an efficient barcode tool to explore plant species discrimination (CBOL Plant Working Group *et al*., 2009).

According to the available literature, no DNA barcoding study has been conducted on the *Physalis* species present in Kenya. Similarly, no study has been conducted to assess the genetic diversity among *Physalis* accessions. The current study aimed at identifying the Kenyan *Physalis* species using *rbcL* and ITS2 barcodes and assessing the efficiency of the two candidate DNA barcodes to identify *Physalis* species. In addition, the phylogenetic relatedness of *Physalis* species was determined using *rbcL* and ITS2 sequences.

6.2 Materials and methods

6.2.1 Study area and collection of plant samples

Leaves of the genus *Physalis* were randomly collected from different locations of Kericho, Elgeyo-Marakwet, Homa Bay, Nakuru, Kajiado, Nyeri and Kiambu Counties of Kenya (Figure 6.1). The leaves were purposively sampled based on the availability, as most of the samples were wild plants growing without human intervention except for those collected from Elgeyo-Marakwet from a farmer. Within specific locations of sampling in the different counties, leaves and ripe fruits were collected and labeled after being placed in collection bags. The collected *Physalis* plant samples were identified by the taxonomist Mr. Patrick Mutiso and the samples were preserved in the University of Nairobi herbarium in the Department of Biology (Codes of Voucher Specimens: KP/UON2019/001- KP/UON2019/064). A Global Positioning System (GPS) device was used to record the location where the samples were collected in different counties; the altitude of the location of sampling was also noted and the assigned species name based on morphological appearance was also recorded (Table 6.1 and Appendix 1 A). Leaves of sixty-four (64) *Physalis* plants were collected between April and June 2019 in triplicate in zip-lock bags. Since it was difficult to identify the samples morphologically, each set of triplicate

plants was given a specific unique identification name based on the location at which they were collected, and a number (Appendix 1 A). Representative images of plants of some of the collected samples are presented in Figure 6.2. The samples were transported within 24 h post-sampling in a cool box with icepacks to the Department of Biochemistry at the University of Nairobi, and kept in the laboratory for genomic DNA extraction.

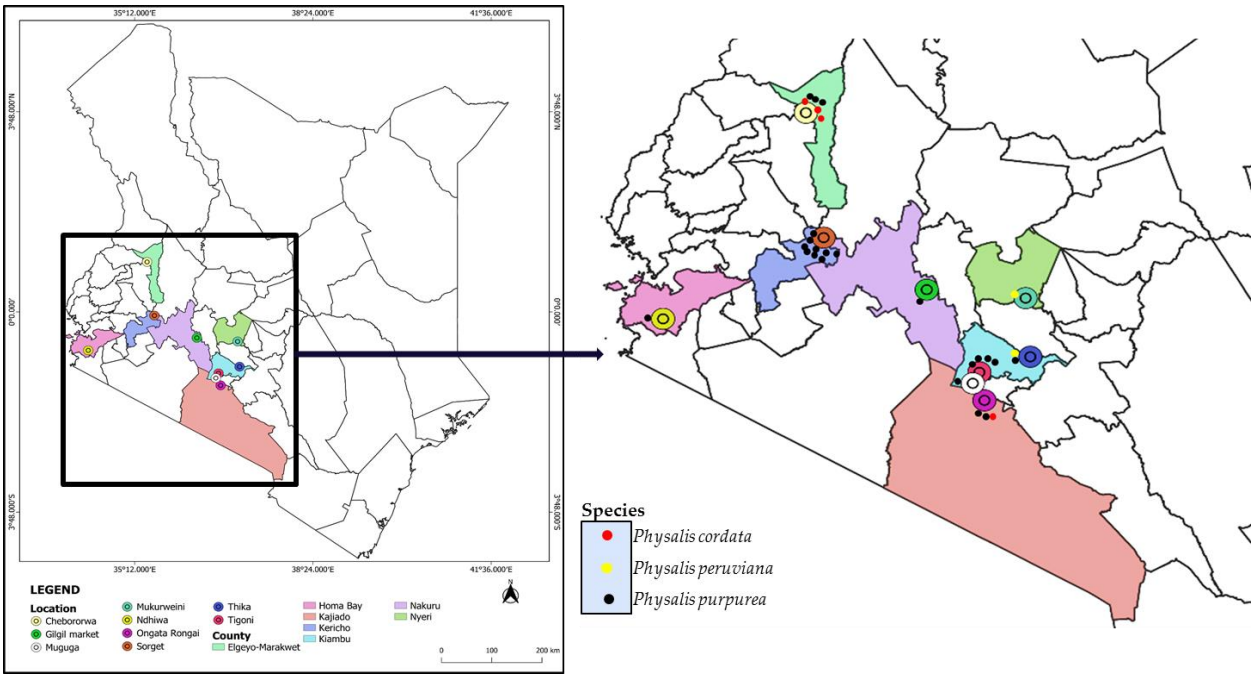


Figure 6.1. The locations of *Physalis* sampling sites in seven counties of Kenya and spatial distribution of the *Physalis* species discriminated based on ITS2 barcoding

Table 6.1. Geographical coordinates and number of samples picked from the various regions of *Physalis* sampling within seven different counties in Kenya

Serial No.	County	Location	Latitude	Longitude	No. of samples collected
1.	Kericho	Londiani - Sorget	0.0684° S	35.5548° E	10
2.	Elgeyo-Marakwet	Cheborорwa	0.9487° N	35.4234° E	13
3.	Homa Bay	Ndhiwa	0.7299° S	34.3671° E	3
4.	Nakuru	Gilgil market	0.4923° S	36.3173° E	1
5.	Kajiado	Ongata Rongai	1.3939° S	36.7442° E	5
6.	Nyeri	Mukurweini	0.5609° S	37.0488° E	5
7.	Kiambu	Tigoni	1.1651° S	36.7065° E	17
		Thika	1.0388° S	37.0834° E	9
		Muguga	1.2551° S	36.6580° E	1

Figure 6.2. Plant morphology of *Physalis* species

(**A**) *P. purpurea*—OQ372009.1; (**B**) *P. microcarpa*—OQ372018.1; (**C**) *P. purpurea*—OQ372013.1; (**D**) *P. purpurea*—OQ372019.1; (**E**) *P. purpurea*—OQ372020.1 and (**F**) *P. cordata*—OQ372012.1) in their natural habits. The *Physalis* species were discriminated based on their ITS2 barcode sequence. Morphological data for the *Physalis* accessions was not recorded. *Physalis* accessions were highly morphologically similar and environmental factors like soil salinity and pH, light exposure and carbon dioxide saturation can affect the physiological growth stages of the plants.

6.2.2 Molecular authentication of *Physalis* plants

6.2.2.1 Genomic DNA extraction

Physalis accessions genomic DNA was isolated, purified, viewed and stored as described in section 3.2.2.1.

6.2.2.2 PCR amplification and sequencing

Polymerase chain reaction (PCR) amplification was performed using ITS2 and *rbcL* DNA barcode markers presented in Table 3.1. PCR amplification, purification of PCR products and sequencing was performed as described in section 3.2.2.2.

6.2.2.3 Sequence alignment, phylogenetic and data analysis

The sequences of only 28 *Physalis* accessions that were successfully sequenced for both ITS2 and *rbcL* primers were used. The sequences of ITS2 and *rbcL* genes for *Physalis* accessions attained were curated, subjected to BLASTn analysis, used to prepare MSAs and phylogenetic trees as described in section 3.2.2.3. The MSA for the construction of MrBayes phylogenetic tree was prepared using the 28 *Physalis* ITS2 and *rbcL* gene sequences that had successfully been sequenced and BLASTn generated reference sequences for the two genes. Two MSAs were also prepared separately for ITS2 and *rbcL* sequences without use of their reference sequences and were used in the genetic diversity, nucleotide polymorphism, neutrality test, and automatic barcode gap discovery (ABGD) analysis.

6.2.3 Analysis of genetic divergence

The DNA divergence between *Physalis* accessions based on ITS2 sequences was assessed as described in section 4.2.4. DNA divergence within *Physalis* accessions based on ITS2 gene sequences was performed as described in section 3.2.3.

6.2.4 Determination of intraspecific and interspecific genetic distance

The intra- and interspecific genetic distances and overall mean distance of *Physalis* accessions based on the ITS2 sequences were determined as described in section 4.2.5. The intraspecific distance of *Physalis* accessions based on *rbcL* genes was determined as described in section 3.2.4.

6.2.5 Nucleotide polymorphism

Nucleotide polymorphism was assessed using the two MSAs for 28 *Physalis* accessions based on ITS2 and *rbcL* sequences as described in section 3.2.5.

6.2.6 Tajima's neutrality tests

Tajima's neutrality test for ITS2 and *rbcL* sequences was performed to assess the frequency of mutations among species and to determine selection in the populations (Tajima, 1989). Tajima's neutrality test was assessed using the MEGA 11.0 software (Tamura *et al*., 2021). The analysis involved 28 nucleotide sequences for the DNA barcode gene sequences analyzed. The codon positions included were the 1st + 2nd + 3rd + Noncoding for the *rbcL* gene sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option) in both analyses based on ITS2 and *rbcL* genes. There were totals of 532 and 716 positions for the ITS2 and *rbcL* genes, respectively, in the final dataset.

6.2.7 Barcoding gap analysis

The ITS2 and *rbcL* multiple sequence alignments for the 28 *Physalis* accessions were separately uploaded to the ABGD website (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html accessed on 20 February 2023) and distance analysis performed based on K80 Kimura measure of distance as described in section 3.2.7.

6.3 Results

6.3.1 Success rates of PCR amplification and sequencing

The success rates of PCR amplification for ITS2 and *rbcL* were 77% and 84%, respectively, while the sequencing success rates were high for *rbcL* (89%) and moderate for ITS2 (65%) (Table 6.2). The lengths of the ITS2 sequences were in the range of 237–707 bp, with an average of 525 bp and mean GC content of 61%, with a range of 55.7–66.9%. Similarly, the lengths of the *rbcL* sequences were in the range of 463–854 bp, with an average of 690 bp and mean GC content of 43.4%, with a range of 42.1–45.5% (Table 6.1, Appendix 1B and C).

Table 6.2. Efficiency of PCR amplification and sequencing for *Physalis* accessions for ITS2 and *rbcL* barcodes.

Barcode region	Samples tested (n)	Number of amplicons produced	Number of sequences produced	% amplification efficiency	% sequencing efficiency	Alignment length (bp)	Mean sequence length (bp)	Mean GC content (%)
ITS2	64	49	32	77	65	841	525	61.00
rbcL	64	54	48	84	89	841	690	43.40

The nucleotide base frequencies at different coding positions in *Physalis* accessions for ITS2 and *rbcL* sequences are indicated in Table 6.3. The percentage GC contents of ITS2 sequences were significantly higher than those of *rbcL* sequences for the *Physalis* accessions used in this study.

Table 6.3. The nucleotide base frequencies of candidate nucleotide sequences at different coding positions in *Physalis* accessions.

Barcode Locus	Base Contents (%)					
	A	T	G	C	AT	GC
ITS2	19.42	19.39	29.78	31.41	39.00	61.00
rbcL	28.22	28.40	23.10	20.28	56.58	43.42

6.3.2 Species discrimination of *Physalis* accessions using BLASTn analysis

Species discrimination used a similarity-based approach based on BLASTn. The results show a high similarity of ITS2 and *rbcL* with other sequences in the GenBank by BLASTn sequence similarity searches. The percentage identity based on ITS2 loci ranged from 80.36 to 97.41%, and the *Physalis* species identified were *P. cordata*, *P. peruviana*, *P. microcarpa*, *P. aff. philadelphica*, *P. minimaculata* and *P. purpurea* (Table 6.4). None of the *Physalis* accessions had 100% identity based on ITS2 sequences for the BLASTn analysis. BLASTn analysis of the *rbcL* sequences identified that all 28 *Physalis* accessions belonged to the genus *Physalis*. Out of the 28 *Physalis* accessions, 7 had 100% identity as *P. minima*, while the rest had percentage identities ranging from 91.10 to 99.86 and were identified as *P. peruviana*, *P. virginiana*, *P. angulata* and *P. minima* (Table 6.4).

Table 6.4. BLASTn analysis results for 28 *Physalis* accessions based on ITS2 and *rbcL* barcode gene

Sample ID	*rbcL*					ITS2				
	Species of best BLAST match	GenBank Accession number (of database)	E-value	Percent identity (%)	Accession number	Species of best BLAST match	GenBank Accession Number (of database)	E-value	Percent identity (%)	Accession Number
Chebororwa C2	*P. peruviana*	NC_026570.1	0.0	94.90	OQ507163.1	*P. purpurea*	MH763740.1	0.0	94.70	OQ371996.1
Chebororwa C5	*P. virginiana*	KT178121	0.0	95.46	OQ507165.1	*P. cordata*	AY665886.1	5e-122	87.78	OQ371997.1
Chebororwa e1in	*P. minima*	NC_048515.1	0.0	93.18	OQ507166.1	*P. cordata*	MH763728.1	1e-81	82.24	OQ371998.1
Chebororwa e4in	*P. minima*	NC_048515.1	0.0	99.46	OQ507167.1	*P. cordata*	AY665886.1	1e-108	86.06	OQ372001.1
Chebororwa e6in	*P. minima*	NC_048515.1	0.0	98.28	OQ507168.1	*P. purpurea*	MH763740.1	0.0	93.75	OQ372003.1
Chebororwa e7in	*P. minima*	NC_048515.1	0.0	99.19	OQ507169.1	*P. purpurea*	MH763740.1	0.0	94.27	OQ372004.1
Gilgil gm	*P. minima*	NC_048515.1	0.0	99.06	OQ507171.1	*P. purpurea*	MH763740.1	5e-35	81.98	OQ372005.1
Muguga DM	*P. minima*	NC_048515.1	0.0	98.74	OQ507177.1	*P. purpurea*	MH763740.1	0.0	93.83	OQ372007.1
Mukurweini ny5	*P. angulata*	NC_039457.1	0.0	92.90	OQ507180.1	*P. peruviana*	AY665914.1	4e-93	88.82	OQ372008.1
Ndhiwa hb2	*P. minima*	NC_048515.1	0.0	93.33	OQ507181.1	*P. purpurea*	MH763740.1	0.0	94.28	OQ372009.1
Ongata Rongai nor3	*P. peruviana*	NC_026570.1	0.0	95.47	OQ507183.1	*P. cordata*	AY665886.1	2e-76	86	OQ372012.1
Ongata Rongai nor4	*P. virginiana*	KT178121.1	0.0	91.10	OQ507184.1	*P. purpurea*	MH763740.1	0.0	94.09	OQ372013.1

Ongata Rongai nor5	*P. peruviana*	NC_026570.1	0.0	98.78	OQ507185.1	*P. purpurea*	MH763740.1	0.0	94.51	OQ372014.1
Thika TK8	*P. minima*	NC_048515.1	0.0	99.04	OQ507192.1	*P. minimaculata*	AY665905.1	2e-93	88.51	OQ372015.1
ThikaTK9	*P. minima*	NC_048515.1	0.0	98.03	OQ507193.1	*P. peruviana*	AY665914.1	1e-35	80.36	OQ372016.1
Tigoni T2	*P. minima*	NC_048515.1	0.0	98.31	OQ507194.1	*P. purpurea*	MH763740.1	7e-160	91.79	OQ372017.1
Tigoni T4	*P.s minima*	NC_048515.1	0.0	96.42	OQ507195.1	*P. microcarpa*	AY665903.1	1e-61	86.78	OQ372018.1
Tigoni T9	*P. minima*	NC_048515.1	0.0	98.13	OQ507199.1	*P. purpurea*	MH763740.1	0.0	92.22	OQ372019.1
Tigoni T11	*P. minima*	NC_048515.1	0.0	99.32	OQ507201.1	*P. purpurea*	MH763740.1	9e-32	84.85	OQ372020.1
Londiani 2	*P. mimima*	NC_048515.1	0.0	100	OQ507153.1	*P. purpurea*	MH763740.1	0.0	92.98	OQ372021.1
Londiani 3	*P. minima*	NC_048515.1	0.0	99.59	OQ507154.1	*P. purpurea*	MH763740.1	0.0	94.06	OQ372022.1
Londiani 4	*P. minima*	NC_048515.1	0.0	100	OQ507155.1	*P. peruviana*	AY665914.1	0.0	97.41	OQ372023.1
Londiani 5	*P. minima*	NC_048515.1	0.0	99.86	OQ507156.1	*P. purpurea*	MH763740.1	0.0	94.45	OQ372024.1
Londiani 6	*P. minima*	NC_048515.1	0.0	100	OQ507157.1	*P. aff. philadelphica*	AY665868.1	2e-144	91.35	OQ372025.1
Londiani 7	*P. minima*	NC_048515.1	0.0	100	OQ507158.1	*P. purpurea*	MH763740.1	1e-148	86.00	OQ372026.1
Londiani 8	*P. minima*	NC_048515.1	0.0	100	OQ507159.1	*P. purpurea*	MH763740.1	0.0	93.07	OQ372027.1
Londiani 9	*P. minima*	NC_048515.1	0.0	100	OQ507160.1	*P. purpurea*	MH763740.1	0.0	88.96	OQ372028.1
Londiani 10	*P. minima*	NC_048515.1	0.0	100	OQ507161.1	*P. purpurea*	MH763740.1	0.0	90.16	OQ372029.1

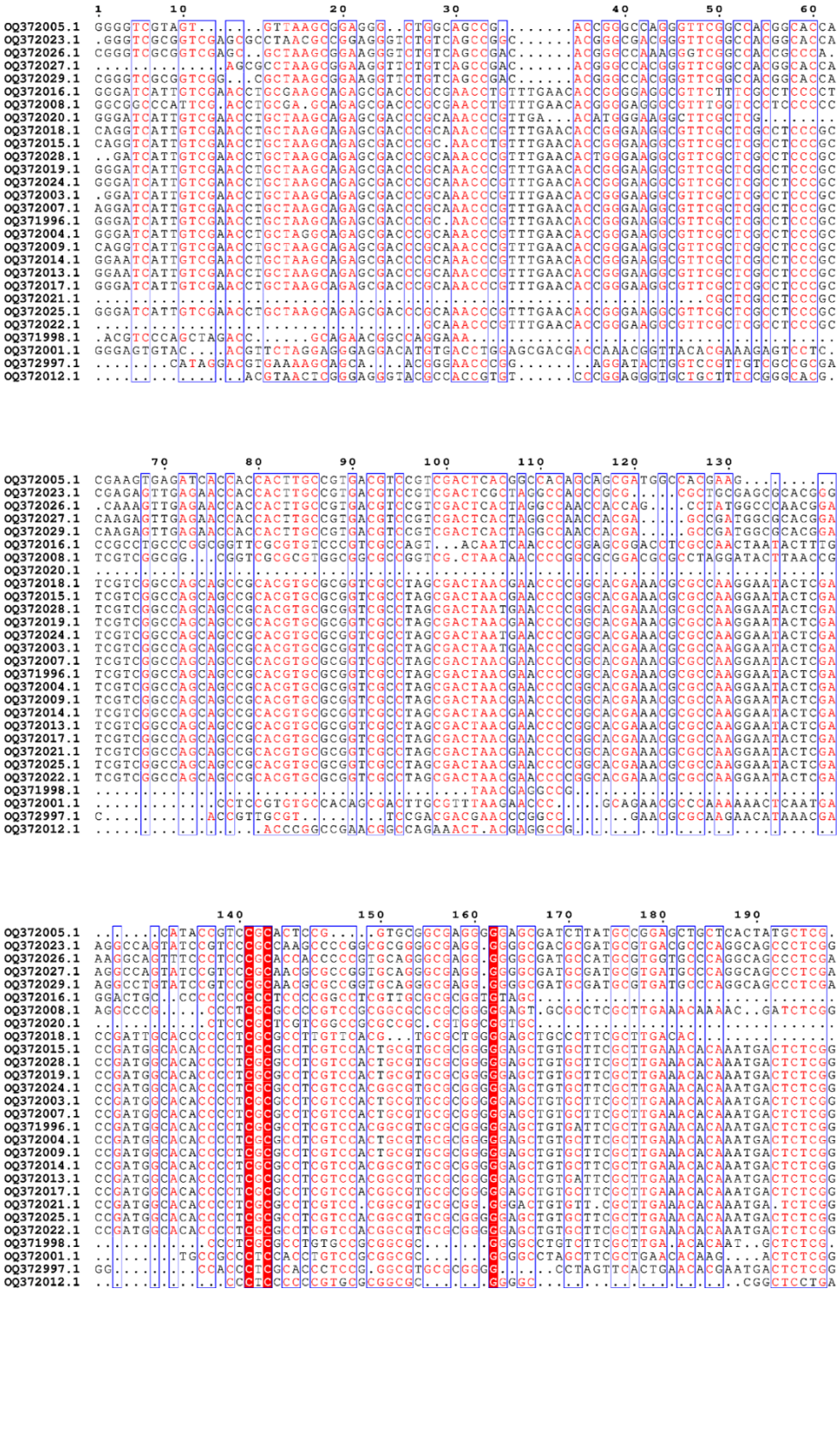
6.3.3 Multiple sequence alignments

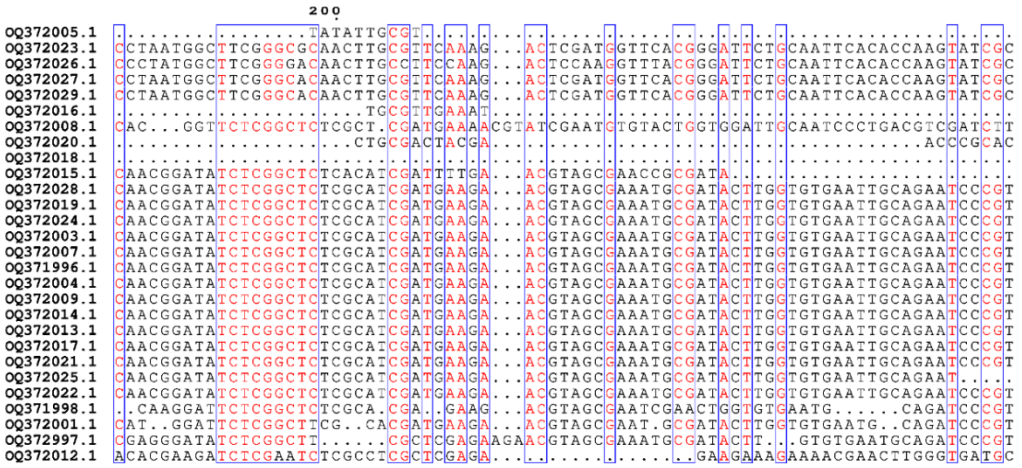
The multiple sequence alignment (MSA) of cleaned ITS2 and *rbcL* sequences as well as their reference sequences from BLASTn analysis was prepared based on MUSCLE had a sequence length of 841 bp. The multiple sequence alignment was compressed using ESPript 3 (Appendix 2D) (https://espript.ibcp.fr/ESPript/temp/1032964064/0-0-1680466160-esp.pdf accessed on 15 February 2023). This MSA had a high rate of nucleotide substitutions, deletions and insertions among and between *Physalis* species based on the ITS2 marker (Appendix 2D). The MSA also shows a high rate of nucleotide sequence conservation among and between the *Physalis* species based on the *rbcL* marker with very few deletions, insertions and substitutions. Substitution transition mutations can be noted at positions 304 and 368 (Appendix 2D). At position 304, we see a transition substitution mutation for the *Physalis* accession OQ507184.1 whereby this sequence has an adenine, but all other *rbcL* sequences and reference sequences have a guanine. At position 368, there is another substitution point mutation for *Physalis* accession OQ507166.1, whereby guanine replaces an adenine base. A transversion point mutation is also noted at position 305 for the *Physalis* accession OQ507184.1, whereby a guanine replaces cytosine (Appendix 2D).

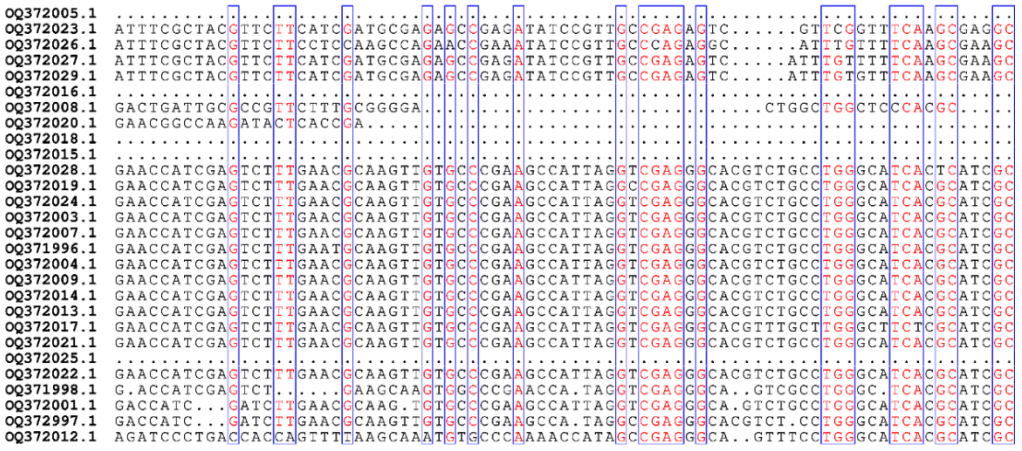
Other transversion point mutations are noted at positions 369 and 419 of the *Physalis* accession OQ507166.1, whereby adenine replaces thymine in both cases (Appendix 2D). An insertion macro-lesion is noted between positions 579 and 580 for *Physalis* accession OQ507166.1, whereby five nucleotides are inserted (Appendix 2D). A deletion macro-lesion is noted for *Physalis* accession OQ507184.1 between positions 530 and 536 whereby seven nucleotides are deleted (Appendix 2D).

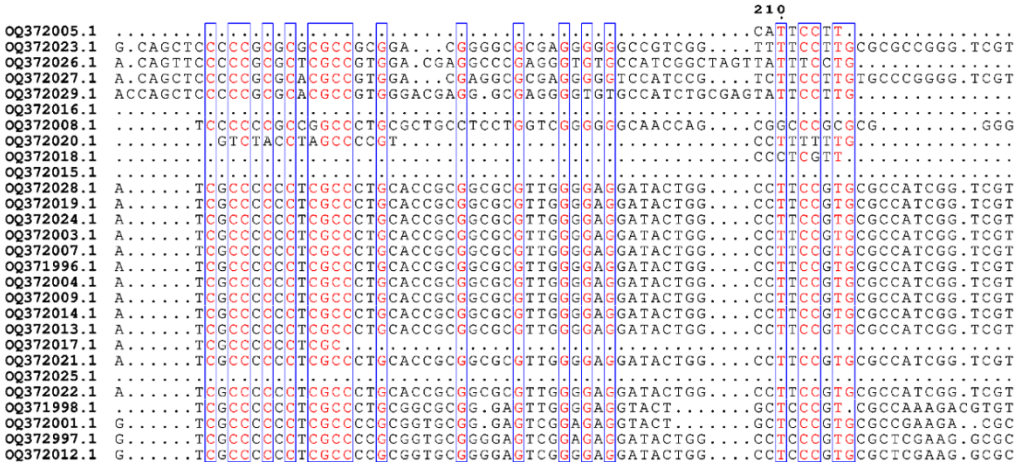
The MSA of the 28 ITS2 sequences based on MUSCLE, trimmed and edited by Jalview version 1.11.2.0, had a sequence length of 532 bp. It was compressed using ESPript (Figure 6.3) (https://espript.ibcp.fr/ESPript/temp/1440398212/0-0-1688383904-esp.pdf accessed on 15 February 2023). This MSA has many substitutions, deletion and insertion mutations (Figure 6.3). The substitution mutations in this MSA are composed of transition and transversion point mutations (Figure 6.3). The MSA of 28 *rbcL* sequences based on MUSCLE, trimmed and edited

by Jalview, had a sequence length of 716 bp. It was compressed using ESPript (Figure 6.4) (https://espript.ibcp.fr/ESPript/temp/1848737578/0-0-1688384397-esp.pdf accessed on 15 February 2023). This MSA is relatively conserved and does not have any insertion or deletion mutations, but it has quite a high number of substitution point mutations; for example, at positions 40, 170, 171, 180, 181, and many others (Figure 6.4). The substitution mutations are composed of transition and transversion point mutations Figure 6.4). The sequence alignments reveal a wide dispersal of sequence similarity for ITS2 sequences and homologous sequences for *rbcL* sequences among the tested *Physalis* accessions.




200
OQ372005.1
OQ372023.1
OQ372026.1
OQ372027.1
OQ372029.1
OQ372016.1
OQ372008.1
OQ372020.1
OQ372018.1
OQ372015.1
OQ372028.1
OQ372019.1
OQ372024.1
OQ372003.1
OQ372007.1
OQ371996.1
OQ372004.1
OQ372009.1
OQ372014.1
OQ372013.1
OQ372017.1
OQ372021.1
OQ372025.1
OQ372022.1
OQ371998.1
OQ372001.1
OQ372997.1
OQ372012.1


OQ372005.1
OQ372023.1
OQ372026.1
OQ372027.1
OQ372029.1
OQ372016.1
OQ372008.1
OQ372020.1
OQ372018.1
OQ372015.1
OQ372028.1
OQ372019.1
OQ372024.1
OQ372003.1
OQ372007.1
OQ371996.1
OQ372004.1
OQ372009.1
OQ372014.1
OQ372013.1
OQ372017.1
OQ372021.1
OQ372025.1
OQ372022.1
OQ371998.1
OQ372001.1
OQ372997.1
OQ372012.1


210
OQ372005.1
OQ372023.1
OQ372026.1
OQ372027.1
OQ372029.1
OQ372016.1
OQ372008.1
OQ372020.1
OQ372018.1
OQ372015.1
OQ372028.1
OQ372019.1
OQ372024.1
OQ372003.1
OQ372007.1
OQ371996.1
OQ372004.1
OQ372009.1
OQ372014.1
OQ372013.1
OQ372017.1
OQ372021.1
OQ372025.1
OQ372022.1
OQ371998.1
OQ372001.1
OQ372997.1
OQ372012.1

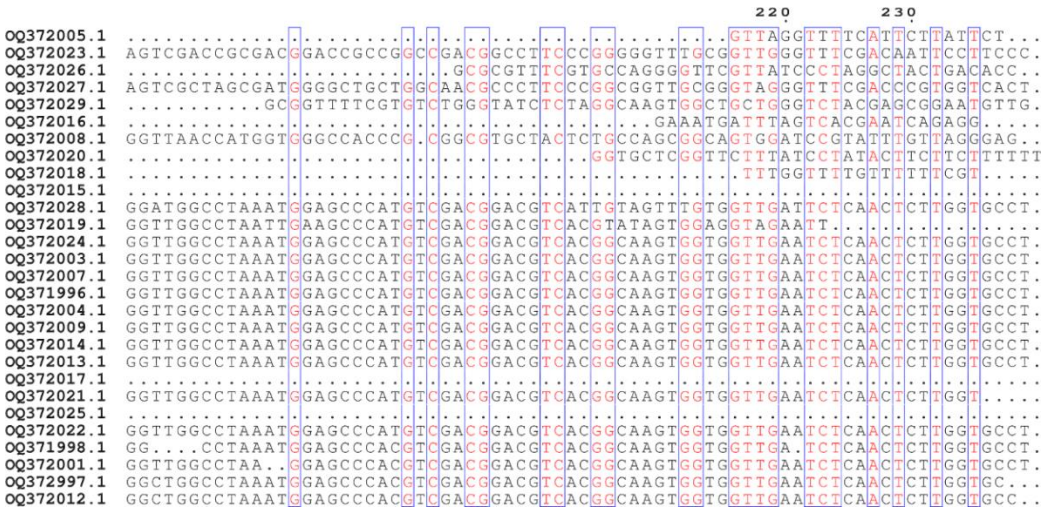
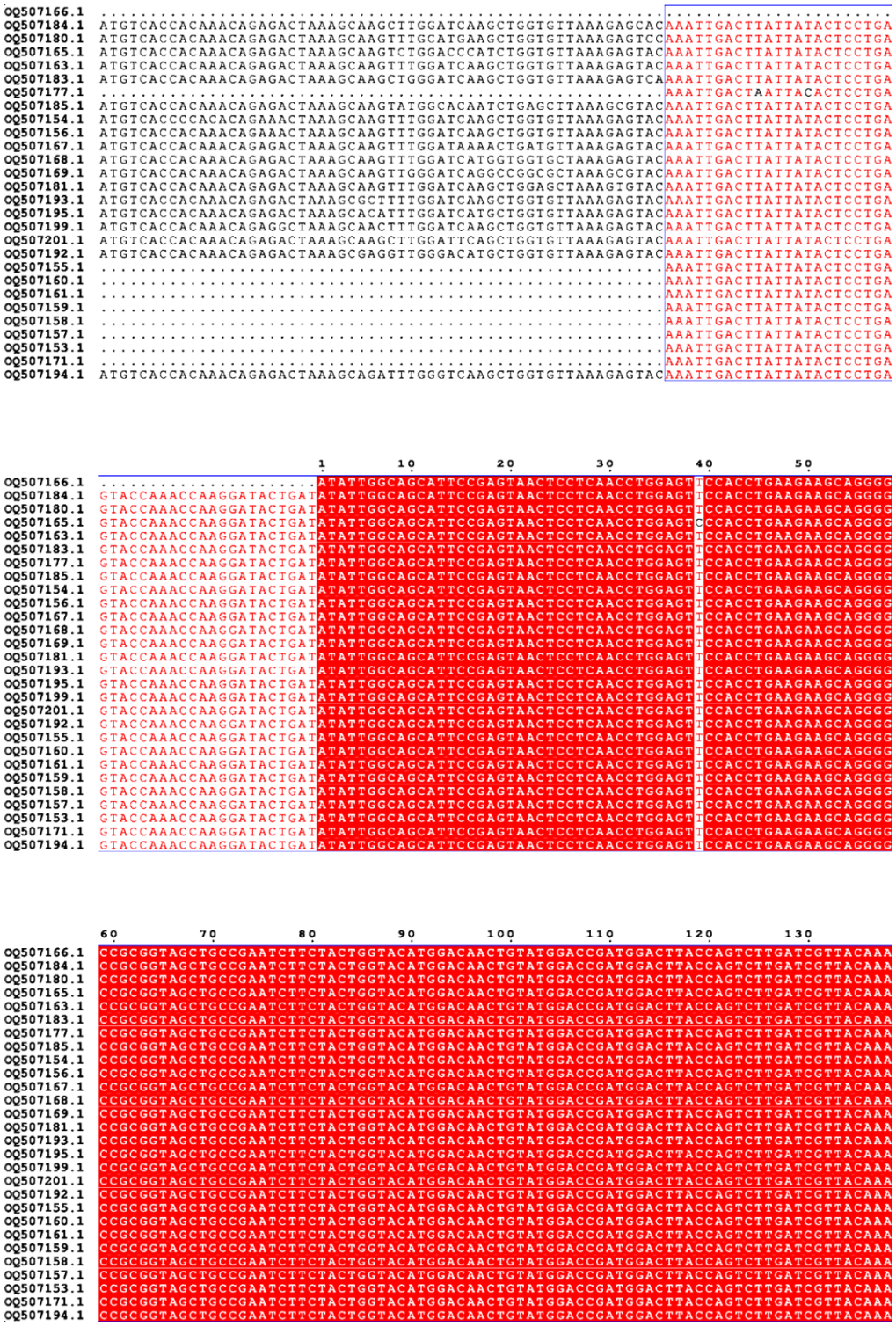
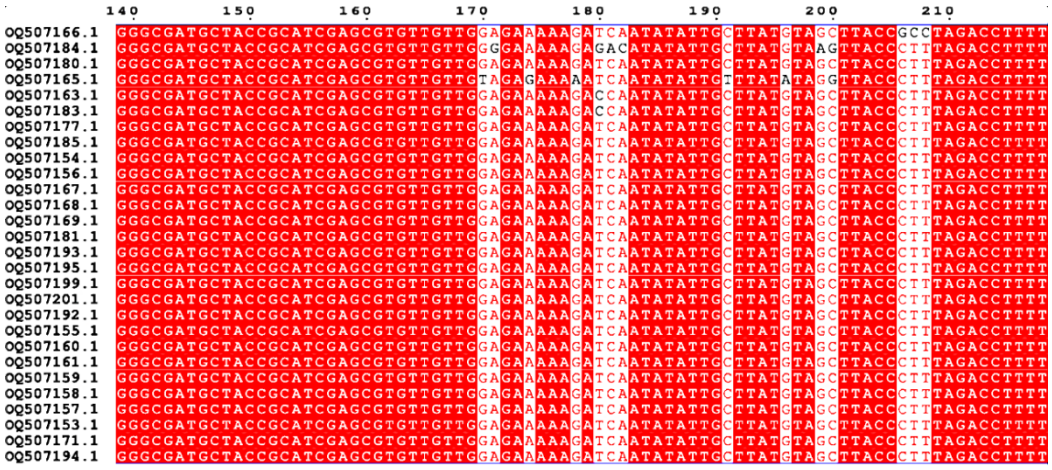
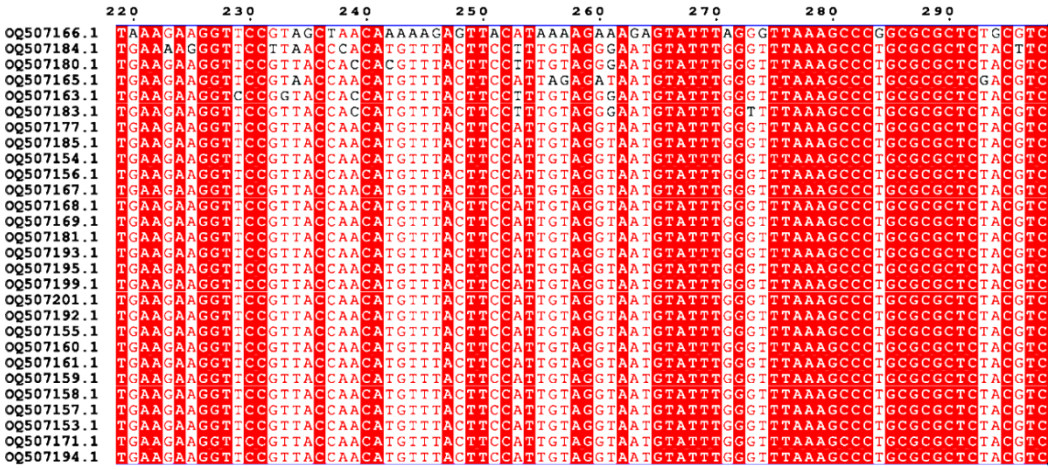
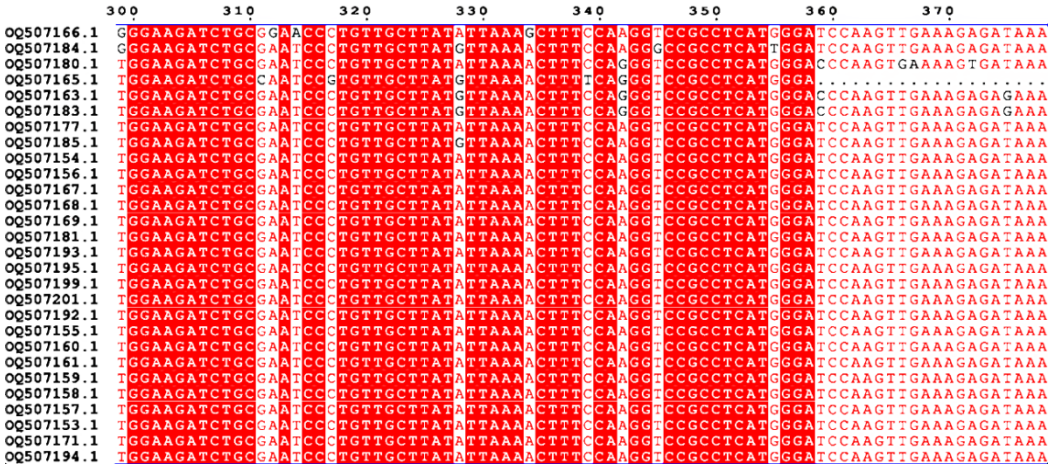


Figure 6.3. Multiple sequence alignment for 28 *Physalis* accessions based on ITS2 marker.




140 150 160 170 180 190 200 210
OQ507166.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCGCCTAGACCTTTT
OQ507184.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGGGAAAAAGACATATATTGCTTATGTAAGTTACCCTTTAGACCTTTT
OQ507180.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507165.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGTAGAGAAAAATCAATATATTGTTTATATAGGTTACCCTTTAGACCTTTT
OQ507163.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGACCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507183.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGACCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507177.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507185.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507154.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507156.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507167.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507168.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507169.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507181.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507193.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507195.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507199.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507201.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507192.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507155.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507160.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507161.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507159.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507158.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507157.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507153.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507171.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT
OQ507194.1 GGGCGATGCTACCGCATCGAGCGTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGGTTACCCTTTAGACCTTTT


220 230 240 250 260 270 280 290
OQ507166.1 TAAAGAAGGTTCCGTAGCTAACAAAAAGAGTTACATAAAAGAAGAGTATTTAGGGTTAAAGCCCGGCGCGCTCTGCGTC
OQ507184.1 TGAAAAGGGTTCCTTAACCCACATGTTTACTTCCTTTGTAGGGAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACTTC
OQ507180.1 TGAAGAAGGTTCCGTTACCACCACGTTTACTTCCTTTGTAGGGAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507165.1 TGAAGAAGGTTCCGTAACCAACATGTTTACTTCCATTAGAGATAATGTATTTGGGTTTAAAGCCCTGCGCGCTCGACGTC
OQ507163.1 TGAAGAAGGTCCGGTACCACCATGTTTACTTCCTTTGTAGGGAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507183.1 TGAAGAAGGTTCCGTTACCACCATGTTTACTTCCTTTGTAGGGAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507177.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507185.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507154.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507156.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507167.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507168.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507169.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507181.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507193.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507195.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507199.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507201.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507192.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507155.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507160.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507161.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507159.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507158.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507157.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507153.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507171.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC
OQ507194.1 TGAAGAAGGTTCCGTTACCAACATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTTAAAGCCCTGCGCGCTCTACGTC


300 310 320 330 340 350 360 370
OQ507166.1 GGGAAGATCTGCGGAACCCTGTTGCTTATATTAAAGGTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507184.1 GGGAAGATCTGCGAATCCCTGTTGCTTATGTTAAAACTTTCCAAGGCCGCCTCATTGGATCCAAGTTGAAAGAGATAAA
OQ507180.1 TGGAAGATCTGCGAATCCCTGTTGCTTATGTTAAAACTTTCCAGGGTCCGCCTCATGGGACCCAAGTGAAAAGTGATAAA
OQ507165.1 TGGAAGATCTGCCAATCCCTGTTGCTTATGTTAAAACTTTTCAGGGTCCGCCTCATGGGA.....................
OQ507163.1 TGGAAGATCTGCGAATCCCTGTTGCTTATGTTAAAACTTTCCAGGGTCCGCCTCATGGGACCCAAGTTGAAAGAGAGAAA
OQ507183.1 TGGAAGATCTGCGAATCCCTGTTGCTTATGTTAAAACTTTCCAGGGTCCGCCTCATGGGACCCAAGTTGAAAGAGAGAAA
OQ507177.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507185.1 TGGAAGATCTGCGAATCCCTGTTGCTTATGTTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507154.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507156.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507167.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507168.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507169.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507181.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507193.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507195.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507199.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507201.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507192.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507155.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507160.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507161.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507159.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507158.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507157.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507153.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507171.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA
OQ507194.1 TGGAAGATCTGCGAATCCCTGTTGCTTATATTAAAACTTTCCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAA

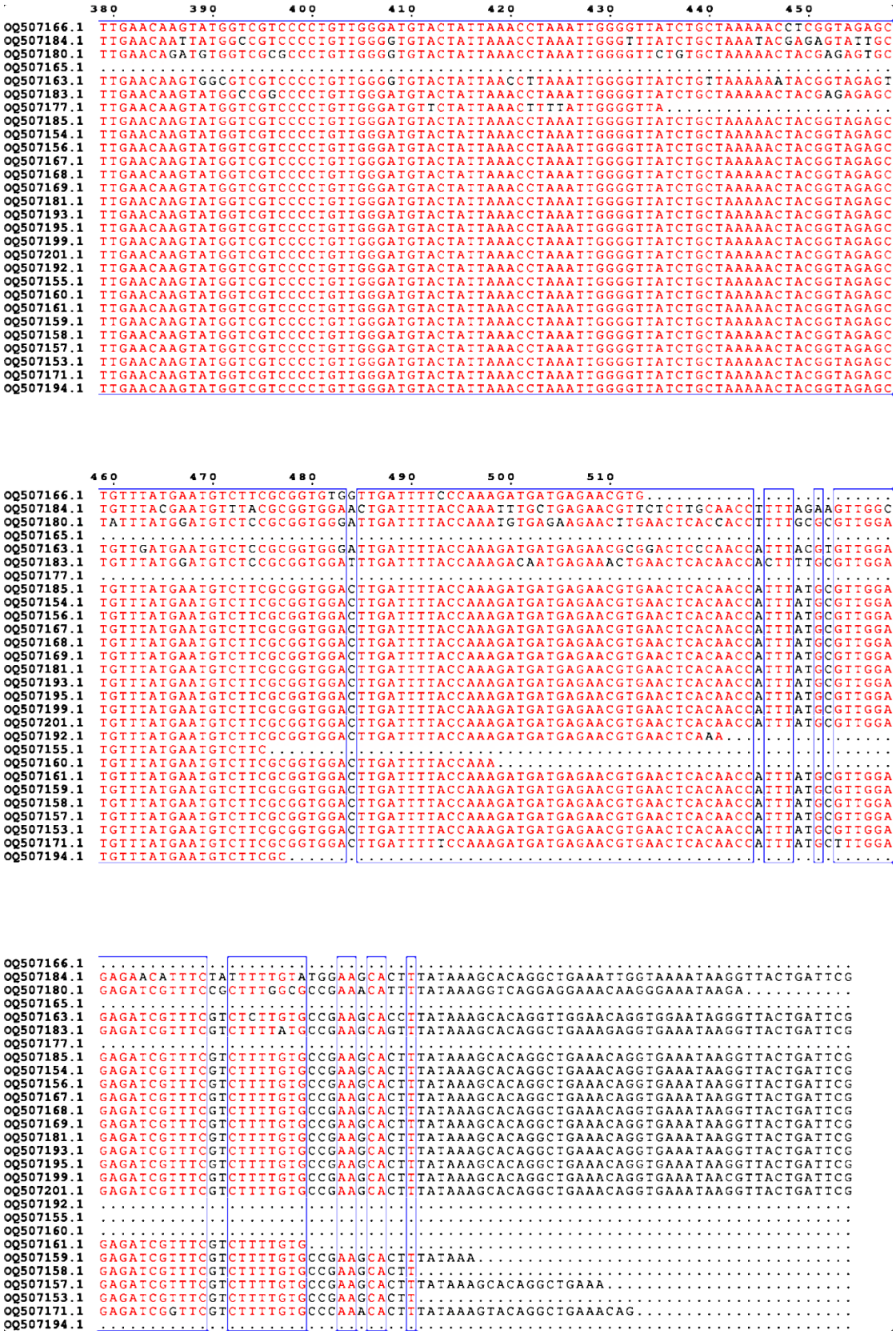


Figure 6.4. Multiple sequence alignment for 28 *Physalis* accessions based on *rbcL* markers.

6.3.4 Species discrimnation of *Physalis* species based on phylogenetic analysis

A phylogenetic tree constructed using combined ITS2 and *rbcL* sequences yielded two major clusters from the BI phylogeny that were robust with 100% posterior probability values (Figure 6.5), with each of the clusters separated based on the ITS2 and *rbcL* nucleotide data matrix. The combination of the two genes on the same phylogenetic tree gave a clear perspective on the *Physalis* species discrimination efficiency of each gene. Species discrimination was only possible with the ITS2 marker, while the discriminatory power of *rbcL* was low and inefficient. The *rbcL* region showed the lowest level of genetic differentiation, with the species samples *P. minima*, *P. peruviana*, *P. angulata* and *P. virginiana* forming a distinct cluster (Figure 6.5). The nucleotide data matrix from *rbcL* reflects the close genetic relationships of these species (Figure 6.5). The nucleotide data matrix of ITS2 splits the *Physalis* accessions into three clades representing three *Physalis* species, namely, *P. cordata* (OQ5372012.1, OQ371998.1, OQ372001.1 and OQ371998.1), *P. peruviana* (OQ372016.1 and OQ372008.1) and *P. purpurea* (OQ371996.1, OQ372003.1, OQ372004.1, OQ372005.1, OQ372007.1, OQ372009.1, OQ372013.1, OQ372014.1, OQ372015.1, OQ372017.1, OQ372018.1- OQ372029) (Figure 6.5).

The clades formed by the ITS2 sequences show longer branch lengths among the *P. peruviana* species, with a posterior probability percentage of 94. The *P. cordata* species is associated with moderate branch lengths on the phylogenetic tree, with a posterior probability percentage of 89. The shortest branch lengths among the ITS2 sequences on the phylogenetic tree are those associated with *P. purpurea*, with a posterior probability percentage of 66.

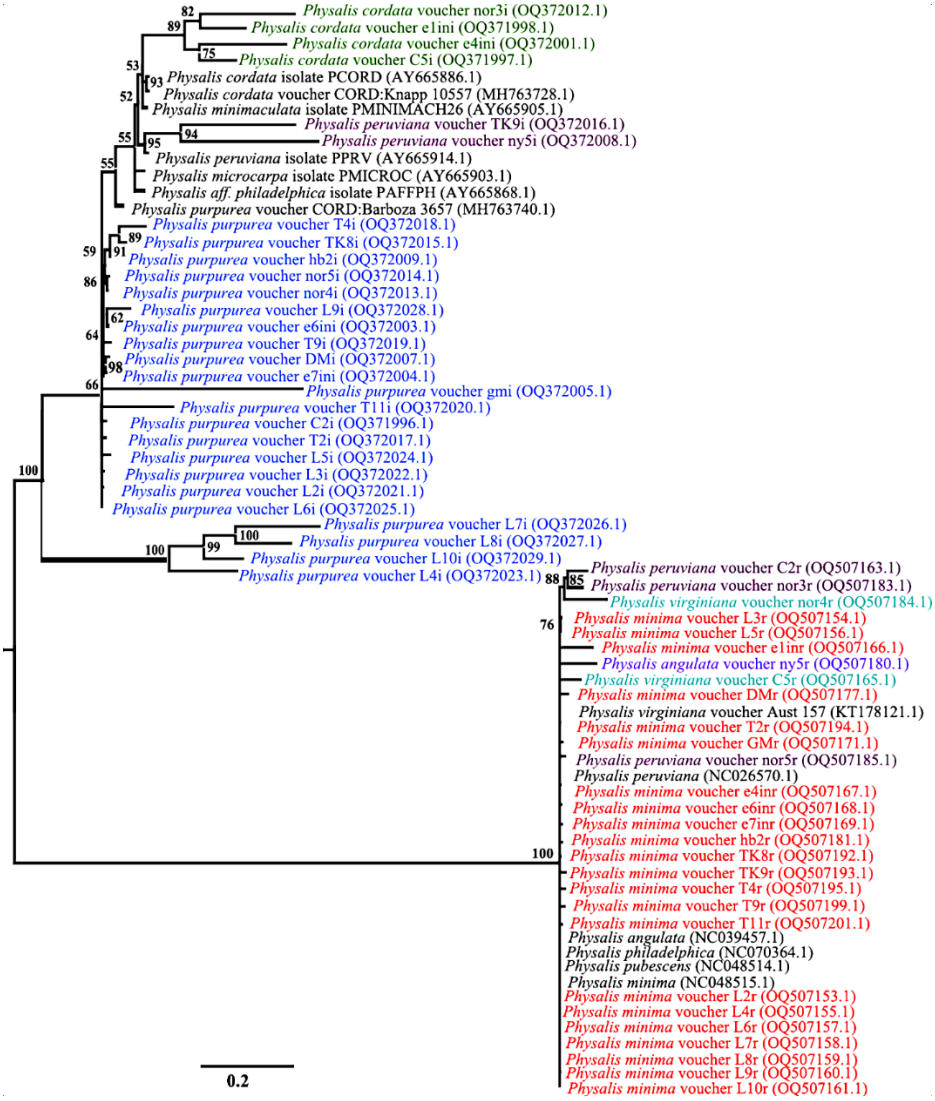


Figure 6.5. Consensus Phylogenetic tree prepared by MrBayes for *Physalis* accessions based on a combination of ITS2 and *rbcL* DNA barcodes

Black represents different *Physalis* species reference sequences retrieved from GenBank after BLASTn analysis, green represents *P. cordata*, plum represents *P. peruviana*, blue represents *P. purpurea*, teal represents *P. virginiana*, purple represents *P. angulata* and orange represents *P. minima*. Numbers above branches indicate the posterior probability percentage statistic for the MrBayes phylogenetic tree. The length of the branches indicates the genetic divergence of the *Physalis* accessions as compared to their counterparts.

6.3.5 Genetic divergence analysis between and within *Physalis* species based on ITS2 sequences

The ITS2 sequence was the only barcode that could be used to differentiate the accessions into *Physalis* species which is necessary for interspecific genetic divergence analysis (Figure 6.5). The *rbcL* genes could not discriminate *Physalis* species and hence it was impossible to assess for interspecific genetic divergence based on this gene (Figure 6.5).

6.3.5.1 DNA divergence between populations based on ITS2 sequences

Varying shared mutations were observed between the *Physalis* populations (Table 6.5). The nucleotide diversity was highest (0.33208) between *P. peruviana* and *P. cordata* and the lowest (0.14821) between *P. cordata* and *P. purpurea*. The average number of nucleotide substitutions per site between populations ranged from 0.24621 to 0.38915. The number of net nucleotide substitutions per site between nucleotides ranged from 0.01299 to 0.12343 (Table 6.5). The total number of fixed (base) differences between populations was: six for *P. peruviana* and *P. cordata*, one for *P. peruviana* and *P. purpurea*, and zero for *P. cordata* and *P. purpurea* (Table 6.5). The number of fixed differences was determined from the total polymorphic sites between populations, and it was observed that the higher the number of polymorphic differences between populations was, the higher the fixed difference would be, and vice versa (Table 6.5).

Table 6.5. DNA divergence between (interspecific) *Physalis* species populations based on ITS2 sequences.

Population	***P. peruviana* (P1)**	***P. cordata* (P2)**	***P. peruviana* (P1)**	***P. purpurea* (P2)**	***P. purpurea* (P1)**	***P. cordata* (P2)**
Polymorphic sites in each population	14	21	12	18	2	4
Total number of polymorphic sites	35		23		4	
Average number of nucleotide differences	17.600		6.351		0.889	
Nucleotide diversity Pi (t)	0.33208		0.18147		0.14821	
Number of fixed differences	6		1		0	
Polymorphic mutations in population 1 (P1) but monomorphic ones in population 2 (P2)	13		10		2	
Polymorphic mutations in P2 but monomorphic ones in P1	28		22		1	
Shared mutations	1		2		2	
Average number of nucleotide differences between populations	20.625		10.158		1.477	
Average nucleotide substitution per site between populations (Dxy)	0.38915		0.29026		0.24621	
Number of net nucleotide substitutions per site between populations (Da)	0.12343		0.03881		0.01299	

6.3.5.2 DNA Divergence within Populations Based on ITS2 Sequences

DNA divergence within each *Physalis* species was assessed using ITS2 sequences by determining the number of polymorphic (segregating) sites (S), the nucleotide diversity and the total number of substitutions (Table 6.6). The *rbcL* marker was not able to facilitate species discrimination and hence the analysis for genetic divergence was only performed on the ITS 2 gene sequences. The nucleotide diversity was highest (0.31250) and lowest (0.14898) within *P. peruviana* and *P. purpurea*, respectively (Table 6.6). The highest (101) and the lowest (26) total numbers of nucleotide substitutions were observed in *P. cordata* and *P. purpurea*, respectively. The numbers of polymorphic segregating sites were highest (83) and lowest (20) within *P. cordata* and *P. purpurea*, respectively (Table 6.6).

Table 6.6. Polymorphism and divergence within (intraspecific) *Physalis* species based on ITS2 sequences.

***Physalis* Species**	***P. peruviana***	***P. cordata***	***P. purpurea***
Total number of sequences	2	4	22
Number of polymorphic (segregating) sites (S)	70	83	20
Nucleotide diversity Pi (Total)	0.31250	0.18095	0.14898
Nucleotide diversity Pi (JC-Total)	0.40425	0.20708	0.16609
Theta (Total)	0.31250	0.19675	0.17396
Total number of substitutions	70	101	26

6.3.6. Genetic distance between and within *Physalis* species based on ITS2 and *rbcL* sequences

The average inter-specific distance between *Physalis* species was determined based on the ITS2 gene sequences only because the *rbcL* marker was not able to facilitate species discrimination. The analysis showed that the highest mean genetic distance (1589.41) was between *P. purpurea* and *P. cordata* (Table 6.7). The lowest mean genetic distance (9.53) was between *P. cordata* and *P. peruviana* (Table 6.7).

Table 6.7. Mean genetic distance between (interspecific) *Physalis* species based on ITS2 sequences.

Groups	***P. purpurea***	***P. peruviana***	***P. cordata***
P. purpurea		198.92	1589.41
P. peruviana	9.58		357.92
P. cordata	21.99	9.53	

The average intra-specific distance within *Physalis* species was determined based on ITS2 sequences. The highest mean intraspecific distance was noted for *P. purpurea* (9.98 ± 12.73), followed by *P. peruviana* (1.31 ± 0.46), while the lowest mean intraspecific distance (0.72 ± 0.13) was recorded for *P. cordata*. The divergence was higher within *P. purpurea* and lowest within *P. cordata*. The average intraspecific distance within *Physalis* accessions was also determined based on *rbcL* sequences. The intraspecific distance within *Physalis* species based on *rbcL* sequences was 0.03 ± 0.00.

6.3.7. Nucleotide Polymorphism

In total, 4 segregation sites (S) were identified within the ITS2 sequences, while 59 segregation sites were identified within the *rbcL* gene sequences (Table 6.8). The nucleotide diversity (Pi) of ITS2 sequences was 0.15917, which is higher than that of *rbcL* sequences (0.01632) (Table 6.8). For the ITS2 sequences, the four polymorphic sites identified had 1 singleton and 3 parsimony informative bases, while the *rbcL* sequences had 48 singletons and 11 parsimony informative sites (Table 6.8).

Table 6.8. DNA polymorphism of *Physalis* accessions based on ITS2 and *rbcL* sequences.

	ITS2			***rbcL***		
Polymorphic Sites/Segregation Sites (S)	**4**	**Position in the Gene**	**Variants**	**59**	**Positions in the Gene**	**Variants**
Singleton	1	177	2	48	141,272,273,276,280,283,284,293,298, 301,308,309,310,322,325,327,331,334, 335,337,339,340,345,346,347,348,350, 353,357,365,366,373,375,376,386,395, 396,398,413,414,416,419,436,441,447, 457 344,359	2 3
Parsimony informative sites	3	179 176 178	2 3 4	11	302,336,341,355,358,362,401,430,444 282,363	2 3
Nucleotide diversity (Pi)	0.15917			0.01632		
Average number of nucleotide differences (k)	0.955			5.844		
Sequence length (base pairs)	532			716		
Number of sequences	28			28		

6.3.8 Tajima's neutrality tests

Tajima's neutrality test was conducted on the ITS2 and *rbcL* barcode sequences in order to establish the existence of a population selection based on the Tajima D value and nucleotide diversity. The Tajima D value of ITS2 sequences (0.870515) was higher compared to that of *rbcL* (−2.73462). The nucleotide diversity based on Tajima's test was also significantly higher for ITS2 sequences ($\pi = 0.176498$) compared to *rbcL* ($\pi = 0.067832$).

6.3.9 Barcoding gap analysis

The Automatic Barcode Gap Discovery (ABGD) results generated by the K80 Kimura measure of distance based on ITS2 and *rbcL* markers for *Physalis* accessions were used to determine the presence of a barcoding gap (Figure 6.6). The histogram ranked pairwise distances by increasing distance values from 0.02 to 1.28 and 0.02 to 0.14 for ITS2 and *rbcL* gene sequences, respectively (Figure 6.6). No barcode gap was detected via ITS2 ABGD analysis, while two barcode gaps were detected by the *rbcL* ABGD analysis (Figure 6.6). The first barcode gap for the *rbcL* gene sequence was detected between distances of 0.02 (2%) and 0.03 (3%), while the second barcode gap was between a distance of 0.12 (12%) and 0.13 (13%).

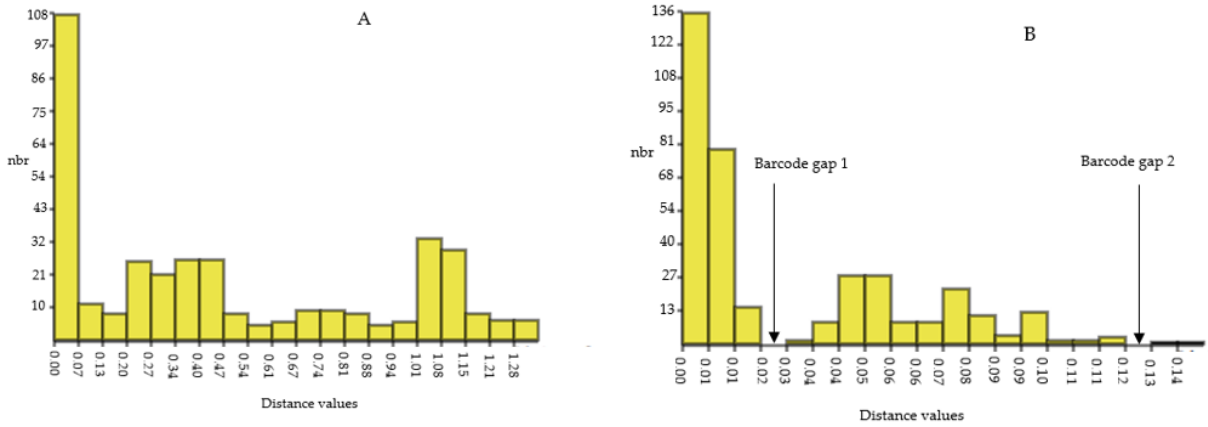


Figure 6.6. Histogram for the hypothetical distribution of pairwise differences of ITS2 and *rbcL* gene sequences for 28 *Physalis* accessions

A represents the histogram prepared using ITS2 genes while B represents the histogram prepared using *rbcL* genes

Low divergence is presumably intraspecific divergence, whereas higher divergence indicates interspecific divergence. The abbreviation nbr on the y-axis of the histogram stands for number of pairwise comparisons.

6.4 Discussion

DNA barcoding is a novel approach for identifying and discriminating species based on the nucleotide diversity of target/specific conserved sequences. Several studies have indicated that the DNA barcodes *rbcL* and ITS2, based on the chloroplast–plastid and nuclear regions, respectively, have been used to identify various plant families with similar morphological traits (Ralte and Singh, 2021). This study aimed at species discrimination in *Physalis* genotypes collected from different regions in Kenya by deploying both *rbcL* and ITS2 barcodes, and evaluated the efficiency of these markers in the barcoding of *Physalis* species. This is the first report to identify *Physalis* in Kenya using chloroplast–plastid and nuclear regions.

In previous studies, DNA barcoding markers, *rbcL, psbA-trnH* and ITS2 have been proven to be efficient in discriminating *Physalis* species from China and India (Ralte and Singh, 2021, Feng *et al*., 2018, Feng *et al*., 2016). These barcode genes were identified as potential candidates for the barcoding of *Physalis* plants. In the current study, the amplification was not universal because 16% and 23% of the samples did not amplify for *rbcL* and ITS2, respectively. Amplification failure can be attributed to DNA degradation during the transit of samples from the field to the laboratory. In addition, failures of DNA amplification and sequencing could also be linked to poor-quality DNA due to the presence of large amounts of secondary metabolites, such as phenolic compounds released during DNA isolation, which are common in *Physalis* species (Kang *et al*., 2017, Medina-Medrano *et al*., 2015).

The *rbcL* region of *Physalis* in this study was amplified more effectively compared to the ITS2 region. This concurs with previous studies, which showed higher amplification and sequencing success rate for *rbcL* compared to ITS2 (Kang *et al*., 2017, Huang *et al*., 2015). The high success rate of *rbcL* amplification is attributed to the high conservation of the gene and its low frequency rates of mutation (Kang *et al*., 2017). Conversely, the lower amplification and sequencing success rate of the ITS2 barcode could be attributed to its incomplete concerted evolution

process, as reported in other species (Simeone *et al*., 2013, Denk *et al*., 2010, Abeysinghe *et al*., 2009).

Basic Local Alignment Search Tool (BLAST) results have been used to identify the genus and facilitate species differentiation. Taxonomic assignments of *Physalis* accessions through BLASTn analyses against publicly available accessions in the databases did not give reliable results. This was probably because of the limited sequence data, since the available sequences in the databases mostly represent the most well-known and broadly studied species with a larger distribution, and to a lesser extent, species from insufficiently studied regions (Ross *et al*., 2008). Therefore, much more information and richer databases are necessary for the reliable application of the BLAST analysis to the Kenyan *Physalis* species.

The levels of genetic discrimination of *Physalis* accessions based on genetic distances differed between the two DNA barcode regions. All *rbcL* sequences and their reference sequences from the database formed a distinct cluster with no differentiation of species, indicating low levels of genetic differentiation in the *Physalis* species. The nucleotide data matrix from the *rbcL* region reflects the close genetic relationships of these species. This indicates the inefficiency of using *rbcL* in discriminating plant species, and thus we consider this region to offer little information relevant to the taxonomic classification of *Physalis*. The inefficiency of *rbcL* in discriminating plant species compared to other barcodes has also been noted in other studies (Kang *et al*., 2017, Huang *et al*., 2015, Tripathi *et al*., 2013). Similar results were presented in other studies, where the phylogenetic tree-based method could not effectively identify species of plants based on *rbcL* sequences (Huang *et al*., 2015). A study that used over 10,000 *rbcL* sequences from the GenBank to identify plant species also came up with similar conclusions to this study—that *rbcL* can only discriminate at the genus level (Newmaster *et al*., 2006). Chloroplast *rbcL* had higher universality but narrow inter-specific genetic divergence, and its species discrimination power was restricted. It is recommended that when *rbcL* is used as a first-tier barcode in species discrimination, and a supplement barcode is also used to increase the efficiency of species discrimination due to the limitations of the *rbcL* barcode (Newmaster *et al*., 2006).

However, the phylogenetic tree constructed based on ITS2 sequences demarcated the *Physalis* accessions into three distinct clades, with each representing a different *Physalis* species namely, *P. peruviana*, *P. cordata* and *P. purpurea*. This could be due to the fact that the ITS2 region possesses high interspecific and low intraspecific divergence (Chen *et al*., 2010). The clades had varying branch lengths, an indication that there was a divergence of the ITS2 sequences among the identified *Physalis* species (Binet *et al*., 2016). The branch lengths of the ITS2 sequences were much longer than those of the *rbcL* sequences, an indication that the ITS2 gene was more divergent, while the *rbcL* gene was more conserved among *Physalis* accessions. This concurs with the results of the genetic diversity studies, which showed a higher divergence among ITS2 as compared to *rbcL* sequences. The phylogenetic tree also showed longer branch lengths among the *P. peruviana* species, an indication that the two *P. peruviana* identified had a high intraspecific divergence. The more divergent the DNA barcode is, the better its ability to provide plant species discrimination among the targeted species (Kartavstev, 2011). Therefore, comparatively, the ITS2 sequences enabled better *Physalis* species discrimination based on Bayesian inference.

Higher nucleotide diversity was obtained for ITS2 compared to *rbcL*, an indication that the *rbcL* barcode is more conserved than ITS2. Therefore, the ITS2 barcode is useful to the interspecific divergence analysis of the *Physalis* accessions used in this study, which is also indicated by its ability to discriminate *Physalis* species. The interspecific divergence analysis of the ITS2 sequences in this study showed the highest nucleotide diversity between *P. peruviana* and *P. cordata* and the lowest between *P. cordata* and *P. purpurea*. One study postulated that a barcode has to exhibit high interspecific divergence to achieve the discrimination of species, especially amongst closely related sister species, while having low intraspecific variation (Rach *et al*, 2008). The current study showed that ITS2 was less conserved and possessed higher interspecific divergence than *rbcL*, indicating the level of species divergence among *Physalis* accessions used in this study.

Genetic distance is a measure of the genetic divergence between species or populations within a species (Beaumont *et al*., 1998). This was significantly higher for the ITS2 barcode compared to that of *rbcL*. This is an indication that there is high genetic divergence and variation among

Physalis species based on the ITS2 barcode. Based on the genetic distance, ITS2 was able to discriminate *Physalis* accessions into various species. The highest and lowest intraspecific distances were obtained within the *P. purpurea* and *P. cordata* populations, respectively. The low genetic distance for *rbcL* sequences is also a confirmation that the barcode is highly conserved in *Physalis* accessions used in this study. The results of the nucleotide polymorphism analysis for the ITS2 and *rbcL* sequences concur with those of the nucleotide divergence analysis, where ITS2's nucleotide diversity was higher than that of *rbcL*. A higher number of singleton and parsimony mutations in the *rbcL* gene indicate higher low-frequency mutations, concurring with the Tajima D value confirming the high level of conservation of the *rbcL* barcode (Carlson *et al*., 2005, Tajima, 1989b). The nucleotide polymorphism of the ITS2 sequences showed fewer low-frequency mutations compared to *rbcL*, and this explains the higher divergence among ITS2 sequences. The Automatic Barcode Gap Discovery (ABDG) was also able to show the intraspecific divergence between ITS2 and *rbcL* sequences of *Physalis* accessions used in this study. The maximum intraspecific distance, P_{max}, was much higher at 0.1 for ITS2 than 0.0219 for *rbcL*. This is an indication that ITS2 is not only more divergent between species, but is also more divergent within species compared to *rbcL*, which is highly conserved between and within *Physalis* species.

An ideal DNA barcode has significantly smaller intraspecific than interspecific distances, with a clear boundary between the two, referred to as the DNA barcoding gap, which can help in the identification of species (Ge *et al*., 2021; Meyer *et al*., 2005). This study confirmed that *rbcL* is highly conserved in *Physalis* plants, as its maximum intraspecific distance based on the automatic barcode gap discovery (ABGD) analysis was $P_{max} = 0.0129$. On the other hand, for the ITS2 marker, the maximum intraspecific distance based on the ABGD analysis was $P_{max} = 0.1$. This confirms that *rbcL* sequences cannot be used to group the *Physalis* accessions into species, and were indeed unable to discriminate *Physalis* species. This has also been reported in studies of other plant species, such as cinnamon, where not only *rbcL* but also other chloroplast-based barcodes such as *matK* and the intergenic sequence *psbA-trnH* were unable to discriminate and identify species of cinnamon (Chandrasekara *et al*., 2021). In other studies, *matK* and *psbA-trnH* have been shown to have better and higher potential as barcodes for the identification of tropical cloud forest trees than *rbcL* (Huang *et al*., 2015). However, other studies have shown that *rbcL* is

useful in the species discrimination of yams (Kipkiror *et al*., 2023). This suggests that *rbcL* species discrimination might differ from one genus of plants to another. The ITS2 sequences of the *Physalis* plants used in this study recorded high intraspecific divergence, as seen in the ABDG analysis (P_{max} = 0.1), probably due to its high variation. The ITS2 sequences were able to discriminate the *Physalis* accessions into three species, and the barcoding gap could be identified for all the three of these species. Their interspecific distance was much higher than that yielded by the ITS2 marker. The presence of a barcoding gap in different species is also an indication that ITS2 is an ideal candidate barcode for use in the discrimination of *Physalis* species and the determination of species diversity.

6.5 Conclusions

The results regarding sequence characteristics, genetic distance and phylogenetic relationships show that ITS2 is a reliable marker for use in the discrimination of *Physalis* species, whereby the accessions used were identified and discriminated into three species, namely, *P. purpurea*, *P. peruviana* and *P. cordata*. The ITS2 barcode was found to possess a sufficient variable region between the different species and accessions for the determination of genetic divergence with high discriminatory ability. These results expand our knowledge of genetic relationships that will benefit future crop improvement strategies in the areas of food, nutrition and therapeutics.

CHAPTER SEVEN

7.0 General discussion, conclusion and recommendations

7.1 General discussion

Physalis species contains nutritional and bioactive compounds of immense importance to public health and represents a potential ingredient for the development of functional foods and beverages. However, no study has reported the chemical composition, nutritional and antioxidant capacity of *Physalis* fruits in the wild and cultivated in Kenya. The lack of this information affects the production and utilization of this important fruit. Based on ITS2 barcode the *Physalis* accessions collected from Kericho County were identified as *P. purpurea* and were used for the analysis of macro-mineral and trace element content. The accessions of *P. purpurea* were found to be rich in both macrominerals (calcium, potassium, magnesium and sodium) and trace elements (zinc, copper, nickel, lithium and manganese). The results indicate that many of the macro- and micro-nutrients necessary for human health are found in *Physalis* fruits. The key phytochemicals identified in the fruits were phenolics, tannins and flavonoids and they facilitate the radical scavenging activity of DPPH radical and hydrogen peroxide.

Over the last decade, commercial *Physalis* plantations in Kenya have significantly increased and the farmers are concerned about the reliability and identity of the planting material. Among the challenges facing the production of planting material is the difficulty in precise identification of the *Physalis* species. A further confounding problem is the lack of discriminatory morphological features in plants of different *Physalis* species. DNA barcoding is a method that can effectively identify species according to the short DNA fragments information. The nuclear barcode, ITS2 has lower length variation and more common primer sites, which can better elucidate the genetic relationship between plant species. Several studies have also demonstrated that *rbcL* is effective and commonly used to identify plant species. Species identification based on the *rbcL* DNA barcode was successful for BLASTn analysis but species discrimination based on phylogenetic analysis was not possible based on the same barcode. Using BLAST algorithm on the NCBI database this study was able to identify all the *Physalis* accessions collected at the genus level, but none were identified at the species level. Despite *rbcL* barcode having high rate of amplification and sequencing among the *Physalis* accessions used in this study, it did not

facilitate the discrimination of *Physalis* species based on phylogenetic analysis as no clear grouping was observed. This concurs with previous studies in other plants where *rbcL* DNA barcode has been reported to have low efficacy in species discrimination of some plant families (Pei *et al*., 2015; Li *et al*., 2014).

Physalis species discrimination by BLAST algorithm on the NCBI database and phylogenetic analysis was effective based on ITS2 DNA barcode. The ITS2 barcode identified three *Physalis* species among the *Physalis* accessions collected in Kenya based on phylogenetic analysis. The three identified species were *P. cordata*, *P. peruviana* and *P. purpurea*. Therefore, ITS2 provided more robust species discrimination than the *rbcL* genes. Other studies have also confirmed the ability of ITS2 to identify and discriminate species (Zhang *et al*., 2015). The combination of multi loci barcodes could improve the species discrimination and several studies have demonstrated the effectiveness (Tran *et al*., 2021; Simeone *et al*., 2013). In the current study, the combination of the two barcode regions failed to increase the *Physalis* species discrimination efficiency compared to the single barcode.

Nucleotide diversity and genetic distance of *Physalis* accessions was higher based on ITS2 barcode as compared to the *rbcL* DNA barcode. The nucleotide polymorphism of ITS2 and *rbcL* barcode sequences for the *Physalis* accessions used in this study was different. Although *rbcL* barcode sequences had higher nucleotide polymorphism than the ITS2 barcode, it did not provide better discrimination of *Physalis* species. Nucleotide polymorphisms observed in the *rbcL* barcode sequences of *Physalis* accessions were singleton mutations which are considered low frequency mutations (Carlson *et al*., 2005). This explained why the low genetic diversity observed in the *rbcL* barcode despite the high nucleotide polymorphism.

7.2 Conclusions

(i) The study confirmed that the fruits of *P. purpurea* contained a high content of minerals, including calcium, sodium, magnesium, and potassium. The fruits were also rich in phenolic acids, tannins, and flavonoids, and exhibited antioxidant properties. The phenolic compounds and flavonoids were the major contributors to the radical scavenging activity of the *P. purpurea* fruits.

(ii) DNA barcode ITS2 was highly efficient in the identification and discrimination of *Physalis* species. The *Physalis* accessions were identified and discriminated into three species namely *P. peruviana*, *P. purpurea* and *P. cordata*. Our study revealed significant variations in ITS2 secondary structure predictions that enhanced discrimination among the three identified *Physalis* species.

(iii) The results regarding sequence characteristics, genetic distance and phylogenetic relationships show that *rbcL* identified the genus *Physalis* but could not discriminate the 56 accessions studied into different *Physalis* species indicating high conservation of *rbcL* gene. Low sequence divergence and low genetic distance among *Physalis* accessions based on *rbcL* gene was also noted, indicating the *rbcL* gene is mostly conserved among the accessions studied.

(iv) The results regarding sequence characteristics, genetic distance and phylogenetic relationships show that ITS2 is a reliable marker compared to *rbcL* for use in the discrimination of *Physalis* species, whereby the accessions used were identified and discriminated based on the ITS2 gene into three species namely, *P. purpurea*, *P. peruviana* and *P. cordata*.

7.3 Recommendations

(i) The underutilized *P. purpurea* was demonstrated to be an excellent source of minerals, phytochemicals and antioxidants and therefore there is need to exploit the accessions for the management of oxidative stress-induced human diseases.

(ii) *Physalis* accessions identified can be utilized for genetic breeding to develop hybrids that are valuable to human nutrition and health.

(iii)There is need to evaluate the efficiency of other DNA barcodes such as *matK*, *trnK* intron, *trnH-psdA* and *psbA-trnH* in species discrimination of *Physalis* plants. These DNA barcodes can be used to assess species discrimination, genetic diversity and nucleotide polymorphism of *Physalis* genus in Kenya.

(iv) There is need for further studies to understand the use of RNA secondary structures for the selection of parents in trait-specific breeding strategies for *Physalis* improvement.

(v) There is need for the development of species-specific DNA barcode for the discrimination of the three *Physalis* species identified in the current study as well as other *Physalis* species.

REFERENCES

Abbaspour, N., Hurrell, R., and Kelishadi, R. (2014). Review on iron and its importance for human health. *Journal of Research in Medical Sciences*, *19*(2), 164.

Abdul-Nasir-Deen, A. Y., Boakye, Y. D., Osafo, N., Agyare, C., Boamah, D., Boamah, V. E., and Agyei, E. K. (2020). Anti-inflammatory and wound healing properties of methanol leaf extract of *Physalis angulata L. South African Journal of Botany*, *133*, 124-131.

Abeysinghe, P. D., Wijesinghe, K. G. G., Tachida, H., Yoshda, T., and Thihagoda, M. (2009). Molecular characterization of *Cinnamon* (*Cinnamomum verum Presl*) accessions and evaluation of genetic relatedness of Cinnamon species in Sri Lanka based on trnL intron region, intergenic spacers between *trnT-trnL*, *trnL-trnF*, *trnH-psbA* and nuclear ITS. *Research Journal of Agriculture and Biological Sciences*, *5*(6), 1079-1088.

Acharya, G. C., Mohanty, S., Dasgupta, M., Sahu, S., Singh, S., Koundinya, A. V., and Sahoo, M. R. (2022). Molecular phylogeny, DNA barcoding, and ITS2 secondary structure predictions in the medicinally important *Eryngium* genotypes of East Coast Region of India. *Genes*, *13*(9), 1678.

Afroz, M., Akter, S., Ahmed, A., Rouf, R., Shilpi, J. A., Tiralongo, E., and Uddin, S. J. (2020). Ethnobotany and antimicrobial peptides from plants of the Solanaceae family: an update and future prospects. *Frontiers in Pharmacology*, *11*, 565.

Ahmad, S., Malik, A., Yasmin, R., Ullah, N., Gul, W., Khan, P. M., and Afza, N. (1999). Withanolides from *Physalis peruviana*. *Phytochemistry*, *50*(4), 647-651.

Álvarez-Flórez, F., López-Cristoffanini, C., Jáuregui, O., Melgarejo, L. M., and López-Carbonell, M. (2017). Changes in ABA, IAA and JA levels during calyx, fruit and leaves development in *Physalis* plants (*Physalis peruviana L.*). *Plant Physiology and Biochemistry*, *115*, 174-182.

Amarowicz, R., Pegg, R. B., Rahimi-Moghaddam, P., Barl, B., and Weil, J. A. (2004). Free-radical scavenging capacity and antioxidant activity of selected plant species from the Canadian prairies. *Food Chemistry*, *84*(4), 551-562.

Amiteye, S. (2021). Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon*, *7*(10).

Anaz, M., Sasidharan, N., Remakanthan, A., and Dilsha, M. V. (2021). ITS2 and RNA secondary structure-based analysis reveals a clear picture on phylogeny of South Indian *Salacia spp*. *Computational Biology and Chemistry*, *91*, 107438.

Anjalam, A., Kalpana, S., Vijai, D., and Premalatha, S. (2016). Documentation of medicinal plants used by Malayali tribes in Kolli Hills. *International Journal of Advanced Research in Biological Sciences*, *3*(3), 101-107.

Arun, M., and Asha, V. V. (2007). Preliminary studies on antihepatotoxic effect of *Physalis peruviana Linn*. (Solanaceae) against carbon tetrachloride induced acute liver injury in rats. *Journal of Ethnopharmacology*, *111*(1), 110-114.

Aryal, S., Baniya, M. K., Danekhu, K., Kunwar, P., Gurung, R., and Koirala, N. (2019). Total phenolic content, flavonoid content and antioxidant potential of wild vegetables from Western Nepal. *Plants*, *8*(4), 96.

Bailey, L. H. (1949). Manual of cultivated plants. *Manual of Cultivated Plants.*

Banerjee, A., and Roychoudhury, A. (2017). Effect of salinity stress on growth and physiology of medicinal plants. *Medicinal Plants and Environmental Challenges*, 177-188.

Barcaccia, G., Lucchin, M., and Cassandro, M. (2015). DNA barcoding as a molecular tool to track down mislabeling and food piracy. *Diversity*, *8*(1), 2.

Barirega, A. (2014). Potential for value chain improvement and commercialization of *Physalis* (*Physalis peruviana L.*) for livelihood improvement in Uganda. *Ethnobotany Research and Applications*, *12*, 131-140.

Bayir, A. G., Kiziltan, H. S., and Kocyigit, A. (2019). Plant Family, Carvacrol, and Putative Protection in Gastric Cancer. In *Dietary Interventions in Gastrointestinal Diseases* (pp. 3-18). Academic Press.

Bazalar Pereda, M. S., Nazareno, M. A., and Viturro, C. I. (2019). Nutritional and antioxidant properties of *Physalis peruviana L*. fruits from the Argentinean northern Andean region. *Plant Foods for Human Nutrition*, *74*(1), 68-75.

Beaumont, M. A., Ibrahim, K. M., Boursot, P., and Bruford, M. W. (1998). Measuring genetic distance. *Molecular Tools for Screening Biodiversity: Plants and Animals*, 315-325.

Berchtold, M. W., Brinkmeier, H., and Muntener, M. (2000). Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity, and disease. *Physiological Reviews*, *80*(3), 1215-1265.

Binet, M., Gascuel, O., Scornavacca, C., Douzery, E. J., and Pardi, F. (2016). Fast and accurate branch lengths estimation for phylogenomic trees. *BMC Bioinformatics*, *17*(1), 1-18.

Brand-Williams, W., Cuvelier, M. E., and Berset, C. L. W. T. (1995). Use of a free radical method to evaluate antioxidant activity. *LWT-Food Science and Technology*, *28*(1), 25-30.

Brown, K. H., Wuehler, S. E., and Peerson, J. M. (2001). The importance of zinc in human nutrition and estimation of the global prevalence of zinc deficiency. *Food and Nutrition Bulletin*, *22*(2), 113-125.

Buckley, D. S., Zasada, J. C., Tappeiner II, J. C., and Stone, D. M. (1997). Plant morphological characteristics as a tool in monitoring response to silvicultural activities. *Communicating the Role of Silviculture in Managing the National Forests*, 37.

Cailes, R. L. 1952. The cultivation of *Physalis*. *Journal of the Department of Agriculture, Western Australia*. *1*(3), 363-365.

Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., and Nickerson, D. A. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, *15*(11), 1553-1565.

CBOL Plant Working Group 1, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., and Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794-12797.

Chacón, S., Isabel, M., Sánchez, Y. D. P., Barrero, M., and Stella, L. (2016). Genetic structure of a Colombian *Physalis* (*Physalis peruviana L.*) collection by means of microsatellite markers. *Agronomía Colombiana*, *34*(1), 5-16.

Chandrasekara, C. B., Naranpanawa, D. N. U., Bandusekara, B. S., Pushpakumara, D. K. N. G., Wijesundera, D. S. A., and Bandaranayake, P. C. (2021). Universal barcoding regions, *rbcL*, *matK* and *trnH-psbA* do not discriminate *Cinnamomum* species in Sri Lanka. *PLoS One*, *16*(2), 1-16.

Chang, C. C., Yang, M. H., Wen, H. M., and Chern, J. C. (2002). Estimation of total flavonoid content in propolis by two complementary colorimetric methods. *Journal of Food and Drug Analysis*, *10*(3), 3.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N., and Savolainen, V. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1889-1895.

Cháves-Gómez, J. L., Becerra-Mutis, L. M., Chávez-Arias, C. C., Restrepo-Díaz, H., and Gómez-Caro, S. (2020). Screening of different *Physalis* genotypes as potential rootstocks or parents against vascular wilt using physiological markers. *Frontiers in Plant Science*, *11*, 806.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., and Luo, K. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one*, *5*(1), e8613.

Chen, S. L. (2012). Traditional Chinese medicine molecular identification by DNA barcode. *Beijing: People's Medical Publishing House*.

China Plant BOL Group 1, Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., and Duan, G. W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences*, *108*(49), 19641-19646.

Chinese Academy of Sciences. (1978). Flora of China. *Science Press*, *67*, 50.

Chinese Pharmacopoeia Editorial Committee. (2015). Pharmacopoeia of the People's Republic of China. *Chemical Industry Press*, *1*, 360–361.

Coleman, A. W. (2003). ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics*, *19*(7), 370-375.

Collins, R. A., and Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, *13*(6), 969-975.

Delgado-Bastidas, N., Lagos-Santander, L. K., and Lagos-Burbano, T. C. (2019). Genetic diversity of 40 genotypes of golden berry *Physalis peruviana L*. using microsatellite markers. *Revista de Ciencias Agrícolas*, *36*(E), 95-107.

Dellaporta, S. L, Wood, J., and Hicks, J. B. (1983). A rapid method for DNA extraction from plant tissue. *Plant Molecular Biology Reporter 1*, 19-21.

Demling, J. H., Eglau, M. C., and Autenrieth, T. (2001). On the physiological function of lithium from a psychiatric view point. *Medical Hypotheses*, *57*(4), 506-509.

Denk, T., and Grimm, G. W. (2010). The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. *Taxon*, *59*(2), 351-366.

de Vere, N., Rich, T. C., Trinder, S. A., and Long, C. (2015). DNA barcoding for plants. *Plant Genotyping*: *Methods and Protocols*, *1245*, 101-118.

Dormontt, E. E., Van Dijk, K. J., Bell, K. L., Biffin, E., Breed, M. F., Byrne, M., and Lowe, A. J. (2018). Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections—an Australian perspective. *Frontiers in Ecology and Evolution*, *6*, 134.

Duan, H., Wang, W., Zeng, Y., Guo, M., and Zhou, Y. (2019). The screening and identification of DNA barcode sequences for *Rehmannia*. *Scientific Reports*, *9*(1), 1-12.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Elansary, O. H., Ashfaq, M., Ali, H. M., and Yessoufou, K. (2017). The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. *PloS one*, *12*(2), e0172170.

El-Beltagi, H. S., Mohamed, H. I., Safwat, G., Gamal, M., and Megahed, B. M. (2019). Chemical composition and biological activity of *Physalis peruviana L*. *Gesunde Pflanzen*, *71*(2), 113-122.

Erkaya, T., Dağdemir, E., and Şengül, M. (2012). Influence of Cape gooseberry (*Physalis peruviana L.*) addition on the chemical and sensory characteristics and mineral concentrations of ice cream. *Food Research International*, *45*(1), 331-335.

Fatima, T., Srivastava, A., Somashekar, P. V., Hanur, V. S., and Rao, M. S. (2019). Development of DNA-based species identification and barcoding of three important timbers. *Bulletin of the National Research Centre*, *43*(1), 1-17.

Feng S. G., Lu J. J., Gao L., Liu J. J., and Wang H. Z. (2014). Molecular phylogeny analysis and species identification of *Dendrobium* (*Orchidaceae*) in China. *Biochemical Genetics, 52*(3-4), 127–136.

Feng S., Jiang Y., Wang S., Jiang M., Chen Z., and Ying Q. (2015). Molecular identification of *Dendrobium* species (*Orchidaceae*) based on the DNA barcode ITS2 region and its application for phylogenetic study. *International Journal of Molecular Sciences, 16*(9), 21975–21988.

Feng, S., Jiang, M., Shi, Y., Jiao, K., Shen, C., Lu, J., and Wang, H. (2016). Application of the ribosomal DNA ITS2 region of *Physalis* (Solanaceae): DNA barcoding and phylogenetic study. *Frontiers in Plant Science, 7*, 1047.

Feng, S., Jiao, K., Zhu, Y., Wang, H., Jiang, M., and Wang, H. (2018). Molecular identification of species of *Physalis* (Solanaceae) using a candidate DNA barcode: the chloroplast *psbA–trnH* intergenic region. *Genome, 61*(1), 15-20.

Feng, S., Zheng, K., Jiao, K., Cai, Y., Chen, C., Mao, Y., and Wang, H. (2020). Complete chloroplast genomes of four *Physalis* species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. *BMC Plant Biology*, *20*(1), 1-14.

Feng, S., Jiao, K., Zhang, Z., Yang, S., Gao, Y., Jin, Y., and Wang, H. (2023). Development of Chloroplast Microsatellite Markers and Evaluation of Genetic Diversity and Population Structure of Cutleaf Groundcherry (*Physalis angulata L.*) in China. *Plants*, *12*(9), 1755.

Franco, L. A., Matiz, G. E., Calle, J., Pinzón, R., and Ospina, L. F. (2007). Antiinflammatory activity of extracts and fractions obtained from *Physalis peruviana L.* calyces. *Biomedica*, *27*(1), 110-5.

Garzón-Martínez, G. A., Osorio-Guarín, J. A., Delgadillo-Durán, P., Mayorga, F., Enciso-Rodríguez, F. E., Landsman, D., and Barrero, L. S. (2015). Genetic diversity and population structure in *Physalis peruviana* and related taxa based on InDels and SNPs derived from COSII and IRG markers. *Plant Gene*, *4*, 29-37.

Ge, Y., Xia, C., Wang, J., Zhang, X., Ma, X., and Zhou, Q. (2021). The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. *Ecology and Evolution*, *11*(10), 5669-5681.

Genchi, G., Carocci, A., Lauria, G., Sinicropi, M. S., and Catalano, A. (2020). Nickel: Human health and environmental toxicology. *International Journal of Environmental Research and Public Health*, *17*(3), 679.

Gengmao, Z., Yu, H., Xing, S., Shihui, L., Quanmei, S., and Changhai, W. (2015). Salinity stress increases secondary metabolites and enzyme activity in safflower. *Industrial Crops and Products*, *64*, 175-181.

Gironés-Vilaplana, A., Baenas, N., Villaño, D., Speisky, H., García-Viguera, C., and Moreno, D. A. (2014). Evaluation of Latin-American fruits rich in phytochemicals with biological effects. *Journal of Functional Foods*, *7*, 599-608.

Golubkina, N. A., Kekina, H. G., Engalichev, M. R., Antoshkina, M. S., Nadezhkin, S. M., and Caruso, G. (2018). Yield, quality, antioxidants and mineral nutrients of *Physalis angulata L.* and *Physalis pubescens L.* fruits as affected by genotype under organic management. *Advances in Horticultural Science*, *32*(4), 541-548.

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics Research International*, *2015*, 1-14

Gruber A. R, Lorenz R., Bernhart S. H., Neuböck R., and Hofacker I. L. (2008). The Vienna Websuite. *Nucleic Acids Research*, (36), (2), 70-74.

Gülçin, İ., Huyut, Z., Elmastaş, M., and Aboul-Enein, H. Y. (2010). Radical scavenging and antioxidant activity of tannic acid. *Arabian Journal of Chemistry*, *3*(1), 43-53.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H., Song, J., and Meng, F. (2013). The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *BioMed Research International, 2013*, 1-7.

Hebert, P. D., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1512), 313-321.

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., and Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, 101*(41), 14812-14817.

Hernández, O. M., Fraga, J. M. G., Jiménez, A. I., Jiménez, F., and Arias, J. J. (2005). Characterization of honey from the Canary Islands: determination of the mineral content by atomic absorption spectrophotometry. *Food Chemistry, 93*(3), 449-458.

Ho, V. T., Tran, T. K. P., Vu, T. T. T., and Widiarsih, S. (2021). Comparison of *matK* and *rbcL* DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. *Journal of Genetic Engineering and Biotechnology, 19*(1), 93.

Hollingsworth, M. L., Andra-Clark, A. L., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., and Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species- level sampling in three divergent groups of land plants. *Molecular Ecology Resources, 9*(2), 439-457.

Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and using a plant DNA barcode. *PloS One, 6*(5), e19254.

Hong, J. M., Kwon, O. K., Shin, I. S., Song, H. H., Shin, N. R., Jeon, C. M., and Ahn, K. S. (2015). Anti-inflammatory activities of *Physalis alkekengi var. franchetii* extract through the inhibition of MMP-9 and AP-1 activation. *Immunobiology, 220*(1), 1-9.

Huang, X. C., Ci, X. Q., Conran, J. G., and Li, J. (2015). Application of DNA barcodes in Asian tropical trees–a case study from Xishuangbanna Nature Reserve, Southwest China. *PLoS One*, *10*(6), e0129295.

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755.

IBM SPSS, I. (2010). SPSS for windows (version 20). Chicago, Ill, USA.

Idrees, M., and Irshad, M. (2014). Molecular markers in plants for analysis of genetic diversity: a review. *European Academic Research*, *2*(1), 1513-1540.

Ismail, M., Ahmad, A., Nadeem, M., Javed, M. A., Khan, S. H., Khawaish, I., and Qamer, S. (2020). Development of DNA barcodes for selected Acacia species by using *rbcL* and *matK* DNA markers. *Saudi Journal of Biological Sciences*, *27*(12), 3735-3742.

Jahnen-Dechent, W., and Ketteler, M. (2012). Magnesium basics. *Clinical Kidney Journal*, *5*(Suppl_1), i3-i14.

Juyó, D., and Sarmiento, F. (2015). Á lvarez M, Brochero H, Gebhardt C, Mosquera T. Genetic diversity and population structure in diploid potatoes of *Solanum tuberosum* group Phureja. *Crop Science*, *55*, 760-769.

Kang, Y., Deng, Z., Zang, R., and Long, W. (2017). DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Scientific Reports*, *7*(1), 1-9.

Karpagasundari, C., and Kulothungan, S. (2014). Free radical scavenging activity of *Physalis minima* Linn. leaf extract (PMLE). *Journal of Medicinal Plant Studies*, *2*(4), 59-64.

Kartavtsev, Y. P. (2011). Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. *Mitochondrial DNA*, *22*(3), 55-65.

Kasali, F. M., Tusiimire, J., Kadima, J. N., Tolo, C. U., Weisheit, A., and Agaba, A. G. (2021). Ethnotherapeutic uses and phytochemical composition of *Physalis peruviana* L.: an overview. *The Scientific World Journal, 2021*.

Kipkiror, N., Muge, E. K., Ochieno, D. M., and Nyaboga, E. N. (2023). DNA barcoding markers provide insight into species discrimination, genetic diversity and phylogenetic relationships of yam (*Dioscorea* spp.). *Biologia*, *78*(3), 689-705.

Khan, W., and Bakht, J. (2015). Development of efficient and optimized protocol for rapid micropropagation of *Physalis ixocarpa*, a medicinal herb. *Pakistan Journal of Weed Science Research*, *21*(3).

Korneliussen, T. S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, *14*(1), 1-14.

Kowey, P.R (2002). The role of potassium. In *Women's Health and Menopause: New Strategies—Improved Quality of Life*, , 151–157.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences*, *102*(23), 8369-8374.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35 (6), 1547–1549.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., and Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, *105*(8), 2923-2928.

Lan, Y. H., Chang, F. R., Pan, M. J., Wu, C. C., Wu, S. J., Chen, S. L., and Wu, Y. C. (2009). New cytotoxic withanolides from *Physalis peruviana*. *Food Chemistry*, *116*(2), 462-469.

Li, Y., Zhou, X. I. N., Feng, G., Hu, H., Niu, L., Hebert, P. D., and Huang, D. (2010). COI and ITS2 sequences delimit species, reveal cryptic taxa and host specificity of fig- associated Sycophila (Hymenoptera, Eurytomidae). *Molecular Ecology Resources*, *10*(1), 31-40.

Li, Y., Feng, Y., Wang, X. Y., Liu, B., and Lv, G. H. (2014). Failure of DNA barcoding in discriminating Calligonum species. *Nordic Journal of Botany*, *32*(4), 511-517.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., and Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*(1), 157-166.

Li, Q., Wu, J., Wang, Y., Lian, X., Wu, F., Zhou, L., and Zhu, S. (2017). The phylogenetic analysis of Dalbergia (Fabaceae: Papilionaceae) based on different DNA barcodes. *Holzforschung*, *71*(12), 939-949.

Li, Y., Kong, D., Fu, Y., Sussman, M. R., and Wu, H. (2020). The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiology and Biochemistry*, *148*, 80-89.

Li, H., Xiao, W., Tong, T., Li, Y., Zhang, M., Lin, X., and Guo, X. (2021). The specific DNA barcodes based on chloroplast genes for species identification of *Orchidaceae* plants. *Scientific Reports*, *11*(1), 1-15.

Lima, M. D. S., Evangelista, A. F., Santos, G. G., Ribeiro, I. M., Tomassini, T. C. B., Pereira Soares, M. B., and Villarreal, C. F. (2014). Antinociceptive properties of physalins from *Physalis angulata*. *Journal of Natural Products*, *77*(11), 2397-2403.

Lee, J. H., Lee, D. H., and Choi, B. H. (2013). Phylogeography and genetic diversity of East Asian Neolitsea sericea (Lauraceae) based on variations in chloroplast DNA sequences. *Journal of Plant Research*, *126*, 193-202.

Lledo, M. D., Crespo, M. B., Cameron, K. M., Fay, M. F., and Chase, M. W. (1998). Systematics of Plumbaginaceae based upon cladistic analysis of *rbcL* sequence data. *Systematic Botany*, 21-29.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*, 1-14.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences*, *101*(19), 7287-7292.

Manzara, T., and Gruissem, W. (1988). Organization and expression of the genes encoding ribulose-1, 5-bisphosphate carboxylase in higher plants. *Molecular Biology of Photosynthesis*, 621-643.

Martínez, M. (1998). Revisión de *Physalis* sección Epiteiorhiza (Solanaceae). *Anales del Instituto de Biología. Serie Botánica*, *69*(2), 71-117.

Medina-Medrano, J. R., Almaraz-Abarca, N., González-Elizondo, M. S., Uribe-Soto, J. N., González-Valdez, L. S., and Herrera-Arrieta, Y. (2015). Phenolic constituents and antioxidant properties of five wild species of *Physalis* (Solanaceae). *Botanical Studies*, *56*, 1-13.

Menzel, M. Y. (1951). The cytotaxonomy and genetics of *Physalis*. *Proceedings of the American Philosophical Society*, *95*(2), 132-183.

Meyer, C. P., and Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), 422.

Mishra, P., Kumar, A., Sivaraman, G., Shukla, A. K., Kaliamoorthy, R., Slater, A., and Velusamy, S. (2017). Character-based DNA barcoding for authentication and conservation of IUCN Red listed threatened species of genus Decalepis (Apocynaceae). *Scientific Reports*, *7*(1), 14910.

Muraguri, P. W., Gesimba, R. M., Wolukau, J. N., and Miheso, M. (2021). Genetic characterization of *Physalis* (*Physalis peruviana L*.) accessions in selected counties in Kenya using SSR markers. *African Journal of Plant Science*, *15*(1), 28-38.

Musinguzi, E., Kikafunda, J. K., and Kiremire, B. T. (2007). Promoting indigenous wild edible fruits to complement roots and tuber crops in alleviating vitamin A deficiencies in Uganda. In *Proceedings of the 13th ISTRC Symposium*, 763-769.

National Research Council. (2000). Copper in Drinking Water. National Research Council: Ottawa, ON, Canada; ISBN-10: 0-309-06939-4.

National Research Council. (2001). Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. *Institute of Medicine/Food and Nutrition Board. National Academy Press. Washington DC*.

Nei M., and Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Newmaster, S. G., Fazekas, A. J., and Ragupathy, S. (2006). DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Botany*, *84*(3), 335-341.

Novoa, R. H., Bojacá, M., Galvis, J. A., and Fischer, G. (2006). Fruit maturity and calyx drying influence post-harvest behavior of *Physalis* (*Physalis peruviana L*.) stored at 12 ºC. *Agronomia Colombiana*, *24*(1), 77-86.

Nurhasanah, S., and Papuangan, N. (2019, May). Amplification and analysis of *rbcL* gene (Ribulose-1, 5-Bisphosphate Carboxylase) of clove in Ternate Island. In *IOP Conference Series: Earth and Environmental Science*, *276*(1), 012061.

O'Dell, R. E., and Rajakaruna, N. (2011). Intraspecific variation, adaptation, and evolution. *Serpentine: Evolution and ecology in a model system*, *97*, 137.

Ogundajo, A. L., Akpome, A. S., Tijani, N. A., and Ogunwande, I. A. (2015). Chemical Constituents of the Leaf Essential Oil of *Physalis angulata* L. *Asian Journal of Applied Sciences*, *3*(4), 652-655.

Okwu, D. E. (2005). Phytochemicals, vitamins and mineral contents of two Nigerian medicinal plants. *International Journal of Molecular Medicine and Advance Sciences*, *1*(4), 375-381.

Ordoñez, C., Ana, Y., Estrada, M. E. M., and Cortés, R. M. (2017). The influence of drying on the physiological quality of *Physalis* (*Physalis peruviana L.*) fruits added with active components. *Acta Agronómica*, *66*(4), 512-518.

Oszmiański, J., Lachowicz, S., Gorzelany, J., and Matłok, N. (2018). The effect of different maturity stages on phytochemical composition and antioxidant capacity of cranberry cultivars. *European Food Research and Technology*, *244*, 705-719.

Pang, X., Song, J., Zhu, Y., Xu, H., Huang, L., and Chen, S. (2011). Applying plant DNA barcodes for Rosaceae species identification. *Cladistics*, *27*(2), 165-170.

Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., and Chen, S. (2012). Utility of the trnH–psbA intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. *PloS One*, *7*(11), 48833.

Pant, P., Pandey, S., and Dall'Acqua, S. (2021). The influence of environmental conditions on secondary metabolites in medicinal plants: A literature review. *Chemistry and Biodiversity*, *18*(11), 2100345.

Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D. W., Kenfack, D., and Hardy, O. J. (2013). How effective are DNA barcodes in the identification of African rainforest trees? *PLoS One*, *8*(4), 54921.

Pei, N., Erickson, D. L., Chen, B., Ge, X., Mi, X., Swenson, N. G., and Kress, W. J. (2015). Closely-related taxa influence woody species discrimination via DNA barcoding: evidence from global forest dynamics plots. *Scientific Reports*, *5*(1), 15127.

Petkova, N. T., Popova, V. T., Ivanova, T. A., Mazova, N. N., Panayotov, N. D., and Stoyanova, A. (2021). Nutritional composition of different *Physalis* genotypes (*Physalis peruviana L.*)—A comparative study. *Food Research*, *5*, 191.

Pinto, M. D. S., Ranilla, L. G., Apostolidis, E., Lajolo, F. M., Genovese, M. I., and Shetty, K. (2009). Evaluation of antihyperglycemia and antihypertension potential of native Peruvian fruits using in vitro models. *Journal of Medicinal Food*, *12*(2), 278-291.

Popenoe H, King S, Leon J, and Kalinowski L (1990) Goldenberry (*Physalis*). Lost Crops of the Incas: Little-known Plants of the Andes with Promise for Worldwide Cultivation, ed by National Research Council National Academy Press, Washington DC. 241.

Puente, L. A., Pinto-Muñoz, C. A., Castro, E. S., and Cortés, M. (2011). *Physalis peruviana Linnaeus*, the multiple properties of a highly functional fruit: A review. *Food Research International*, *44*(7), 1733-1740.

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. J. M. E. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*(8), 1864-1877.

Qian, Z. H., Munywoki, J. M., Wang, Q. F., Malombe, I., Li, Z. Z., and Chen, J. M. (2022). Molecular Identification of African *Nymphaea* Species (Water Lily) Based on ITS, trnT-trnF and rpl16. *Plants*, *11*(18), 2431.

Rach, J., DeSalle, R., Sarkar, I. N., Schierwater, B., and Hadrys, H. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1632), 237-247.

Ralte, L., and Singh, Y. T. (2021). Use of *rbcL* and *ITS2* for DNA barcoding and identification of Solanaceae plants in hilly state of Mizoram, India. *Research on Crops*, *22*(3).

Ramadan, M. F., and Mörsel, J. T. (2003). Oil goldenberry (*Physalis peruviana L.*). *Journal of Agricultural and Food Chemistry*, *51*(4), 969-974.

Ramadan, M. F., and Moersel, J. T. (2007). Impact of enzymatic treatment on chemical composition, physicochemical properties and radical scavenging activity of goldenberry (*Physalis peruviana L.*) juice. *Journal of the Science of Food and Agriculture*, *87*(3), 452-460.

Reddy, C. V. K., Sreeramulu, D., and Raghunath, M. (2010). Antioxidant activity of fresh and dry fruits commonly consumed in India. *Food Research International*, *43*(1), 285-288.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(1), 320-324.

Rodrigues, E., Rockenbach, I. I., Cataneo, C., Gonzaga, L. V., Chaves, E. S., and Fett, R. (2009). Minerals and essential fatty acids of the exotic fruit *Physalis peruviana L*. *Food Science and Technology*, *29*, 642-645.

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. 9.

Rosa, O. L., Contreras-López, E., Juárez, E. F., Barron, U. G., Muñoz, A. M., and Ramos-Escudero, F. (2023). Nutritional and antioxidant profile of the *Physalis* fruit grown in three Andean regions of Peru. *Roczniki Państwowego Zakładu Higieny*, *74*(1).

Rosario, L. H., Padilla, J. O. R., Martínez, D. R., Grajales, A. M., Reyes, J. A. M., Feliu, G. J. V., and Siritunga, D. (2019). DNA barcoding of the Solanaceae family in Puerto Rico including endangered and endemic species. *Journal of the American Society for Horticultural Science*, *144*(5), 363-374.

Ross, H. A., Murugan, S., and Sibon, L. W. L. (2008). Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, *57*(2), 216-230.

Ross, A.C.; Taylor, C.L.; Yaktine, A.L.; Valle, H.B. (2012). Overview of calcium—Dietary reference intakes for calcium and vitamin D—NCBI bookshelf. *Pediatrics*, *130*, 10–542. https://doi.org/10.17226/13050.

Ruch, R. J., Cheng, S. J., and Klaunig, J. E. (1989). Prevention of cytotoxicity and inhibition of intercellular communication by antioxidant catechins isolated from Chinese green tea. *Carcinogenesis*, *10*(6), 1003-1008.

Ruiz, L. M., Libedinsky, A., and Elorza, A. A. (2021). Role of copper on mitochondrial function and metabolism. *Frontiers in Molecular Biosciences*, *8*, 711227.

Saddhe, A. A., and Kumar, K. (2018). DNA barcoding of plants: selection of core markers for taxonomic groups. *Plant Science Today*, *5*(1), 9-13.

Sadiya, H., Ashari, S., Waluyo, B., and Soegianto A. (2021). Genetic diversity and relationship of husk tomato (*Physalis spp*.) from East Java Province revealed by SSR markers. *Biodiversitas Journal of Biological Diversity*, *22*(1), 184-192.

Sahoo, S., Ghosh, G., Das, D., and Nayak, S. (2013). Phytochemical investigation and in vitro antioxidant activity of an indigenous medicinal plant Alpinia nigra BL Burtt. *Asian Pacific Journal of Tropical Biomedicine*, *3*(11), 871-876.

Saidon, N. A., Wagiran, A., Samad, A. F. A., Mohd, S. F., Mohamed, F., Jani, J., and Linatoc, A. C. (2023). DNA Barcoding, Phylogenetic Analysis and Secondary Structure Predictions of *Nepenthes ampullaria, Nepenthes gracilis* and *Nepenthes rafflesiana*. *Genes*, *14*(3), 697.

Sandoval-Padilla, I., del Pilar Zamora-Tavares, M., Ruiz-Sánchez, E., Pérez-Alquicira, J., and Vargas-Ponce, O. (2022). Characterization of the plastome of *Physalis cordata* and comparative analysis of eight species of *Physalis* sensu stricto. *PhytoKeys*, *210*, 109-134.

Sang-Ngern, M., Youn, U. J., Park, E. J., Kondratyuk, T. P., Simmons, C. J., Wall, M. M., and Chang, L. C. (2016). Withanolides derived from *Physalis peruviana* (Poha) with potential

anti-inflammatory activity. *Bioorganic and Medicinal Chemistry Letters*, *26*(12), 2755-2759.

Santos, K. S. D., Passos, A. R., Silva, L. C. C., Silva, A. L. D., and Tanan, T. T. (2022). Genetic variability of *Physalis ixocarpa* and *P. philadelphica* from physicochemical fruit traits. *Pesquisa Agropecuária Brasileira*, *56*, 01534.

Sathyadevi, M., and Subramanian, S. (2015). Extraction, isolation and characterization of bioactive flavonoids from the fruits of *Physalis peruviana Linn* extract. *Asian Journal of Pharmaceutical and Clinical Research*, *8*(1), 152-157.

Schindel, D. E., and Miller, S. E. (2005). DNA barcoding a useful tool for taxonomists. *Nature. 435*, 17.

Seleem, E. A., and Nassar, R. (2021). Morphological and Anatomical Studies on *Physalis peruviana L*. and *Physalis ixocarpa* Brot. Exhornem. *Journal of Plant Production*, *12*(11), 1179-1183.

Shah, P., and Bora, K. S. (2019). Phytochemical and therapeutic potential of *Physalis* species: A review. *IOSR Journal of Pharmaceutical and Biological Sciences*, *14*, 34.

Sharmila, S., Kalaichelvi, K., Rajeswari, M., and Anjanadevi, N. (2014). Studies on the folklore medicinal uses of some indigenous plants among the tribes of Thiashola, Manjoor, Nilgiris South Division, Western Ghats. *International Journal of Plant, Animal and Environmental Sciences*, *4*(3), 14-22.

Shenstone, E., Lippman, Z., and Van Eck, J. (2020). A review of nutritional properties and health benefits of *Physalis* species. *Plant Foods for Human Nutrition*, *75*(3), 316-325.

Shinwari, Z. K., Jan, S. A., Khalil, A. T., Khan, A. D. I. L., Ali, M., Qaiser, M., and Zahra, N. B. (2018). Identification and phylogenetic analysis of selected medicinal plant species from Pakistan: DNA barcoding approach. *Pakistan Journal of Botany*, *50*(2), 553-560.

Shrestha, P. M., and Dhillion, S. S. (2006). Diversity and traditional knowledge concerning wild food species in a locally managed forest in Nepal. *Agroforestry Systems*, *66*, 55-63.

Sica, D. A., Struthers, A. D., Cushman, W. C., Wood, M., Banas Jr, J. S., and Epstein, M. (2002). Importance of potassium in cardiovascular disease. *The Journal of Clinical Hypertension*, *4*(3), 198-206.

Simbaqueba, J., Sanchez, P., Sanchez, E., Nunez Zarantes, V. M., Chacon, M. I., Barrero, L. S., and Mariño-Ramírez, L. (2011). Development and characterization of microsatellite markers for the *Physalis peruviana*. *PloS One*, *6*(10), e26719.

Simeone, M. C., Piredda, R., Papini, A., Vessella, F., and Schirone, B. (2013). Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, *Fagaceae*): problems, prospects and phylogenetic implications. *Botanical Journal of the Linnean Society*, *172*(4), 478-499.

Singh, D. P., Kumar, A., Rodrigues, V., Prabhu, K. N., Kaushik, A., Mani, D. N., and Sundaresan, V. (2020). DNA barcoding coupled with secondary structure information enhances *Achyranthes* species resolution. *Journal of Applied Research on Medicinal and Aromatic Plants*, *19*, 100269.

Singleton, V. L., Orthofer, R., and Lamuela-Raventós, R. M. (1999). Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. In *Methods in Enzymology 299*, 152-178.

Somani, S. J., Modi, K. P., Majumdar, A. S., and Sadarani, B. N. (2015). Phytochemicals and their potential usefulness in inflammatory bowel disease. *Phytotherapy Research*, *29*(3), 339-350.

Tajima F. (1989) a. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics 123*, 585-595.

Tajima, F. (1989) b. The effect of change in population size on DNA polymorphism. *Genetics*, *123*(3), 597-601.

Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* https://doi.org/10.1093/molbev/msab120.

Tan, W. H., Chai, L. C., and Chin, C. F. (2020). Efficacy of DNA barcode internal transcribed spacer 2 (ITS 2) in phylogenetic study of Alpinia species from Peninsular Malaysia. *Physiology and Molecular Biology of Plants*, *26*(9), 1889-1896.

Thitikornpong, W., Palanuvej, C., and Ruangrungsi, N. (2018). DNA barcoding for authentication of the endangered plants in genus Aquilaria. *Thai Journal of Pharmaceutical Sciences (TJPS)*, *42*(4).

Toledo, J. M. (2013). *Physalis victoriana* (Solanaceae) a new species from Northern Argentina. *Phytotaxa*, *124*(1), 60-64.

Tran, T. K. P., Vu, T. T. T., and Widiarsih, S. (2021). Comparison of *matK* and *rbcL* DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. *Journal of Genetic Engineering and Biotechnology*, *19*(1), 1-8.

Tripathi, A. M., Tyagi, A., Kumar, A., Singh, A., Singh, S., Chaudhary, L. B., and Roy, S. (2013). The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. *PloS One*, *8*(2), 57934.

Troshin, P. V., Procter, J. B., and Barton, G. J. (2011). Java bioinformatics analysis web services for multiple sequence alignment—JABAWS: MSA. *Bioinformatics*, *27*(14), 2001-2002.

Troshin, P. V., Procter, J. B., Sherstnev, A., Barton, D. L., Madeira, F., and Barton, G. J. (2018). JABAWS 2.2 distributed web services for Bioinformatics: protein disorder, conservation and RNA secondary structure. *Bioinformatics*, *34*(11), 1939-1940.

Usaizan, N., Abdullah, N. A. P., Ahmad, S. H., and Saleh, G. (2014). Preliminary phytochemical screening and GC-MS analysis of ethanol extract of *Physalis minima* L.(Solanaceae). *Journal of Advanced Agricultural Technologies*, *1*(2), 100-103.

Usaizan, N., Ashikin, N., Abdullah, P., Saleh, G., and Kashiani, P. (2018). Genetic diversity of *Physalis minima L*. accessions based on morphological traits. *Advances in Plants and Agriculture Research*, *8*(2), 151-157.

Vargas, O., Martínez, M., and Dávila, P. (2001). Two new species of *Physalis* (Solanaceae) endemic to Jalisco, Mexico. *Brittonia*, 505-510.

Vargas-Ponce, O., Pérez-Álvarez, L. F., Zamora-Tavares, P., and Rodríguez, A. (2011). Assessing genetic diversity in Mexican husk tomato species. *Plant Molecular Biology Reporter*, *29*, 733-738.

Vargas-Ponce, O., Sánchez Martínez, J., Zamora Tavares, M. D. P., and Valdivia Mares, L. E. (2016). Traditional management of a small-scale crop of *Physalis angulata* in Western Mexico. *Genetic Resources and Crop Evolution*, *63*(8), 1383-1395.

Vdovenko S. A., Polutin O. O., Muliarchuk O. I., Hareba O.V., and Havrys I.L. (2021) Peculiarities of tomatillo (*Physalis philadelphica*) field production in Ukraine with the use of different elements of technology. *Research on Crops* 22(1): 116–128. https://doi.org/10.31830/2348- 7542.2021.044

Wei, J. L., Hu, X. R., Yang, J. J., and Yang, W. C. 2012. Identification of single-copy orthologous genes between P*hysalis* and *Solanum lycopersicum* and analysis of genetic diversity in *Physalis* using molecular markers. *PLoS One*, 7(11): 50164.

Whitson, M., and Manos, P. S. (2005). Untangling *Physalis* (Solanaceae) from the Physaloids: a two-gene phylogeny of the Physalinae. *Systematic Botany*, *30*(1), 216-230.

Wu, S. J., Tsai, J. Y., Chang, S. P., Lin, D. L., Wang, S. S., Huang, S. N., and Ng, L. T. (2006). Supercritical carbon dioxide extract exhibits enhanced antioxidant and anti-inflammatory activities of *Physalis peruviana*. *Journal of Ethnopharmacology*, *108*(3), 407-413.

Xu, J., Zang, F., Wu, Q., Wang, Y., Wang, B., Huang, P., and Zheng, Y. (2021). Analysis of the genetic diversity and molecular phylogeography of the endangered wild rose (*Rosa*

rugosa) in China based on chloroplast genes. *Global Ecology and Conservation*, *28*, 01653.

Yang, F., Ding, F., Chen, H., He, M., Zhu, S., Ma, X., and Li, H. (2018). DNA barcoding for the identification and authentication of animal species in traditional medicine. *Evidence-Based Complementary and Alternative Medicine*, *2018*.

Yao, H., and Song, J. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals, *PLoS One*, *5*, e13102.

Yen C. Y., Chiu C. C., Chang F. R., Chen J. Y., and Hwang C. C. (2010) 4betaHydroxywithanolide E from *Physalis peruviana* (golden berry) inhibits growth of human lung cancer cells through DNA damage, apoptosis and G2/M arrest. *BMC Cancer 10*, 46.

Yin, J., Zhao, H., Wu, X., Ma, Y., Zhang, J., Li, Y., and Xu, Z. (2023). SSR marker-based analysis for identification and of genetic diversity of non-heading Chinese cabbage varieties. *Frontiers in Plant Science*, *14*, 1112748.

Yu, J., Wu, X. I., Liu, C., Newmaster, S., Ragupathy, S., and Kress, W. J. (2021). Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety*, *208*, 111691.

Zamora-Tavares, P., Vargas-Ponce, O., Sánchez-Martínez, J., and Cabrera-Toledo, D. (2015). Diversity and genetic structure of the husk tomato (*Physalis philadelphica Lam*.) in Western Mexico. *Genetic Resources and Crop Evolution*, *62*, 141-153.

Zhang, Y. J., Deng, G. F., Xu, X. R., Wu, S., Li, S., and Li, H. B. (2013). Chemical components and bioactivities of *Physalis* (*Physalis peruviana*). *Internal Journal of Food Nutrition and Safety*, *3*(1), 15-24.

Zhang, W., Yuan, Y., Yang, S., Huang, J., and Huang, L. (2015). ITS2 secondary structure improves discrimination between medicinal “Mu Tong” species when using DNA barcoding. *PLoS One*, *10*(7), 0131185.

Zhang, W., Yang, S., Zhao, H., and Huang, L. (2016). Using the ITS2 sequence-structure as a DNA mini-barcode: A case study in authenticating the traditional medicine "Fang Feng". *Biochemical Systematics and Ecology*, *69*, 188-194.

Zhang, W. N., and Tong, W. Y. (2016). Chemical constituents and biological activities of plants from the genus *Physalis*. *Chemistry and Biodiversity*, *13*(1), 48-65.

Zhang, W., Tian, W., Gao, Z., Wang, G., and Zhao, H. (2020). Phylogenetic utility of rRNA ITS2 sequence-structure under functional constraint. *International Journal of Molecular Sciences*, *21*(17), 6395.

Zhao, Y. (2007). Berry fruit: value-added products for health promotion. CRC press.

Zhao, S., Chen, X., Song, J., Pang, X., and Chen, S. (2015). Internal transcribed spacer 2 barcode: a good tool for identifying Acanthopanacis cortex. *Frontiers in Plant Science*, *6*, 840.

Zhao, L. L., Feng, S. J., Tian, J. Y., Wei, A. Z., and Yang, T. X. (2018). Internal transcribed spacer 2 (ITS 2) barcodes: A useful tool for identifying Chinese Zanthoxylum. *Applications in Plant Sciences*, *6*(6), 01157.

APPENDICES

Appendix 1: *Physalis* accessions sample collection data, PCR and sequencing success rates for ITS 2 and *rbcL* genes

Appendix 1A: Table showing all sample collection areas and their geographical coordinates for *Physalis* accessions used in this study.

No.	Sample ID	County of sample collection	Specific Location of collection	Latitude	Longitude
1.	Londiani L1	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
2.	Londiani L2	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
3.	Londiani L3	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
4.	Londiani L4	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
5.	Londiani L5	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
6.	Londiani L6	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
7.	Londiani L7	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
8.	Londiani L8	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
9.	Londiani L9	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
10.	Londiani L10	Kericho	Londiani-Sorget	0.0532° S	35.5548° E
11.	Chebororwa C1	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
12.	Chebororwa C2	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
13.	Chebororwa C3	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
14.	Chebororwa C4	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
15.	Chebororwa C5	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
16.	Chebororwa e1in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
17.	Chebororwa e2in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
18.	Chebororwa e3in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
19.	Chebororwa e4in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
20.	Chebororwa e5in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
21.	Chebororwa e6in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
22.	Chebororwa e7in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
23.	Chebororwa e8in	Elgeyo Marakwet	Chebororwa	0.9487° N	35.4234° E
24.	Ndhiwa hb1	Homabay	Ndhiwa	0.7299° S	34.3671° E
25.	Ndhiwa hb2	Homabay	Ndhiwa	0.7299° S	34.3671° E
26.	Ndhiwa hb3	Homabay	Ndhiwa	0.7299° S	34.3671° E
27.	Mukurweini ny1	Nyeri	Mukurweini	0.5609° S	37.0488° E
28.	Mukurweini ny2	Nyeri	Mukurweini	0.5609° S	37.0488° E
29.	Mukurweini ny3	Nyeri	Mukurweini	0.5609° S	37.0488° E
30.	Mukurweini ny4	Nyeri	Mukurweini	0.5609° S	37.0488° E
31.	Mukurweini ny5	Nyeri	Mukurweini	0.5609° S	37.0488° E
32.	Ongata Rongai nor1	Kajiado	Ongata Rongai	1.3939° S	36.7442° E
33.	Ongata Rongai nor2	Kajiado	Ongata Rongai	1.3939° S	36.7442° E
34.	Ongata Rongai nor3	Kajiado	Ongata Rongai	1.3939° S	36.7442° E
35.	Ongata Rongai nor4	Kajiado	Ongata Rongai	1.3939° S	36.7442° E
36.	Ongata Rongai nor5	Kajiado	Ongata Rongai	1.3939° S	36.7442° E
37.	Gilgil gm	Nakuru	Gilgil market	0.4923° S	36.3173° E
38.	Muguga DM	Kiambu	Muguga	1.2551° S	36.6580° E
39.	Thika TK1	Kiambu	Thika	1.0388° S	37.0834° E

40.	Thika TK2	Kiambu	Thika	1.0388° S	37.0834° E
41.	Thika TK3	Kiambu	Thika	1.0388° S	37.0834° E
42.	Thika TK4	Kiambu	Thika	1.0388° S	37.0834° E
43.	Thika TK5	Kiambu	Thika	1.0388° S	37.0834° E
44.	Thika TK6	Kiambu	Thika	1.0388° S	37.0834° E
45.	Thika TK7	Kiambu	Thika	1.0388° S	37.0834° E
46.	Thika TK8	Kiambu	Thika	1.0388° S	37.0834° E
47.	Thika TK9	Kiambu	Thika	1.0388° S	37.0834° E
48.	Tigoni T1	Kiambu	Tigoni	1.1651° S	36.7065° E
49.	Tigoni T2	Kiambu	Tigoni	1.1651° S	36.7065° E
50.	Tigoni T3	Kiambu	Tigoni	1.1651° S	36.7065° E
51.	Tigoni T4	Kiambu	Tigoni	1.1651° S	36.7065° E
52.	Tigoni T5	Kiambu	Tigoni	1.1651° S	36.7065° E
53.	Tigoni T6	Kiambu	Tigoni	1.1651° S	36.7065° E
54.	Tigoni T7	Kiambu	Tigoni	1.1651° S	36.7065° E
55.	Tigoni T8	Kiambu	Tigoni	1.1651° S	36.7065° E
56.	Tigoni T9	Kiambu	Tigoni	1.1651° S	36.7065° E
57.	Tigoni T10	Kiambu	Tigoni	1.1651° S	36.7065° E
58.	Tigoni T11	Kiambu	Tigoni	1.1651° S	36.7065° E
59.	Tigoni T12	Kiambu	Tigoni	1.1651° S	36.7065° E
60.	Tigoni T13	Kiambu	Tigoni	1.1651° S	36.7065° E
61.	Tigoni T14	Kiambu	Tigoni	1.1651° S	36.7065° E
62.	Tigoni T15	Kiambu	Tigoni	1.1651° S	36.7065° E
63.	Tigoni T16	Kiambu	Tigoni	1.1651° S	36.7065° E
64.	Tigoni T17	Kiambu	Tigoni	1.1651° S	36.7065° E
65.	Kahuho KK1	Kiambu	Kahuho	1.2082° S	36.6795° E
66.	Kahuho KK2	Kiambu	Kahuho	1.2082° S	36.6795° E
67.	Banana Hill JCBK	Kiambu	Banana Hill	1.1760° S	36.7550° E
68.	Tala KT1	Machakos	Tala	1.2670° S	37.3201° E
69.	Tala KT2	Machakos	Tala	1.2670° S	37.3201° E
70.	Tala KT3	Machakos	Tala	1.2670° S	37.3201° E
71.	Tala KT4	Machakos	Tala	1.2670° S	37.3201° E
72.	Kilimani KL1	Nairobi	Kilimani	1.2893° S	36.7869° E
73.	Kilimani KL2	Nairobi	Kilimani	1.2893° S	36.7869° E
74	Kilimani ANR	Nairobi	Kilimani	1.2893° S	36.7869° E
75.	Kilimani SHR	Nairobi	Kilimani	1.2893° S	36.7869° E
76.	Valley Arcade Van1	Nairobi	Kilimani	1.2907° S	36.7692° E
77.	Valley Arcade Van 2	Nairobi	Kilimani	1.2907° S	36.7692° E
78.	Maji Mazuri mm	Baringo	Maji Mazuri	0.0076° S	35.6861° E

Appendix 1B: Amplification and sequencing rates for the ITS2 barcode gene of *Physalis* accessions

No	PCR amplicon ID	Sequencing success	Sequence length (base pairs)	GC content (%)
1.	Londiani L1i	Not successful	N/A	N/A
2.	Londiani L2i	Successful	466	61.6
3.	Londiani L3i	Successful	602	62.1
4.	Londiani L4i	Successful	663	65.2
5.	Londiani L5i	Successful	656	61
6.	Londiani L6i	Successful	310	60
7.	Londiani L7i	Successful	511	60.3
8.	Londiani L8i	Successful	642	60.4
9.	Londiani L9i	Successful	624	59.6
10.	Londiani L10i	Successful	578	59.9
11.	Chebororwa C2i	Successful	679	61.1
12.	Chebororwa C5i	Successful	469	61.6
13.	Chebororwa e1ini	Successful	373	62.5
14.	Chebororwa e2ini	Successful	383	61.4
15.	Chebororwa e3ini	Successful	683	60.6
16.	Chebororwa e4ini	Successful	508	60.2
17.	Chebororwa e5ini	Successful	707	61
18.	Chebororwa e6ini	Successful	664	60.4
19.	Chebororwa e7ini	Successful	692	60.8
20.	Magwagwa mw3i	Successful	688	65.6
21.	Ndhiwa hb2i	Successful	676	60.8
22.	Ndhiwa hb3i	Not successful	N/A	N/A
23.	Mukurweini ny4i	Not successful	N/A	N/A
24.	Mukurweini ny5i	Successful	514	66.9
25	Gilgil gmi	Successful	274	59.5
26.	Nyathuna NKi	Successful	314	55.1
27.	Muguga DMi	Successful	595	61.7
28.	Thika TK1i	Not successful	N/A	N/A
29.	Thika TK2i	Not successful	N/A	N/A
30.	Thika TK3i	Not successful	N/A	N/A
31.	Thika TK4i	Not successful	N/A	N/A
32.	Thika TK5i	Not successful	N/A	N/A
33.	Thika TK6i	Not successful	N/A	N/A
34.	Thika TK7i	Not successful	N/A	N/A
35.	Thika TK8i	Successful	310	61
36.	Thika TK9i	Successful	272	62.5
37.	Tigoni T1i	Not successful	N/A	N/A
38.	Tigoni T2i	Successful	431	59.6
39.	Tigoni T4i	Successful	274	60.6
40.	Tigoni T6i	Not successful	N/A	N/A
41.	Tigoni T7i	Not successful	N/A	N/A
42.	Tigoni T8i	Not successful	N/A	N/A
43.	Tigoni T9i	Successful	545	60.4
44.	Tigoni T10i	Not successful	N/A	N/A
45.	Tigoni T11i	Successful	237	55.7
46.	Tigoni T12i	Not successful	N/A	N/A

47.	Tigoni T17i	Not successful	N/A	N/A
48.	Ongata Rongai nor1i	Successful	679	60.2
49.	Ongata Rongai nor3i	Successful	415	63.6
50.	Ongata Rongai nor4i	Successful	677	60.4
51.	Ongata Rongai nor5i	Successful	683	60.8

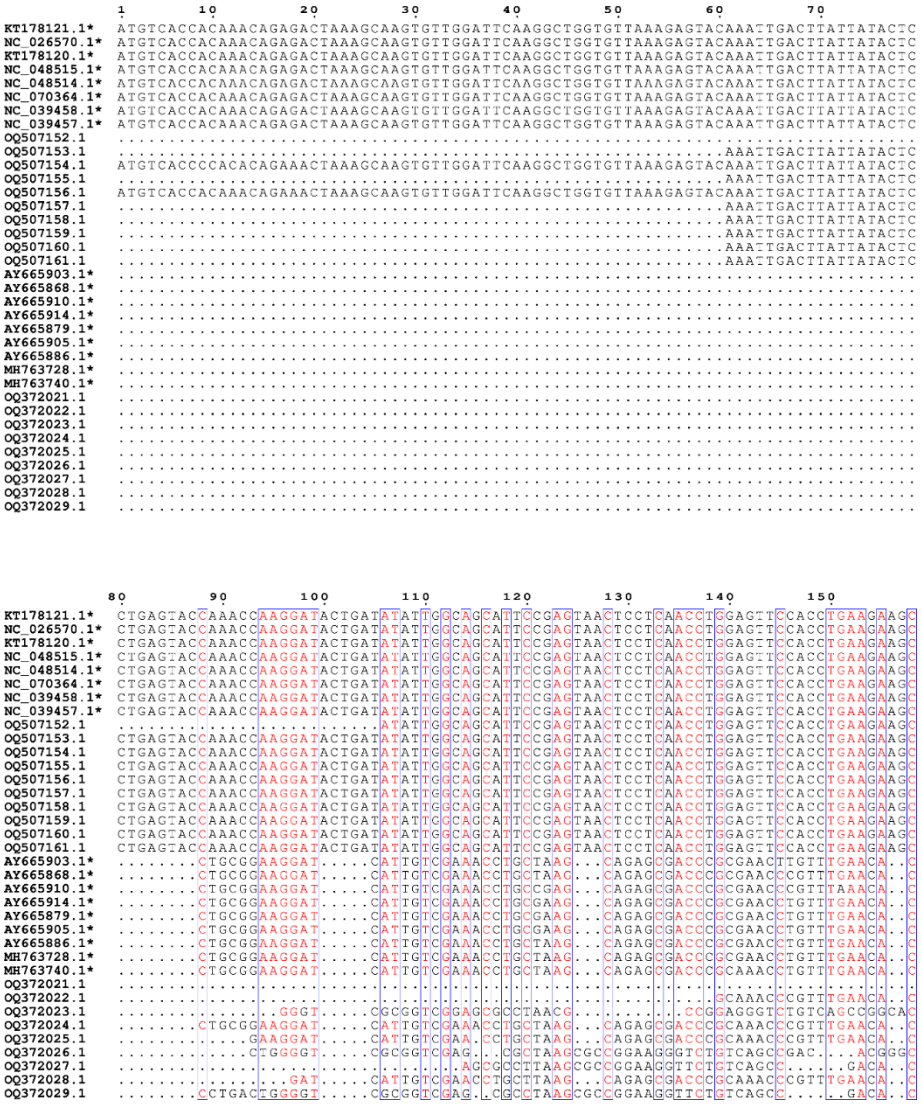
Appendix 1 C: Amplification and sequencing rates for the *rbcL* barcode gene of *Physalis* accessions

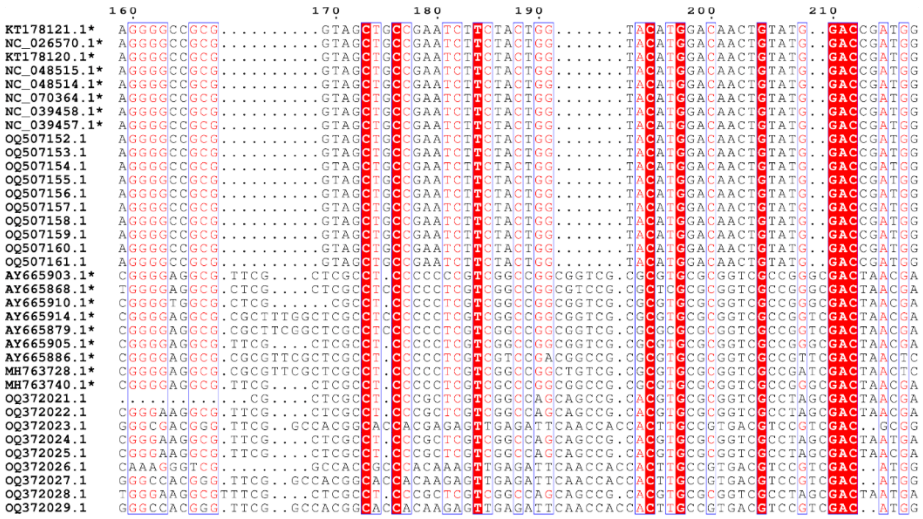
No	PCR amplicon ID	Sequencing success	Sequence length (base pairs)	GC content %
1.	Londiani L1r	Successful	547	43.9
2.	Londiani L2r	Successful	615	43.3
3.	Londiani L3r	Successful	733	43.1
4.	Londiani L4r	Successful	520	43.1
5.	Londiani L5r	Successful	730	42.7
6.	Londiani L6r	Successful	634	43.1
7.	Londiani L7r	Successful	615	43.3
8.	Londiani L8r	Successful	621	42.8
9.	Londiani L9r	Successful	543	43.1
10.	Londiani L10r	Successful	604	43
11.	Chebororwa C2r	Successful	732	45.5
12.	Chebororwa C4r	Successful	745	43.5
13.	Chebororwa C5r	Successful	463	44.5
14.	Chebororwa e1inr	Successful	513	45.2
15.	Chebororwa e2inr	Not successful	N/A	N/A
16.	Chebororwa e3inr	Not successful	N/A	N/A
17.	Chebororwa e4inr	Successful	743	42.8
18	Chebororwa e6inr	Successful	815	43.2
19	Chebororwa e7inr	Successful	745	43.9
20.	Chebororwa e8inr	Successful	717	43.5
21.	Ndhiwa hb1r	Not successful	N/A	N/A
22.	Ndhiwa hb2r	Successful	745	43
23.	Ndhiwa hb3r	Successful	744	43
24.	Mukurweini ny1r	Successful	854	42.3
25.	Mukurweini ny3r	Not successful	N/A	N/A
26.	Mukurweini ny4r	Successful	724	44.8
27.	Mukurweini ny5r	Successful	712	44.7
28.	Gilgil gmr	Successful	637	42.9
29.	Maji Mazuri mmr	Successful	565	45.5
30.	Muguga DMr	Successful	475	44
31.	Thika TK1r	Not successful	N/A	N/A
32.	Thika TK2r	Successful	733	43.4
33.	Thika TK3r	Successful	792	43.2
34.	Thika TK4r	Successful	815	43.4
35.	Thika TK5r	Not successful	N/A	N/A
36.	Thika TK6r	Successful	789	43.5

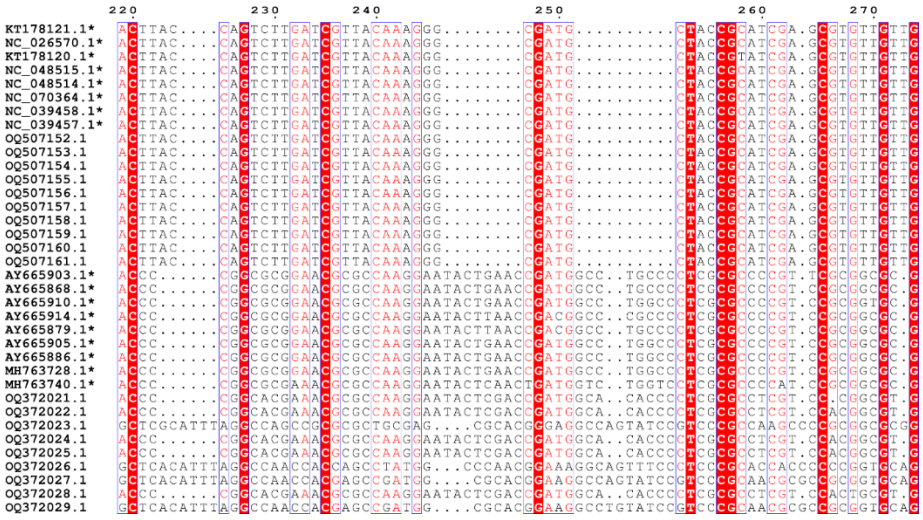
37.	Thika TK7r	Successful	733	42.8
38.	Thika TK8r	Successful	626	43.1
39.	Thika TK9r	Successful	815	43.1
40.	Tigoni T2r	Successful	582	43.1
41.	Tigoni T4r	Successful	841	43.6
42.	Tigoni T6r	Successful	464	44.4
43.	Tigoni T7r	Successful	807	43.1
44.	Tigoni T8r	Successful	540	43.7
45.	Tigoni T9r	Successful	802	42.8
46.	Tigoni T10r	Successful	540	45
47.	Tigoni T11r	Successful	745	43.5
48.	Tigoni T12r	Successful	545	43.1
49.	Tigoni T13r	Successful	773	43.5
50.	Tigoni T14r	Successful	775	42.7
51.	Tigoni T16r	Successful	742	42.3
52.	Tigoni T17r	Successful	803	43.3
53.	Kahuho KK1r	Successful	553	42.9
54.	Banana Hill JCBK	Successful	729	44.2
55.	Tala KT1r	Successful	668	43.9
56.	Ongata Rongai nor3r	Successful	818	43.9
57.	Ongata Rongai nor4r	Successful	776	42.1
58.	Ongata Rongai nor5r	Successful	745	43.2
59.	Kilimani KL1r	Successful	744	43
60.	Kilimani KL2r	Not successful	N/A	N/A
61.	Kilimani SHRr	Successful	703	43.1
62.	Kilimani ANRr	Successful	722	42.5
63.	Valley Arcade Van 1	Successful	847	43.3
64.	Valley Arcade Van 2	Not successful	N/A	N/A

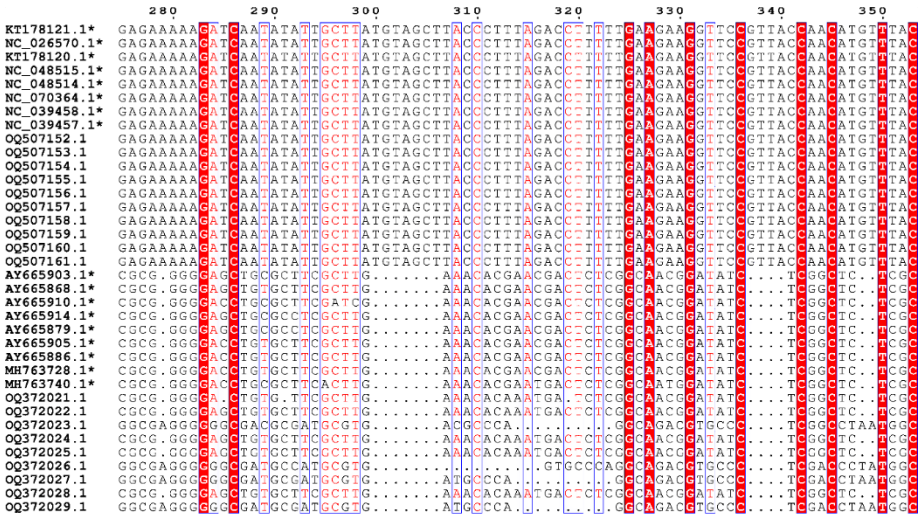
Appendix 2: Multiple sequences alignments for *Physalis* accessions.

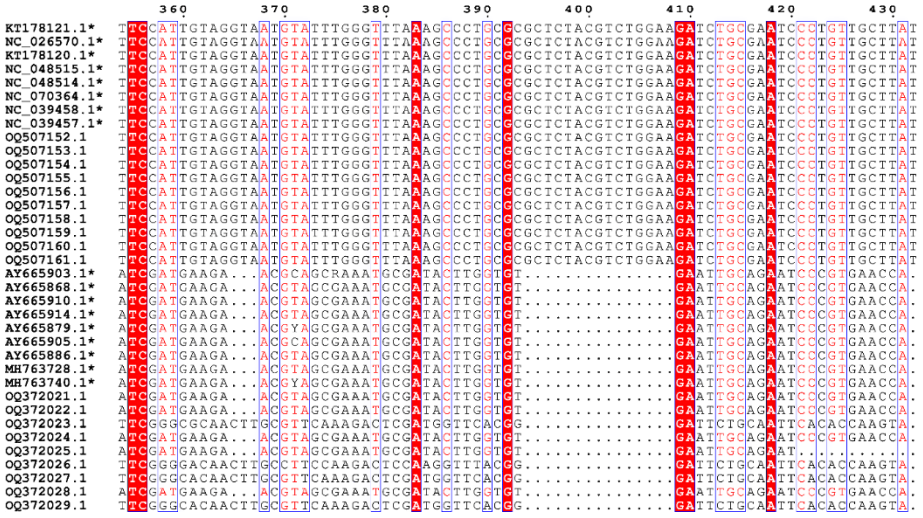
Appendix 2 A: Multiple sequence alignment of indigenous *Physalis* accessions from Sorget Forest based on their ITS2 and *rbcL* gene sequences as well as their reference sequences based on BLASTn analysis (https://espript.ibcp.fr/ESPript/temp/1818399141/0-0-1680464456-esp.pdf)

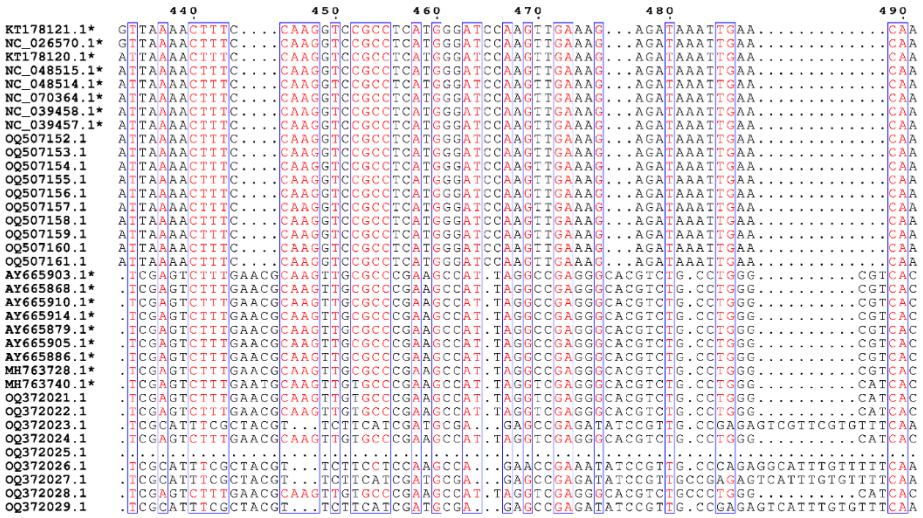


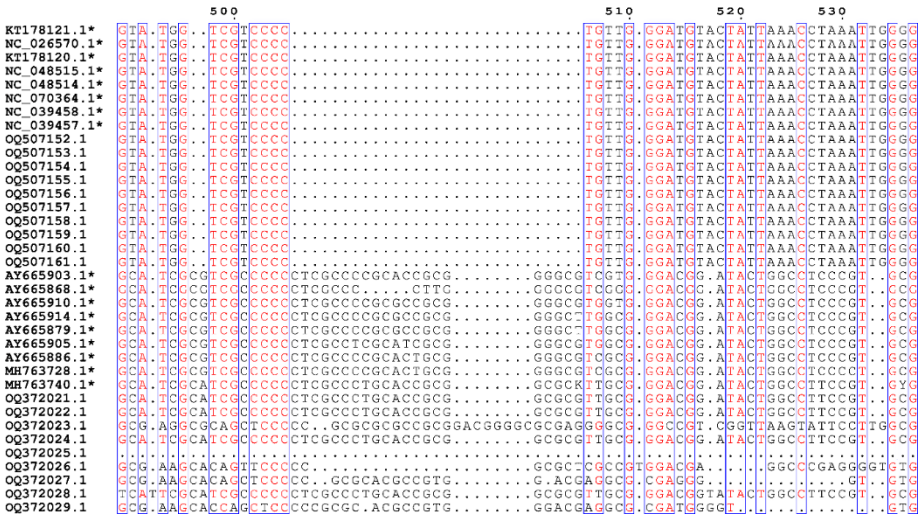


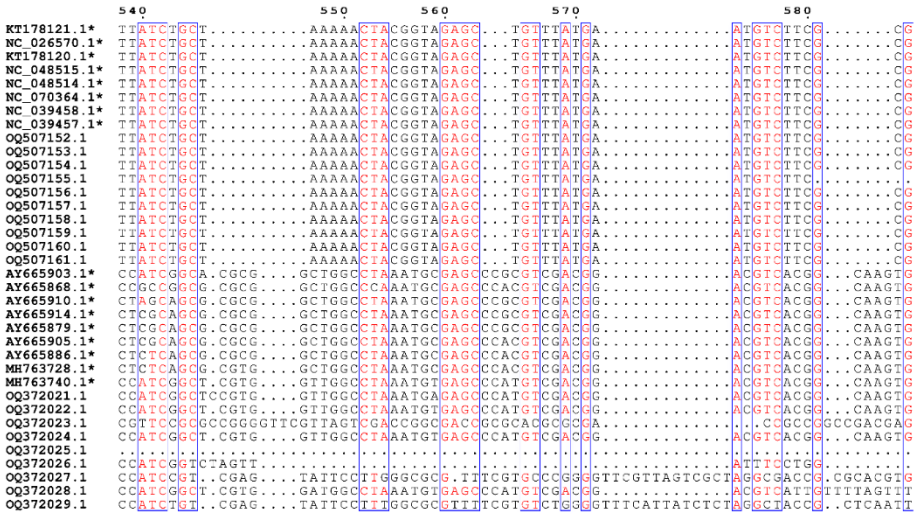


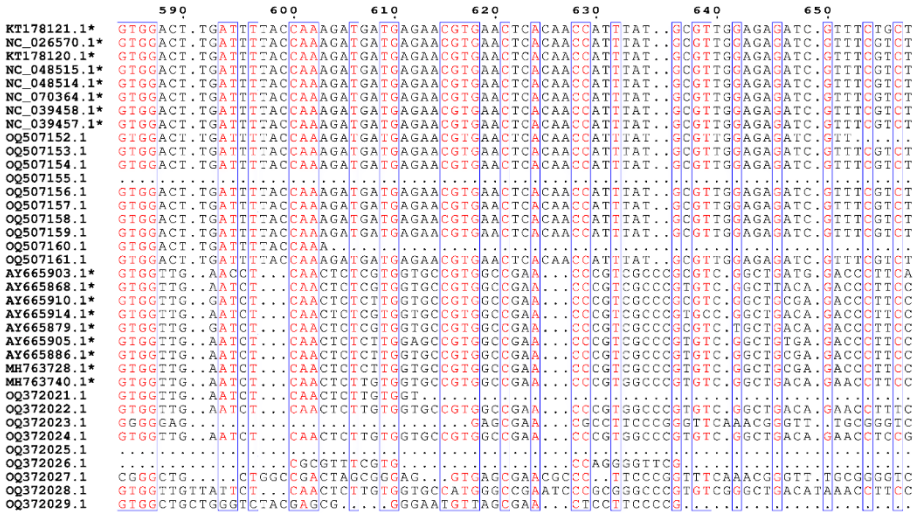


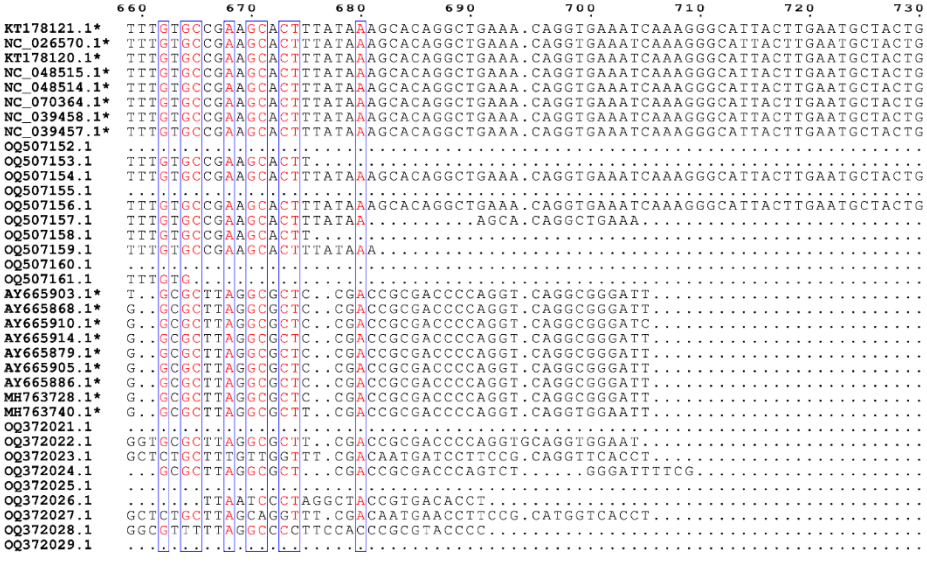




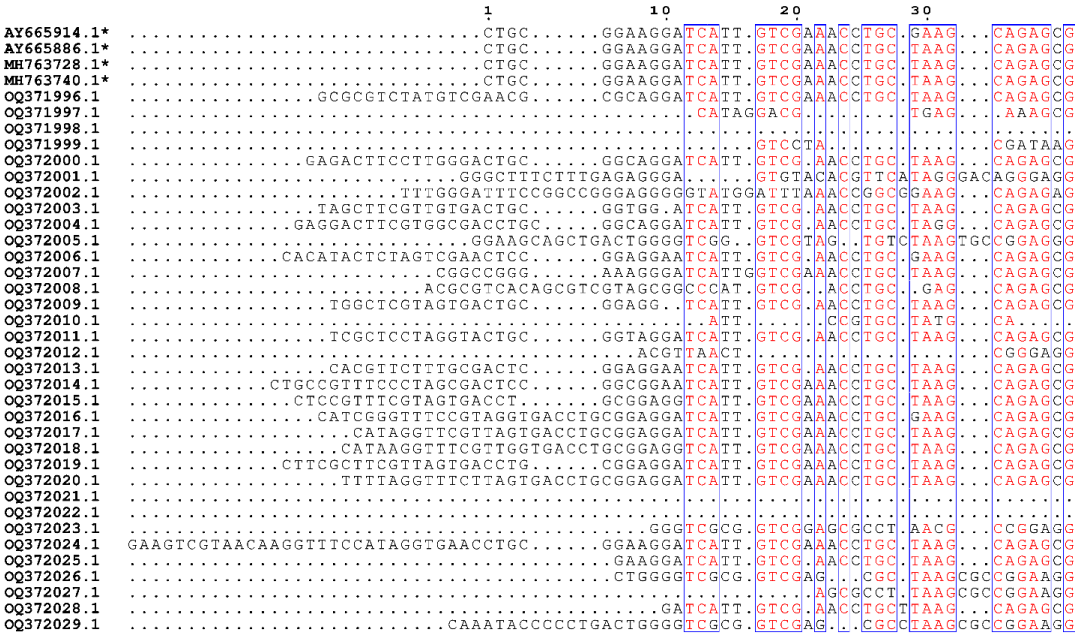


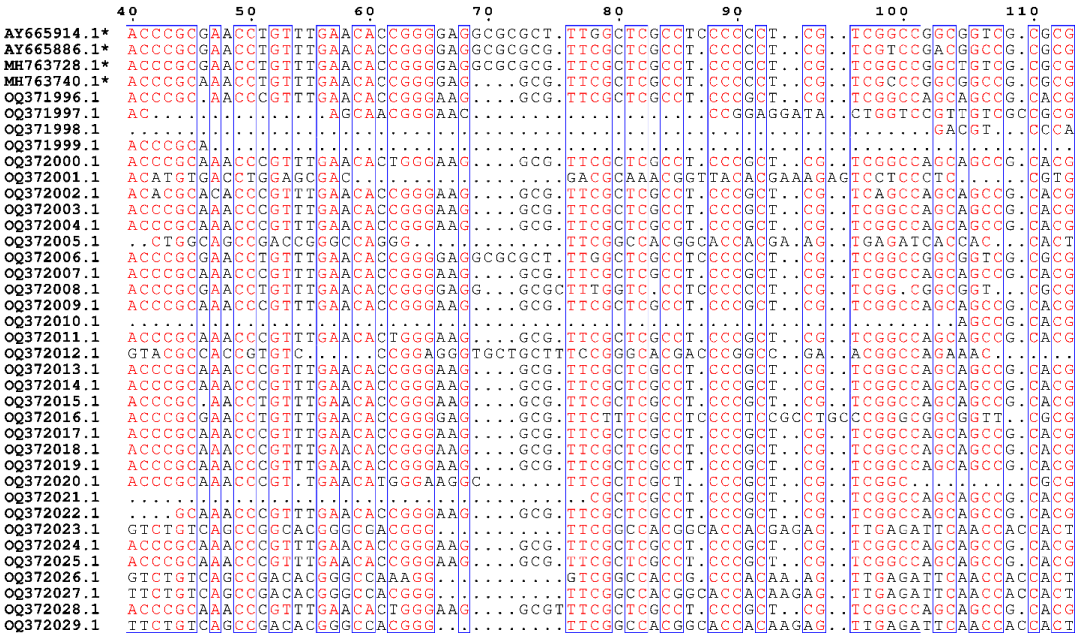


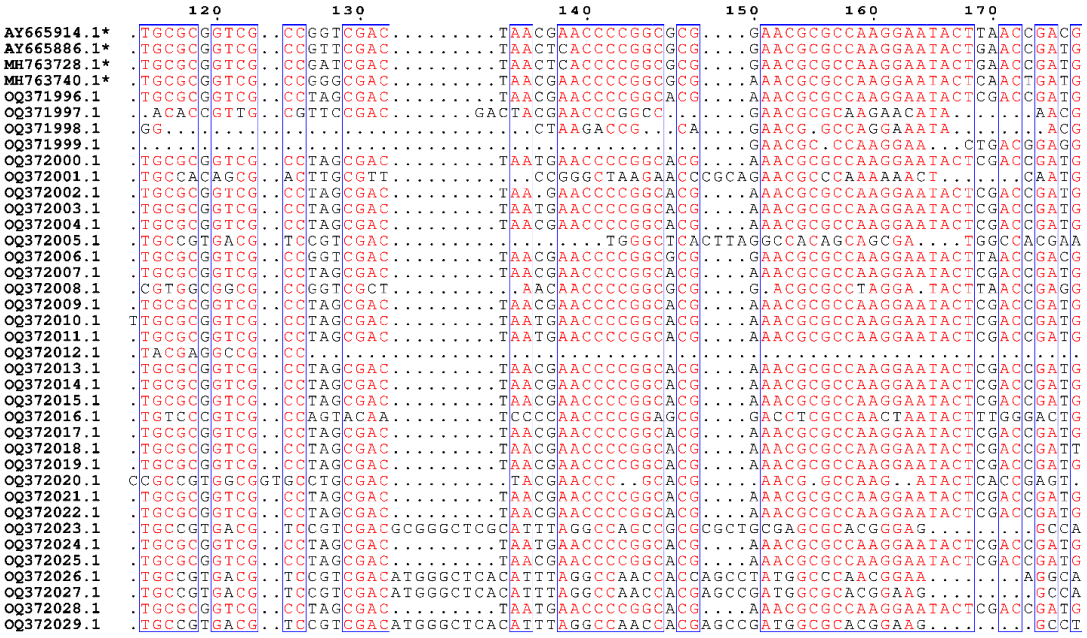


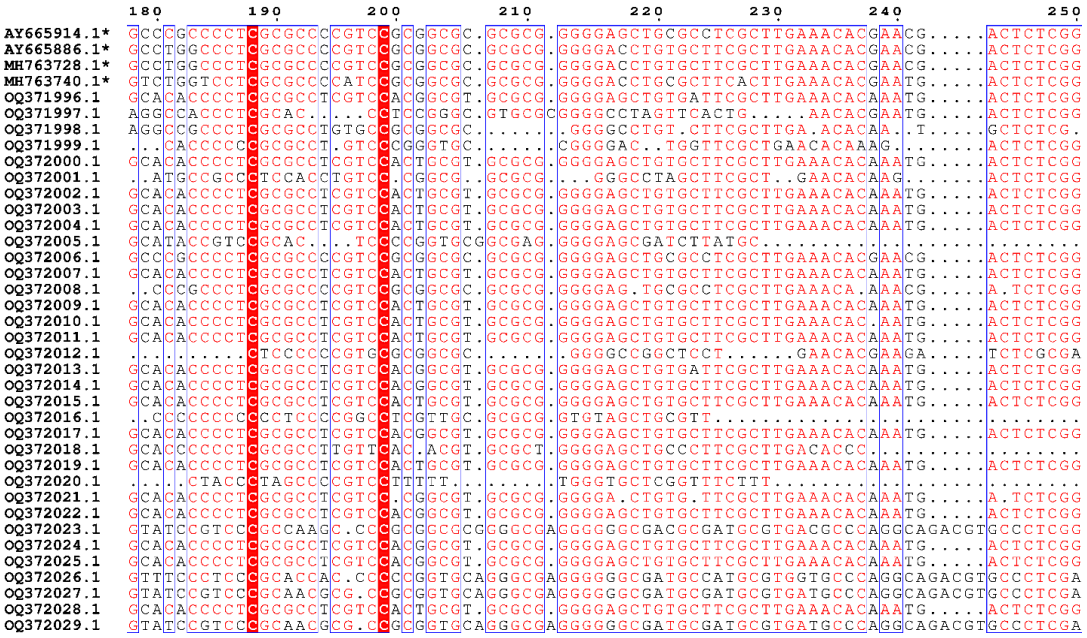


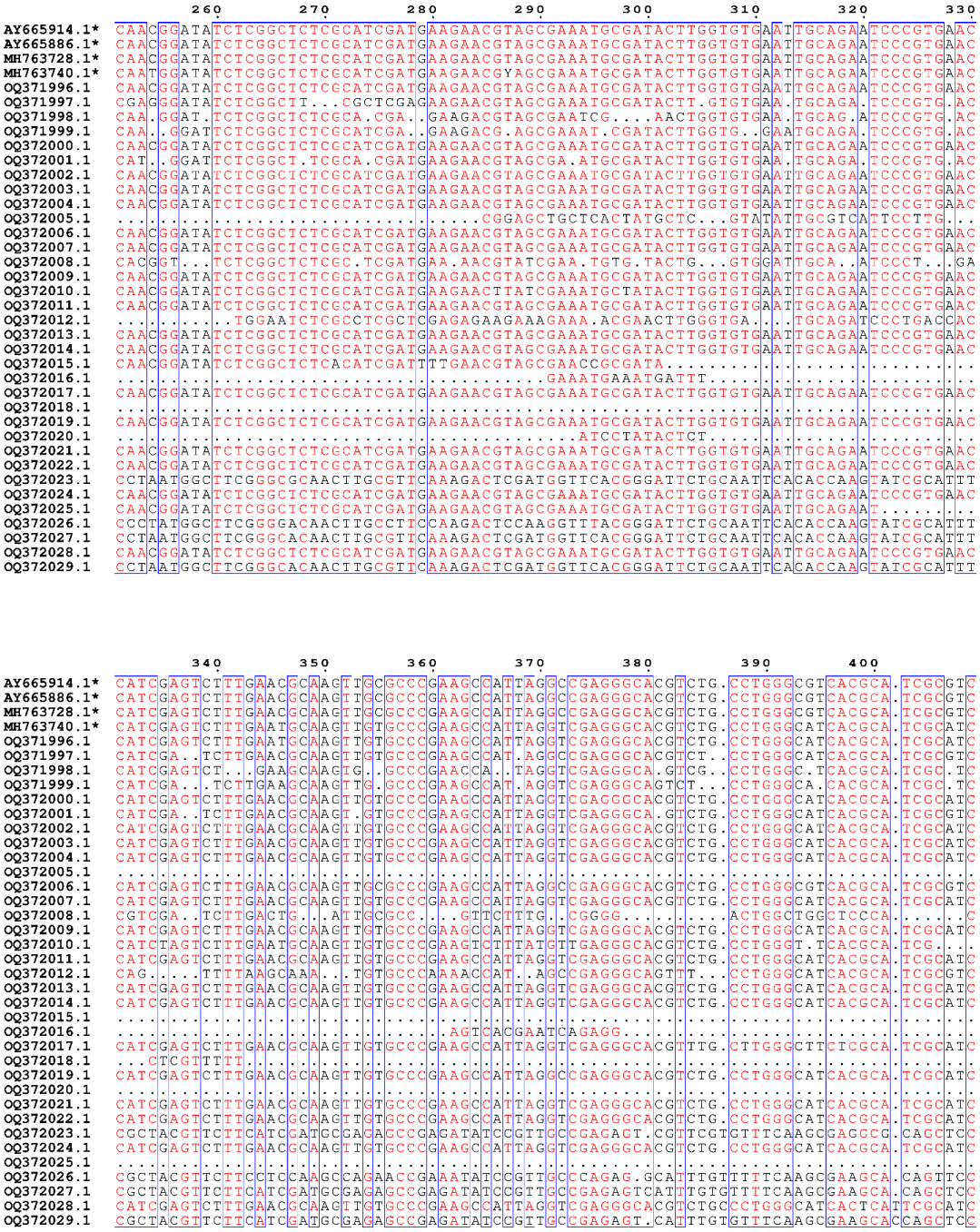
Appendix 2 B: Multiple sequence alignment (MSA) for ITS2 sequences of the *Physalis* accessions as well as reference sequence (https://espript.ibcp.fr/ESPript/temp/1101891838/0-0-1680467018-esp.pdf)



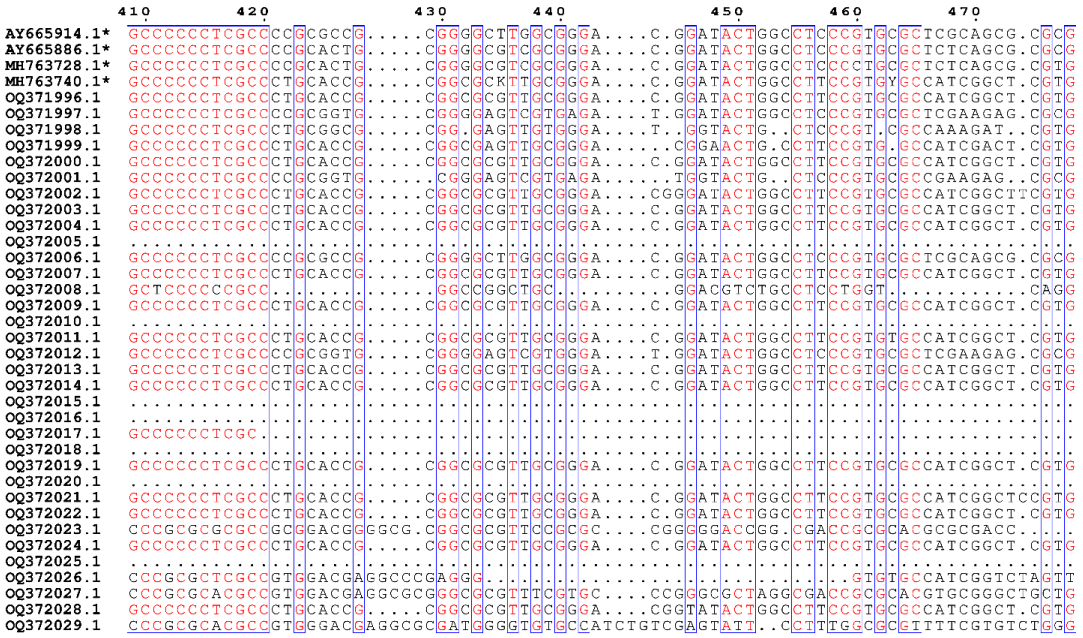


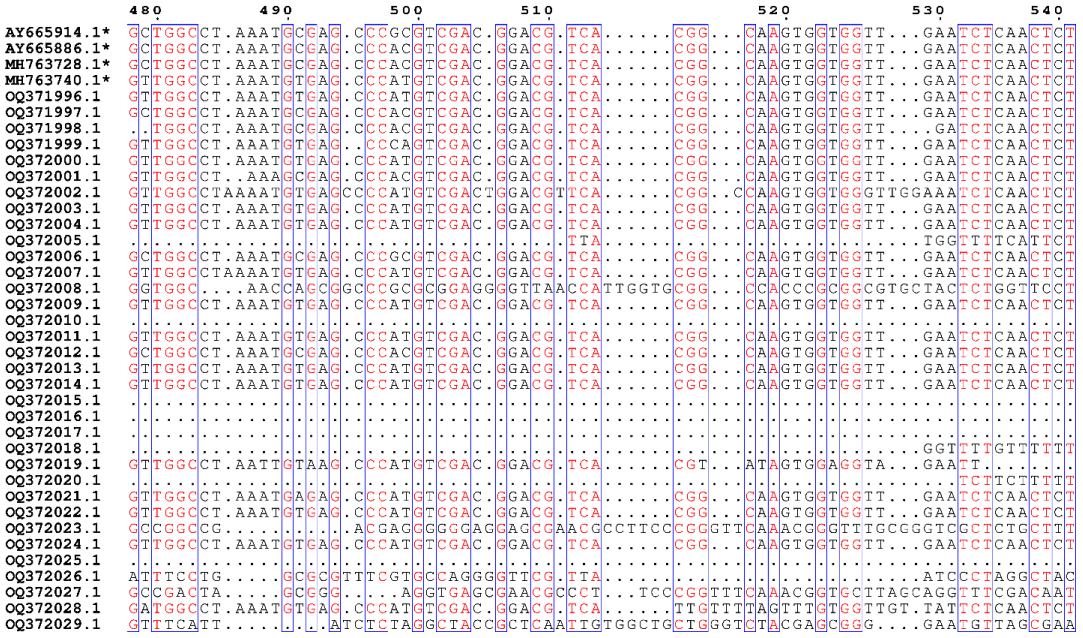





260 270 280 290 300 310 320 330
AY665914.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ371999.1
OQ372000.1
OQ372001.1
OQ372002.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372006.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372010.1
OQ372011.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

340 350 360 370 380 390 400
AY665914.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ371999.1
OQ372000.1
OQ372001.1
OQ372002.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372006.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372010.1
OQ372011.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

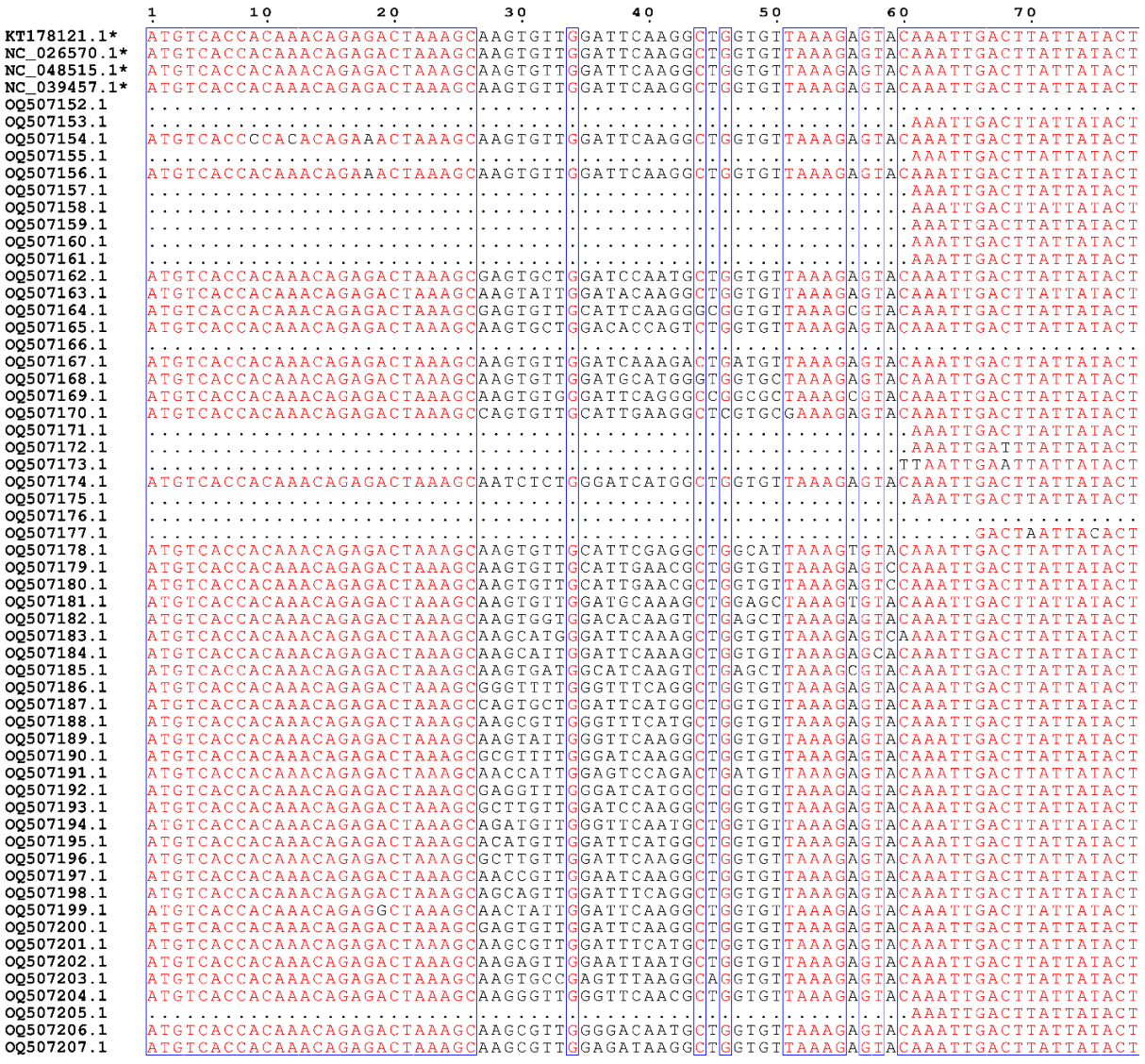

410 420 430 440 450 460 470
AY665914.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ371999.1
OQ372000.1
OQ372001.1
OQ372002.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372006.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372010.1
OQ372011.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

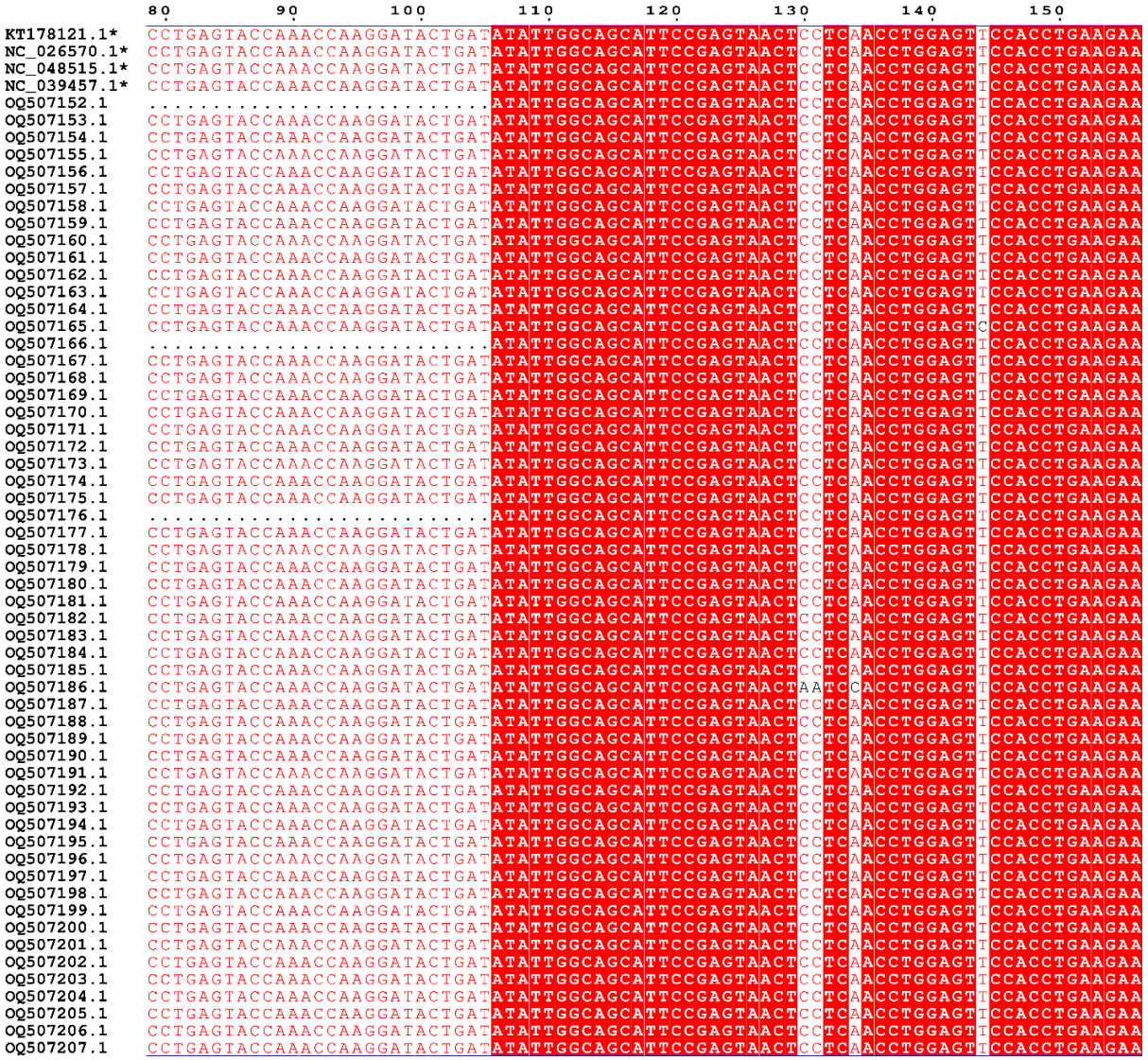

480 490 500 510 520 530 540
AY665914.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ371999.1
OQ372000.1
OQ372001.1
OQ372002.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372006.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372010.1
OQ372011.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

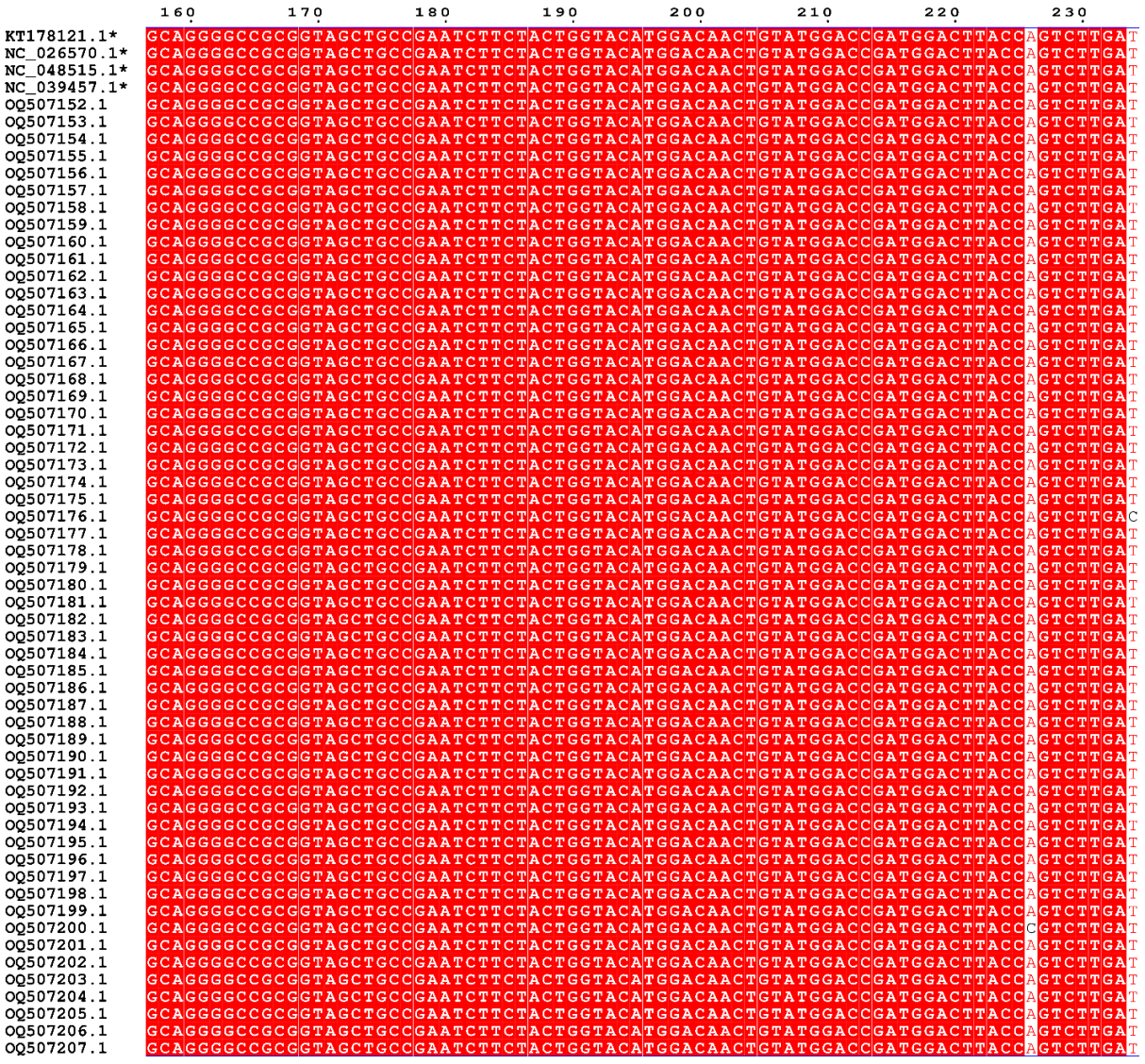
```
                 550         560         570         580         590         600
AY665914.1*  CGT.GGTGCCG..TGGCCGAACCCGTCG..CCCGTGCC.GGCTGACA...GACCCTTCCG..GCGCTTAGG...CGCTCC
AY665886.1*  CTT.GGTGCCG..TGGCCGAACCCGTCG..CCCGTGTC.GGCTGCGA...GACCCTTCCG..GCGCTTAGG...CGCTCC
MH763728.1*  CTT.GGTGCCG..TGGCCGAACCCGTCG..CCCGTGTC.GGCTGCGA...GACCCTTCCG..GCGCTTAGG...CGCTCC
MH763740.1*  TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCG..GCGCTTAGG...CGCTTC
OQ371996.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCCG.GCGCTTAGG...CGCTTC
OQ371997.1   CTT.GGTGC.....................................................................
OQ371998.1   CTT.GGTGCCGT....................................................................
OQ371999.1   TGT.GGTGCCGTG...................................................................
OQ372000.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCG..GCGCTTAGG...CGCTTC
OQ372001.1   CTT.GGTGCCG..TGG................................................................
OQ372002.1   TGTGGGTGCCGGTGGGCCGAACCCGTGGGCCCGGTGTC.GGCTGGAACATAAACCTTTCCGGGCGCTTTAGGGCGGCTTC
OQ372003.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCGGCGC...TTAG...CGCTTC
OQ372004.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTTCCG.GCGCTTAGG...CGCTTC
OQ372005.1   TTT..ATTCAT.....................................................................
OQ372006.1   CGT.GGTGCCG..TGGCCGAACCCGTCG..CCCGTGCC.GGCTGACA...GACCCTTCCG..GCGCTTAGG...CGCTCC
OQ372007.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTCGGGCTGACA...GAACCTTTCCG.................
OQ372008.1   CACCAGCGGCA.....................................................................
OQ372009.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCG..GCGCTTAGG...CGCTTC
OQ372010.1   ...............................................................................
OQ372011.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTCCG..GCGCTTAGG...CGCTTC
OQ372012.1   CTT.GGTGCCG.....................................................................
OQ372013.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.AGCTGACA...GAACCTTCCG..GCGCTTAGG...CGCTTC
OQ372014.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.AGCTGACA...GAACCTTCCGGCGC...TTAG...CGCTTC
OQ372015.1   ...............................................................................
OQ372016.1   ...............................................................................
OQ372017.1   ...............................................................................
OQ372018.1   CGT............................................................................
OQ372019.1   ...............................................................................
OQ372020.1   CTT............................................................................
OQ372021.1   TGT.GGT........................................................................
OQ372022.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTTTCGGTGCGCTTAGG...CGCTTC
OQ372023.1   GTT.GGTTTCG.....................ACAATGATCCTTCCGCAGGTTCACCTATCGAAACCTT.......TGTTAC
OQ372024.1   TGT.GGTGCCG..TGGCCGAACCCGTGG..CCCGTGTC.GGCTGACA...GAACCTCCG...GCGCTTAGG...CGC.TC
OQ372025.1   ...............................................................................
OQ372026.1   CGT.GACACCT.....................................................................
OQ372027.1   GAA.CCTTCCGCATGGTCACCTACGGGAAACTTTTGTT..........TTGAACTTCCTC....................
OQ372028.1   TGT.GGTGCCA.TGGGCCGAATCCCGCGGGCCCGTGTCGGGCTGACA..TAAACCTTCCGGCGTTTTTAGGC..CCCTTC
OQ372029.1   CTCCTTCCCG......................................................................
```

```
             610         620         630
AY665914.1*  GACC..GCGACCCC.....AGGTCAGGCGGGATT.....................................................
AY665886.1*  GACC..GCGACCCC.....AGGTCAGGCGGGATT.....................................................
MH763728.1*  GACC..GCGACCCC.....AGGTCAGGCGGGATT.....................................................
MH763740.1*  GACC..GCGACCCC.....AGGTCAGGTGGAATT.....................................................
OQ371996.1   GACC..GCGACCCC.....AGGT.CAGGTGGAATTACCCGCTGAGTTTAAG..CATATCAATAGGCGGAGGGA.......
OQ371997.1   ...............................................................................
OQ371998.1   ...............................................................................
OQ371999.1   ...............................................................................
OQ372000.1   GACC..GCGACCCC.....AGGTCAGGTGGAAATTACCCGCTGAGTTTAAG..CATATCAATAAGGCGGGAGGGAA....
OQ372001.1   ...............................................................................
OQ372002.1   GAACCGGCGACCCCCCAGGGTGCAGGGTGGGAATTTAACCCCGGGCTTGAA..GGTTTTAATGCCCATAA..........
OQ372003.1   GA.C..GCGACCCA.....GGTC....AGGTGATTACCCGCTGAGTTAAG...CATAATGAATAGC.............
OQ372004.1   GACC..GCGACCCC.....AGGTCAGGTGGAAATTACCCCGCTGAGTTTAAGCCATATCAATAAGCGGGAAGGAAATTTC
OQ372005.1   ...............................................................................
OQ372006.1   GACC..GCGACCCC.....AGGTCAGGCGGGATTACCCGCTGAGTTTAAGCATATCAATAAAGCGGGAGGAA........
OQ372007.1   ...............................................................................
OQ372008.1   ......GTGGATCCGTATTTGTTATAGGGAGA.............................................
OQ372009.1   GACC..GCGACCCC.....AGGT..CAGTGGAATTACCCGCTGAGTTTAAG..CATATCAATAAGGCGGAGGAAA.....
OQ372010.1   ...............................................................................
OQ372011.1   GACC..GCGACCCC.....AGGT.CAGGTGGAATTACCCGCTGAGTTTAAG..CATATCAATAAGCGGGAGGAAA.....
OQ372012.1   ...............................................................................
OQ372013.1   GACC..GCGACCCC.....AGGT..CAGTGGAATTACCCGCTGAGTTTAAG..CATATCAATAAGCCGGAGGAA......
OQ372014.1   GACC..GCGACCCC.....AGGT..CAGTGGAATTACCCGCTGAGTTTAAG..CATATCAATAAGCCGGAGGAAA.....
OQ372015.1   ...............................................................................
OQ372016.1   ...............................................................................
OQ372017.1   ...............................................................................
OQ372018.1   ...............................................................................
OQ372019.1   ...............................................................................
OQ372020.1   ...............................................................................
OQ372021.1   ...............................................................................
OQ372022.1   GACC..GCGACCCC.....AGGTGCAGGTGGAATTACCCCGCTGAGTT.............................
OQ372023.1   GAACTGGGGCCTTCTTTTAAGGGATAAGG.................................................
OQ372024.1   GACC..GCGACCCA.....GTCT......GGGATTTTCG.......................................
OQ372025.1   ...............................................................................
OQ372026.1   ...............................................................................
OQ372027.1   ...............................................................................
OQ372028.1   CACCCGCGTACCCC.................................................................
OQ372029.1   ...............................................................................
```

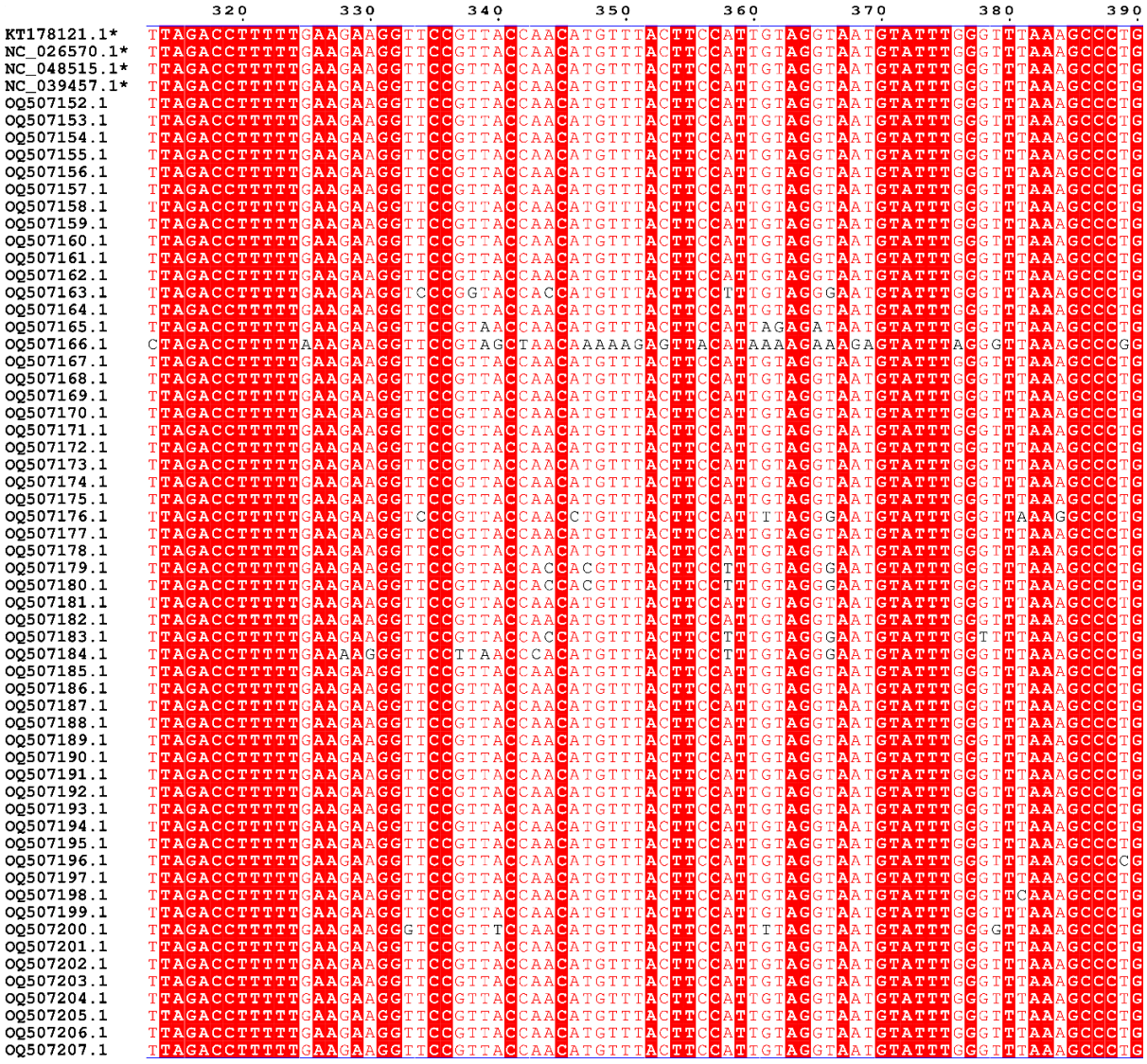
Appendix 2 C: Multiple alignment sequence for *Physalis* accessions gene sequence as well as reference sequences from BLASTn analysis based on *rbcL* marker (https://espript.ibcp.fr/ESPript/temp/1694482275/0-0-1680468378-esp.pdf).



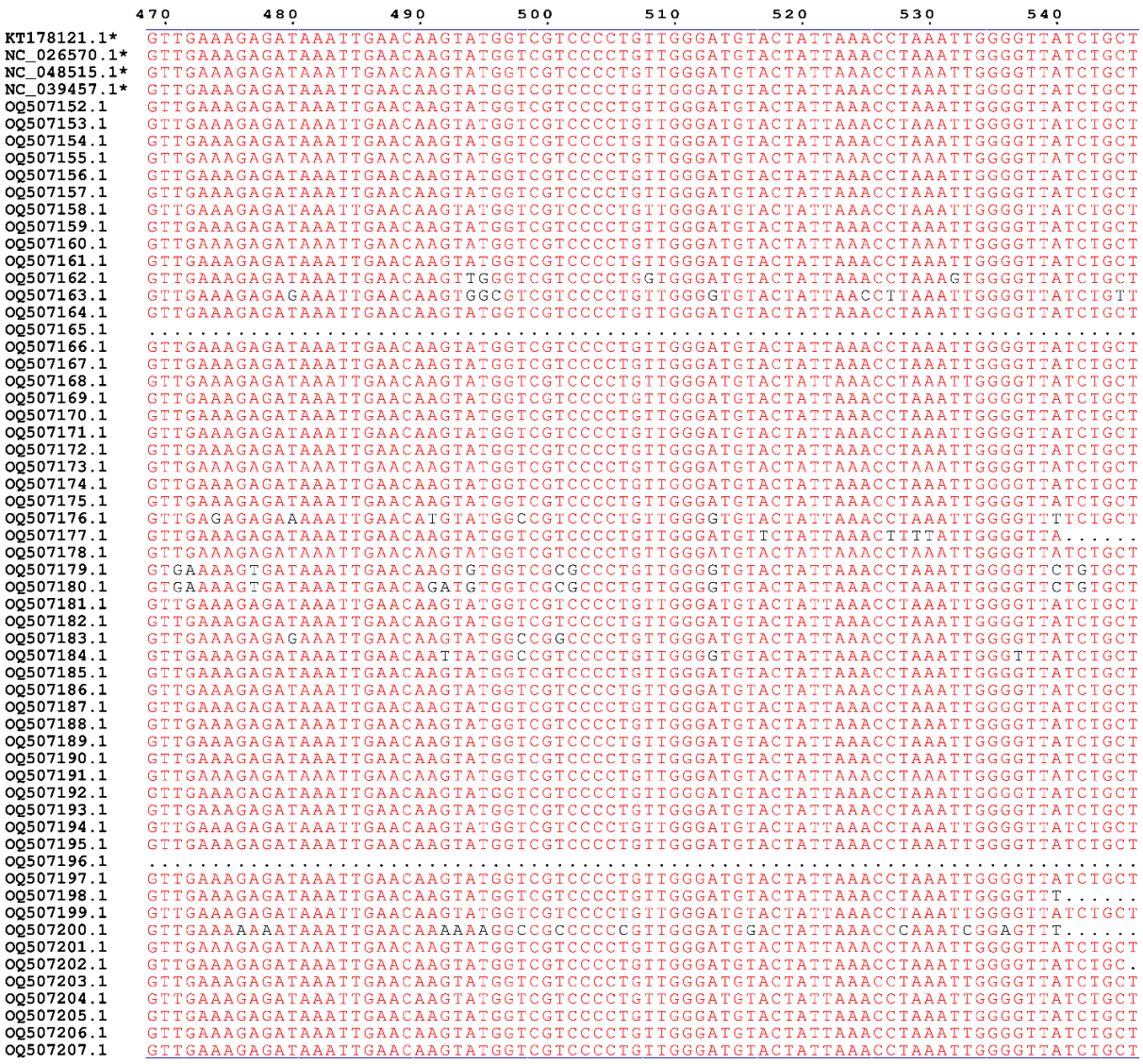

80 90 100 110 120 130 140 150
KT178121.1*
NC_026570.1*
NC_048515.1*
NC_039457.1*
OQ507152.1
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507162.1
OQ507163.1
OQ507164.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507172.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507176.1
OQ507177.1
OQ507178.1
OQ507179.1
OQ507180.1
OQ507181.1
OQ507182.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507186.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507196.1
OQ507197.1
OQ507198.1
OQ507199.1
OQ507200.1
OQ507201.1
OQ507202.1
OQ507203.1
OQ507204.1
OQ507205.1
OQ507206.1
OQ507207.1

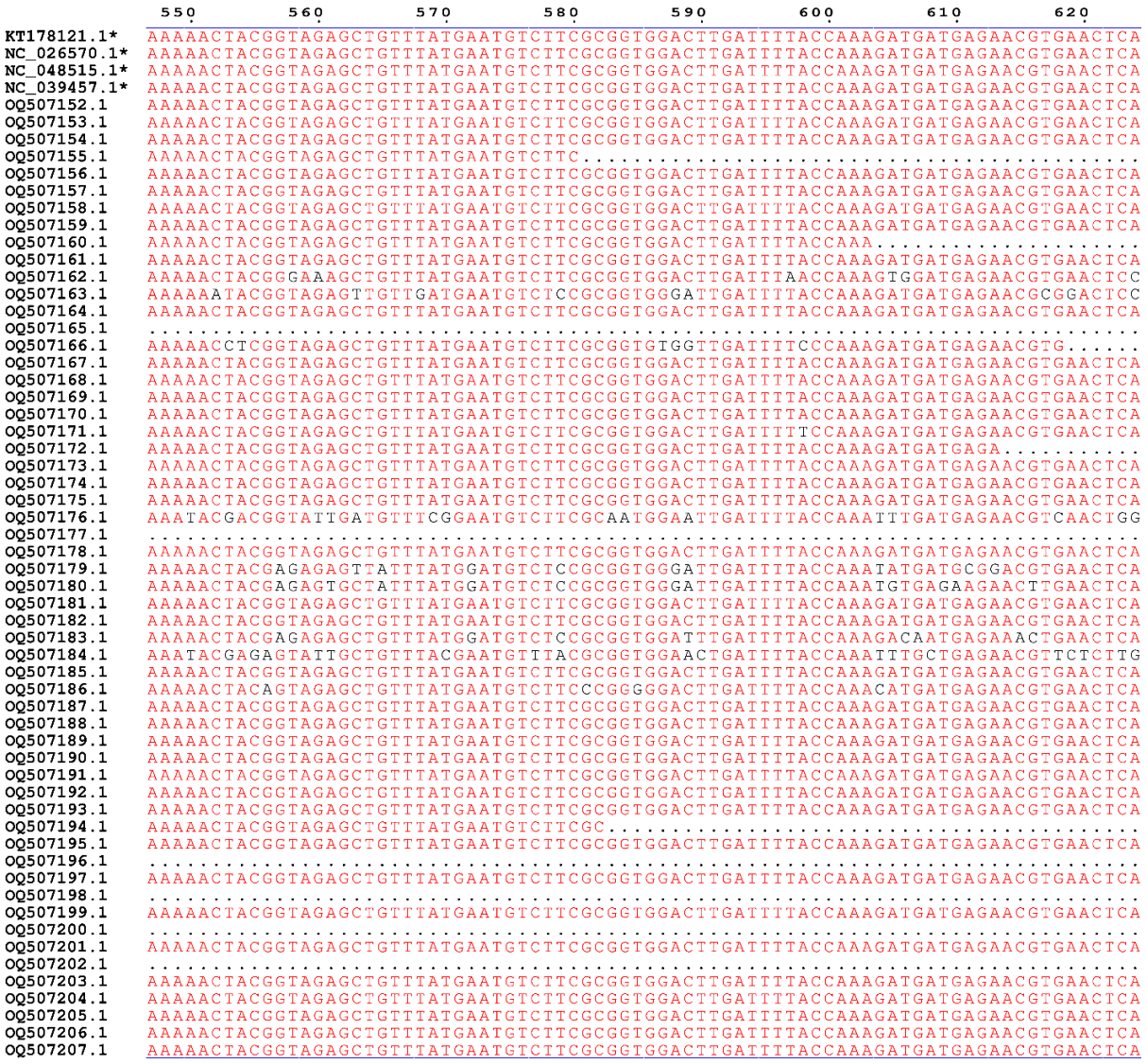

160 170 180 190 200 210 220 230
KT178121.1* GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
NC_026570.1* GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
NC_048515.1* GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
NC_039457.1* GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507152.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507153.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507154.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507155.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507156.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507157.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507158.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507159.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507160.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507161.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507162.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507163.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507164.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507165.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507166.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507167.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507168.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507169.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507170.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507171.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507172.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507173.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507174.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507175.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507176.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAC
OQ507177.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507178.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507179.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507180.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507181.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507182.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507183.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507184.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507185.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507186.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507187.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507188.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507189.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507190.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507191.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507192.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507193.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507194.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507195.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507196.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507197.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507198.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507199.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507200.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCCGTCTTGAT
OQ507201.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507202.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507203.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507204.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507205.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507206.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT
OQ507207.1 GCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTGTATGGACCGATGGACTTACCAGTCTTGAT

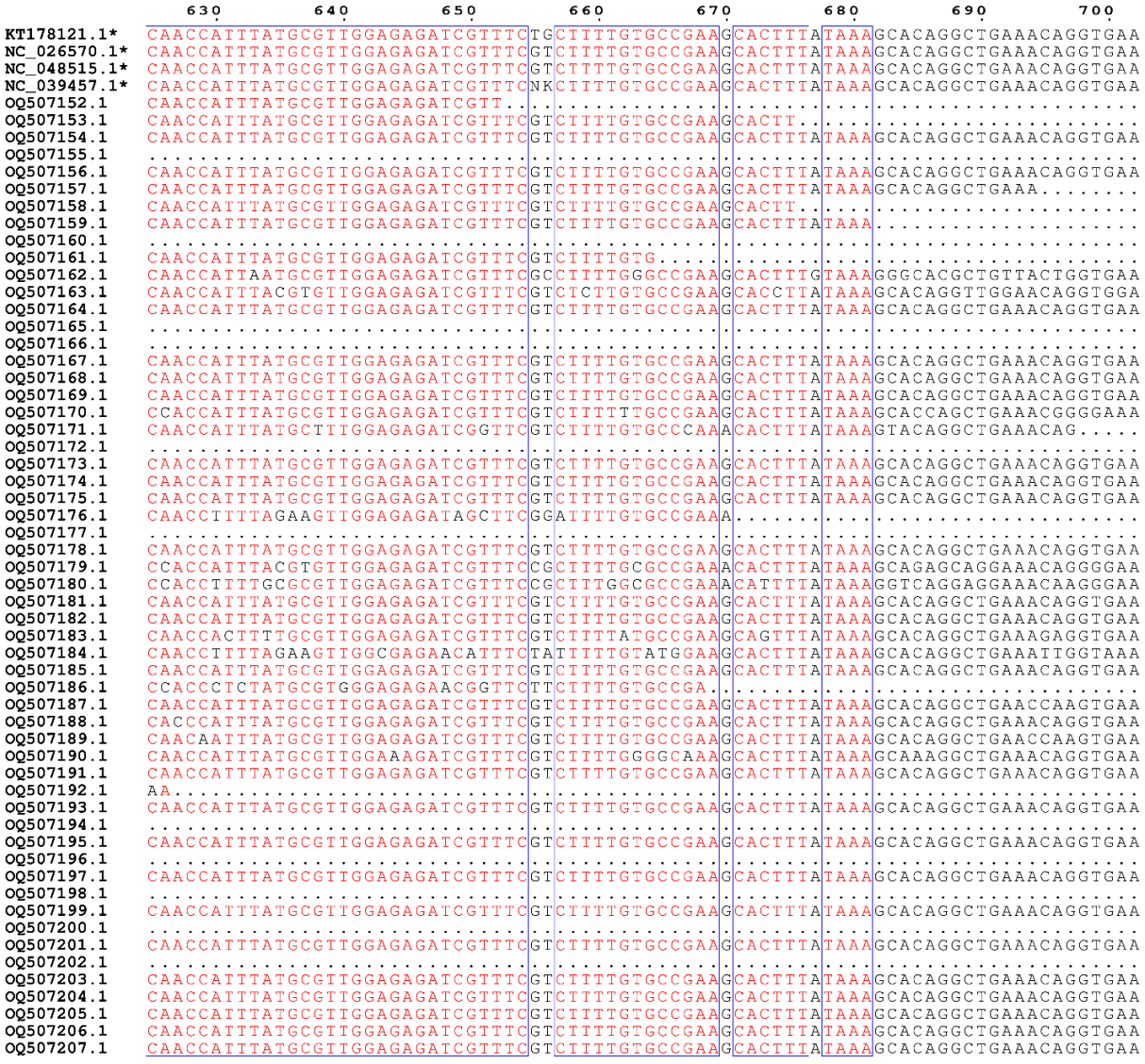

240
250
260
270
280
290
300
310
KT178121.1*
NC_026570.1*
NC_048515.1*
NC_039457.1*
OQ507152.1
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507162.1
OQ507163.1
OQ507164.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507172.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507176.1
OQ507177.1
OQ507178.1
OQ507179.1
OQ507180.1
OQ507181.1
OQ507182.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507186.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507196.1
OQ507197.1
OQ507198.1
OQ507199.1
OQ507200.1
OQ507201.1
OQ507202.1
OQ507203.1
OQ507204.1
OQ507205.1
OQ507206.1
OQ507207.1


320
330
340
350
360
370
380
390
KT178121.1*
NC_026570.1*
NC_048515.1*
NC_039457.1*
OQ507152.1
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507162.1
OQ507163.1
OQ507164.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507172.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507176.1
OQ507177.1
OQ507178.1
OQ507179.1
OQ507180.1
OQ507181.1
OQ507182.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507186.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507196.1
OQ507197.1
OQ507198.1
OQ507199.1
OQ507200.1
OQ507201.1
OQ507202.1
OQ507203.1
OQ507204.1
OQ507205.1
OQ507206.1
OQ507207.1


400
410
420
430
440
450
460
KT178121.1*
NC_026570.1*
NC_048515.1*
NC_039457.1*
OQ507152.1
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507162.1
OQ507163.1
OQ507164.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507172.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507176.1
OQ507177.1
OQ507178.1
OQ507179.1
OQ507180.1
OQ507181.1
OQ507182.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507186.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507196.1
OQ507197.1
OQ507198.1
OQ507199.1
OQ507200.1
OQ507201.1
OQ507202.1
OQ507203.1
OQ507204.1
OQ507205.1
OQ507206.1
OQ507207.1


470 480 490 500 510 520 530 540
KT178121.1* GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
NC_026570.1* GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
NC_048515.1* GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
NC_039457.1* GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507152.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507153.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507154.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507155.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507156.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507157.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507158.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507159.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507160.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507161.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507162.1 GTTGAAAGAGATAAATTGAACAAGTTGGGTCGTCCCCTGGTGGGATGTACTATTAAACCTAAAGTGGGGTTATCTGCT
OQ507163.1 GTTGAAAGAGAGAAATTGAACAAGTGGCGTCGTCCCCTGTTGGGGTGTACTATTAACCTTAAATTGGGGTTATCTGTT
OQ507164.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507165.1 ....
OQ507166.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507167.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507168.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507169.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507170.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507171.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507172.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507173.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507174.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507175.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507176.1 GTTGAGAGAGAAAAATTGAACATGTATGGCCGTCCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTTTCTGCT
OQ507177.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTTCTATTAAACTTTTATTGGGGTTA......
OQ507178.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507179.1 GTGAAAAGTGATAAATTGAACAAGTGTGGTCGCGCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTCTGTGCT
OQ507180.1 GTGAAAAGTGATAAATTGAACAGATGTGGTCGCGCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGGTTCTGTGCT
OQ507181.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507182.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507183.1 GTTGAAAGAGAGAAATTGAACAAGTATGGCCGGCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507184.1 GTTGAAAGAGATAAATTGAACAATTATGGCCGTCCCCTGTTGGGGTGTACTATTAAACCTAAATTGGGTTTATCTGCT
OQ507185.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507186.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507187.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507188.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507189.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507190.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507191.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507192.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507193.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507194.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507195.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507196.1 ....
OQ507197.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507198.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTT......
OQ507199.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507200.1 GTTGAAAAAAATAAATTGAACAAAAAAGGCCGCCCCCCGTTGGGATGGACTATTAAACCCAAATCGGAGTTT......
OQ507201.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507202.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGC.
OQ507203.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507204.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507205.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507206.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT
OQ507207.1 GTTGAAAGAGATAAATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCT

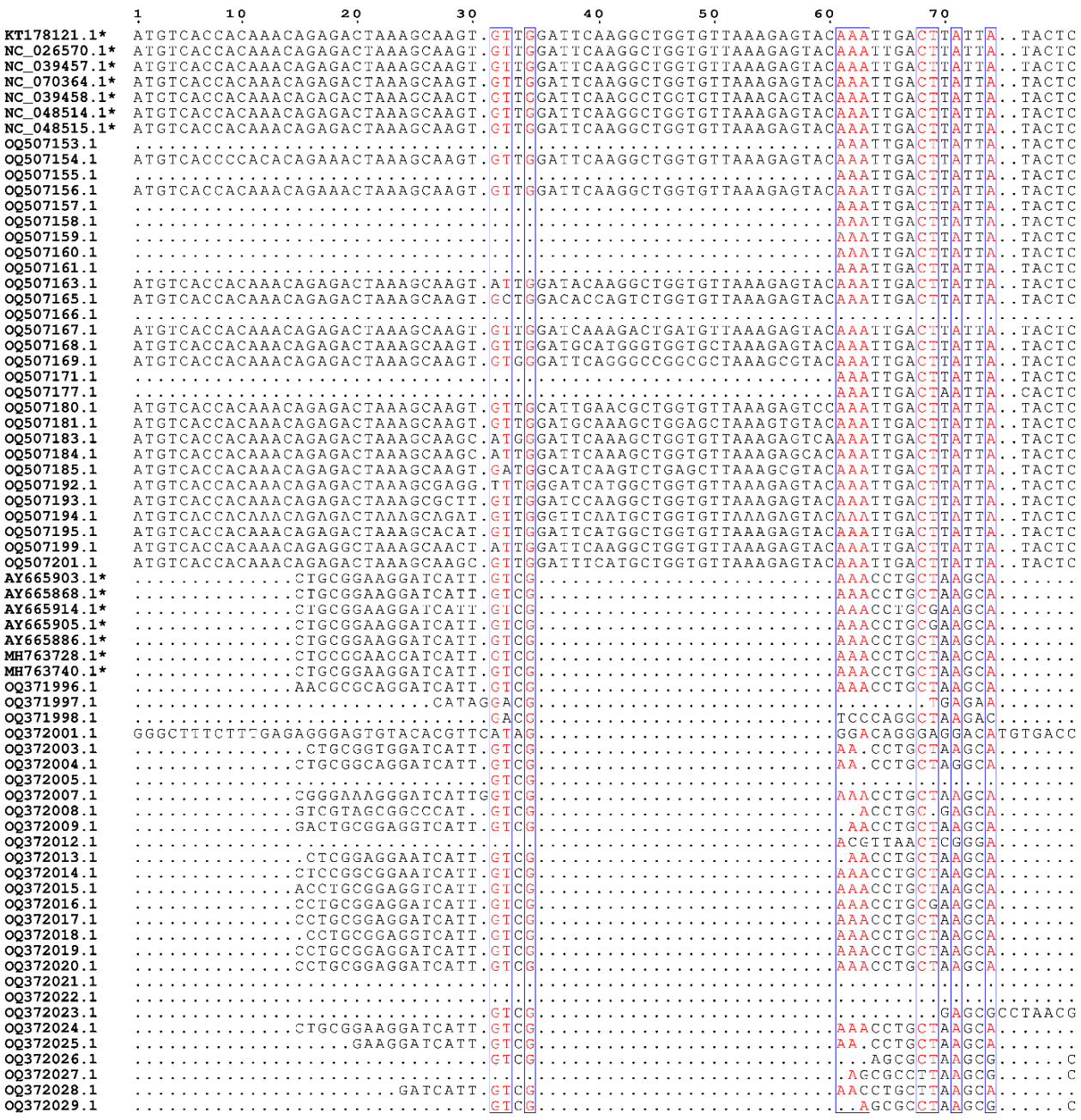

550 560 570 580 590 600 610 620
KT178121.1*
NC_026570.1*
NC_048515.1*
NC_039457.1*
OQ507152.1
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507162.1
OQ507163.1
OQ507164.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507170.1
OQ507171.1
OQ507172.1
OQ507173.1
OQ507174.1
OQ507175.1
OQ507176.1
OQ507177.1
OQ507178.1
OQ507179.1
OQ507180.1
OQ507181.1
OQ507182.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507186.1
OQ507187.1
OQ507188.1
OQ507189.1
OQ507190.1
OQ507191.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507196.1
OQ507197.1
OQ507198.1
OQ507199.1
OQ507200.1
OQ507201.1
OQ507202.1
OQ507203.1
OQ507204.1
OQ507205.1
OQ507206.1
OQ507207.1

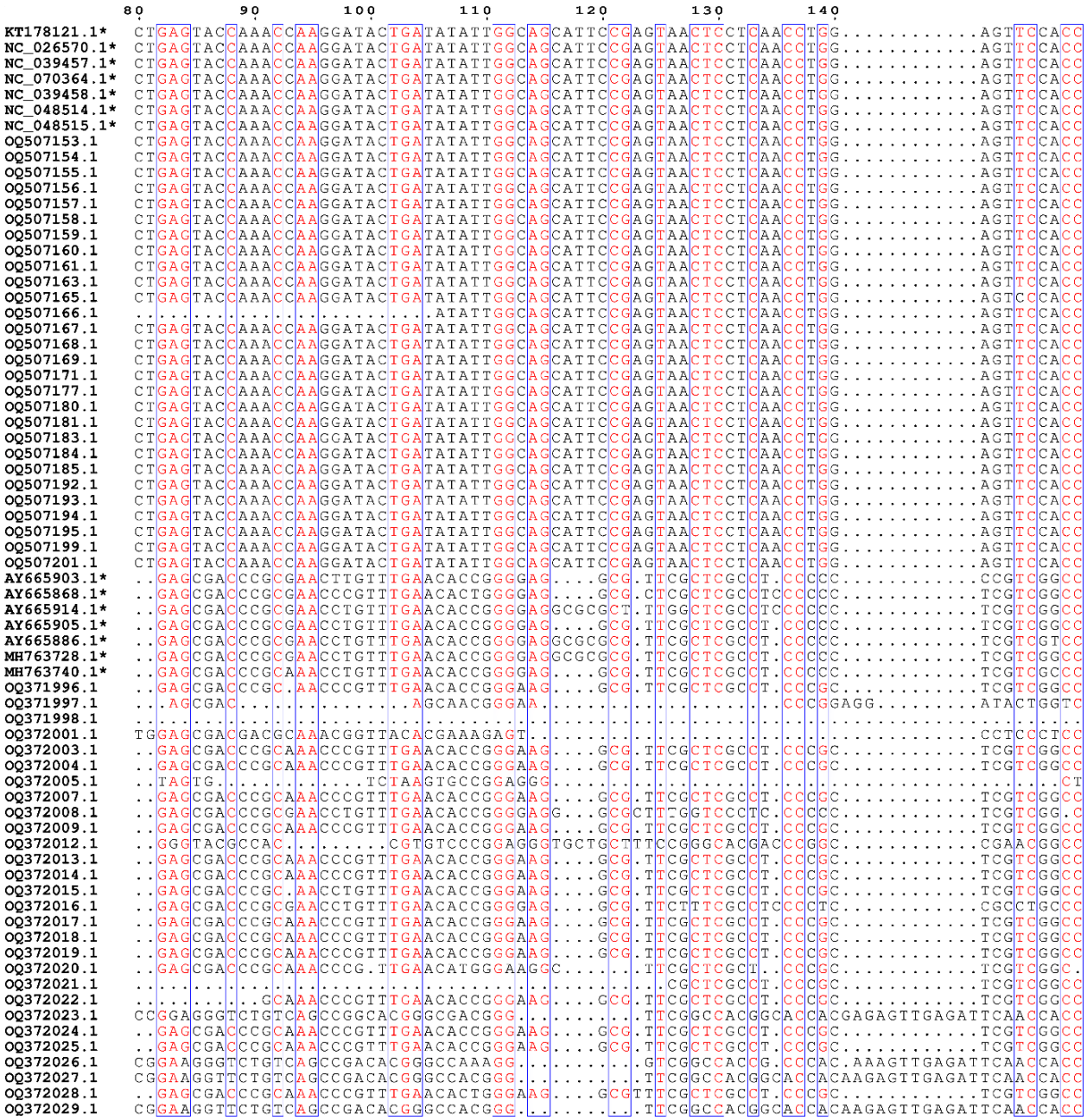

630 640 650 660 670 680 690 700
KT178121.1* CAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
NC_026570.1* CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
NC_048515.1* CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
NC_039457.1* CAACCATTTATGCGTTGGAGAGATCGTTTCNKCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507152.1 CAACCATTTATGCGTTGGAGAGATCGTT.....
OQ507153.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTT.....
OQ507154.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507155.1 .....
OQ507156.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507157.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAA.....
OQ507158.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTT.....
OQ507159.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAA.....
OQ507160.1 .....
OQ507161.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTG.....
OQ507162.1 CAACCATTAATGCGTTGGAGAGATCGTTTCGCCTTTTGGGCCGAAGCACTTTGTAAAGGGCACGCTGTTACTGGTGAA
OQ507163.1 CAACCATTTACGTGTTGGAGAGATCGTTTCGTCTCTTGTGCCGAAGCACCTTATAAAGCACAGGTTGGAACAGGTGGA
OQ507164.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507165.1 .....
OQ507166.1 .....
OQ507167.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507168.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507169.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507170.1 CCACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTTTGCCGAAGCACTTTATAAAGCACCAGCTGAAACGGGGAAA
OQ507171.1 CAACCATTTATGCTTTGGAGAGATCGGTTCGTCTTTTGTGCCCAAACACTTTATAAAGTACAGGCTGAAACAG.....
OQ507172.1 .....
OQ507173.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507174.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507175.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507176.1 CAACCTTTTAGAAGTTGGAGAGATAGCTTCGGATTTTGTGCCGAAA.....
OQ507177.1 .....
OQ507178.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507179.1 CCACCATTTACGTGTTGGAGAGATCGTTTCCGCTTTTGCGCCGAAACACTTTATAAAGCAGAGCAGGAAACAGGGGAA
OQ507180.1 CCACCTTTTGCGCGTTGGAGAGATCGTTTCCGCTTTGGCGCCGAAACATTTTATAAAGGTCAGGAGGAAACAAGGGAA
OQ507181.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507182.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507183.1 CAACCACTTTTGCGTTGGAGAGATCGTTTCGTCTTTTATGCCGAAGCAGTTTATAAAGCACAGGCTGAAAGAGGTGAA
OQ507184.1 CAACCTTTTAGAAGTTGGCGAGAACATTTCTATTTTTGTATGGAAGCACTTTATAAAGCACAGGCTGAAATTGGTAAA
OQ507185.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507186.1 CCACCCTCTATGCGTGGGAGAGAACGGTTCTTCTTTTGTGCCGA.....
OQ507187.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAACCAAGTGAA
OQ507188.1 CACCCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507189.1 CAACAATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAACCAAGTGAA
OQ507190.1 CAACCATTTATGCGTTGGAAAGATCGTTTCGTCTTTTGGGGCAAAGCACTTTATAAAGCAAAGGCTGAAACAGGTGAA
OQ507191.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507192.1 AA.....
OQ507193.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507194.1 .....
OQ507195.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507196.1 .....
OQ507197.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507198.1 .....
OQ507199.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507200.1 .....
OQ507201.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507202.1 .....
OQ507203.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507204.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507205.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507206.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA
OQ507207.1 CAACCATTTATGCGTTGGAGAGATCGTTTCGTCTTTTGTGCCGAAGCACTTTATAAAGCACAGGCTGAAACAGGTGAA

```
                710        720        730        740        750        760        770        780
KT178121.1*     ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTG
NC_026570.1*    ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTG
NC_048515.1*    ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTG
NC_039457.1*    ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTG
OQ507152.1      ................................................................................
OQ507153.1      ................................................................................
OQ507154.1      ATCAAAGGGCATTACTTGAATGCTACTGCAG.................................................
OQ507155.1      ................................................................................
OQ507156.1      ATCAAAGGGCATTACTTGAATGCTACTG....................................................
OQ507157.1      ................................................................................
OQ507158.1      ................................................................................
OQ507159.1      ................................................................................
OQ507160.1      ................................................................................
OQ507161.1      ................................................................................
OQ507162.1      ATCCAAGGGCATTACTTTTATGCTTTA.....................................................
OQ507163.1      ATCAAGGAGCATTACTTGGATGCTTCGGCC..................................................
OQ507164.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAA.....................................
OQ507165.1      ................................................................................
OQ507166.1      ................................................................................
OQ507167.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA.......................................
OQ507168.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATATTCAAAAGAGGTGTCTTTGCCAGGGAATTG
OQ507169.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAG.....................................
OQ507170.1      ATCAAAGGGCATTAC.................................................................
OQ507171.1      ................................................................................
OQ507172.1      ................................................................................
OQ507173.1      ATCAAAGGGCATTACTTGAATGCTTCTGCATGTACATGCGAAAAACTGATTAAAAGAGCTGTATCTGCGAGACAATTG
OQ507174.1      ATCAAAGCATACTGCTTGAATGCTACTGCAGTTACATGCGAA.......................................
OQ507175.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACTTGCCAAGACGTGATCAAAATAGCAG...................
OQ507176.1      ................................................................................
OQ507177.1      ................................................................................
OQ507178.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGTTACATGCGAAGAAACTATTAAAAGGGCTGTCTTTGATAGAGAACGG
OQ507179.1      ATCAAAGGGTATTACTGCAATG..........................................................
OQ507180.1      ATTGAAGTAT......................................................................
OQ507181.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCAAAA.....................................
OQ507182.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATTGCAA.......................................
OQ507183.1      ATCAAAGGGTCTTACTTGAATGCTACTGCAGGTACCCGCGAAGAAATGATCAAAAGAGCTGTATTTGTTAGAGAATTG
OQ507184.1      ATCAAAGGGGGTTACTTGAATTTTACTGCAGGTACCGGCGAAGAAATGAGTAAAAGAGCTGTATTTAGTAGAGG....
OQ507185.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGGCAAA.....................................
OQ507186.1      ................................................................................
OQ507187.1      ATCAAAGGGCATTACTTGGATGCTATCCGTG.................................................
OQ507188.1      ATTAAAGGGCATTACCTGAATGCTACTGCAGTTCTCTGCGAAGAAATGATCAAAAACGCTACGTTTGATAGAGAATTG
OQ507189.1      ATCACAGGGCATTACTGGAATGCCATTGCAGGTACATTCCAAGAATTGGCCAAGAGAAATGTTCTTGCGAGAGAATTG
OQ507190.1      ATCAACGCGCATTACTTGAATGCTACTGCAGTTACCTGCGAAGAAATGGCCAAATGGGCTGTATTTGTAAGAGAACTG
OQ507191.1      ATCAAAGGGCATTACTTGAATGCTACTGAGT..................................................
OQ507192.1      ................................................................................
OQ507193.1      ATCAAAGGGCATTACCTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAGAGTTCTATATTTGCTAGAGAATTG
OQ507194.1      ................................................................................
OQ507195.1      ATCAAAGGGCATTACTTGGATGCTACTGCAGGTGCACTTGAAGAACTATTCAAAAGAGCTTCTTTTGCCACGGAATTG
OQ507196.1      ................................................................................
OQ507197.1      TTCAACGCATACTGCTTGAATGCTACTGCAGTCACATGCGAAGAAGCGATCATGTTTTCCCTACACACAAGAAAACCG
OQ507198.1      ................................................................................
OQ507199.1      ATTCAACGGCATTACTTGAATGCTACTGCAGTTACATGGGAAGAAATTCTCAAAAGAGCTCTATTCACTAGAGAATTG
OQ507200.1      ................................................................................
OQ507201.1      ATCAAAGGGCATTACTTGGATGCTACTGCAGGTCCAGGGGGAA.....................................
OQ507202.1      ................................................................................
OQ507203.1      ATCAAAGGGTACTACTTGAATGCTATTGCAGGCACTCGCGAAGAGATCATCTGGAGAGAGGTATTTGCTGC.......
OQ507204.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGTTACATGCGAAAACATGGTCAAAAGTGTTGAATTTGCTACAT.....
OQ507205.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAGATGCTTAAAAACGCTGTATTTGCTTTAGCATTG
OQ507206.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGTTCCATGCGAAGAAATGATCAAAAGTTCTTTGTTTGCGAGAGACTTG
OQ507207.1      ATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATTTTCAAAAGAGCCCTATTTGAGAGAGAAACG
```

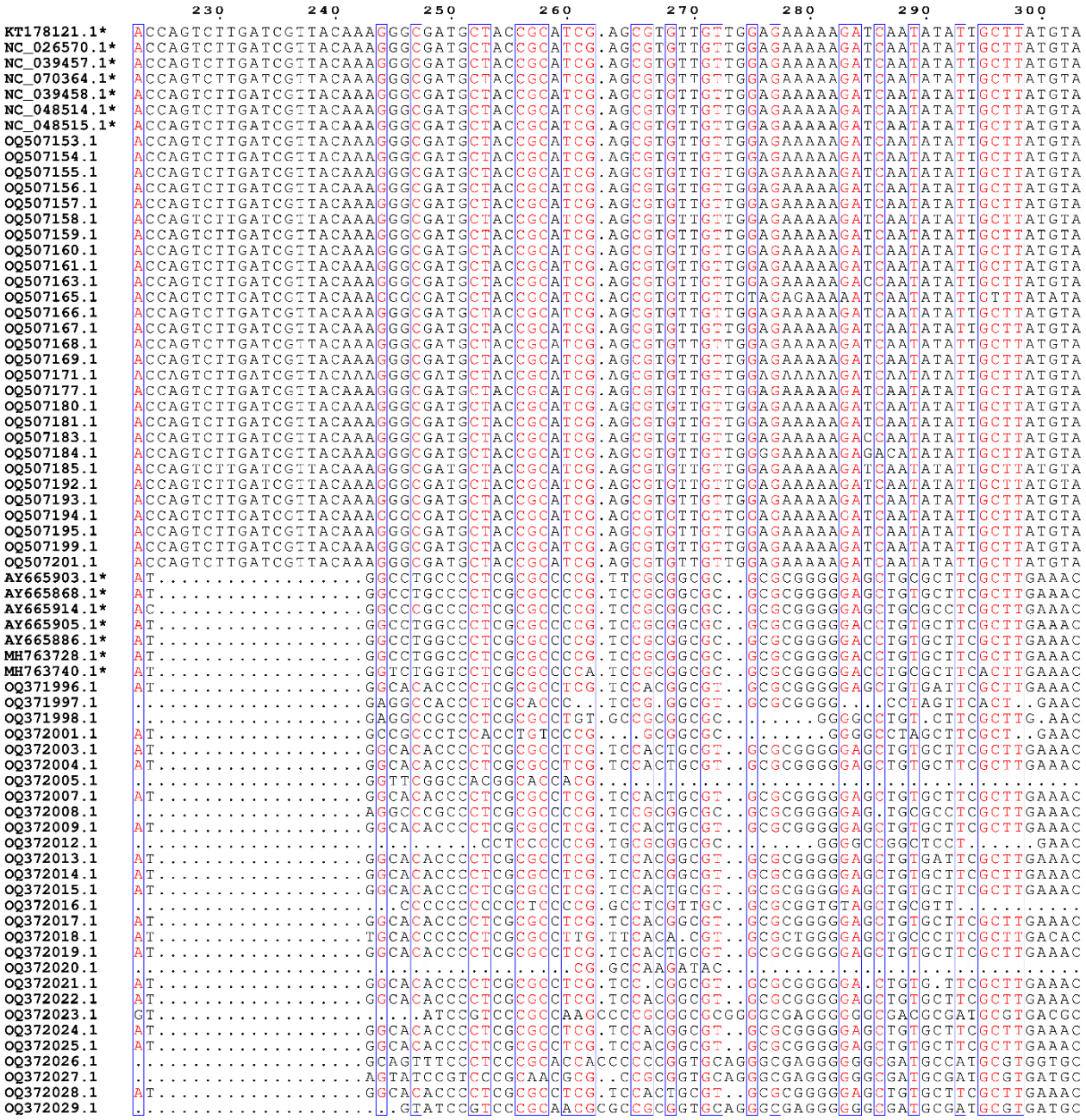
```
                    790       800       810       820       830       840       850
KT178121.1*   GGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGGCTCATTATTGCCG
NC_026570.1*  GGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGGCTCATTATTGCCG
NC_048515.1*  GGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGGCTCATTATTGCCG
NC_039457.1*  GGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGGCTCATTATTGCCG
OQ507152.1    ................................................................................
OQ507153.1    ................................................................................
OQ507154.1    ................................................................................
OQ507155.1    ................................................................................
OQ507156.1    ................................................................................
OQ507157.1    ................................................................................
OQ507158.1    ................................................................................
OQ507159.1    ................................................................................
OQ507160.1    ................................................................................
OQ507161.1    ................................................................................
OQ507162.1    ................................................................................
OQ507163.1    ................................................................................
OQ507164.1    ................................................................................
OQ507165.1    ................................................................................
OQ507166.1    ................................................................................
OQ507167.1    ................................................................................
OQ507168.1   GGCGTTCCGGTCGTAATATATGACCACTTAACGTT.....................................
OQ507169.1    ................................................................................
OQ507170.1    ................................................................................
OQ507171.1    ................................................................................
OQ507172.1    ................................................................................
OQ507173.1   G...............................................................................
OQ507174.1    ................................................................................
OQ507175.1    ................................................................................
OQ507176.1    ................................................................................
OQ507177.1    ................................................................................
OQ507178.1   GGTATTCCAATTATAATTTATGACTATTCCAAGGGGGGATTCACCACAAATACTAGCTTATCTCATCATTACCG
OQ507179.1    ................................................................................
OQ507180.1    ................................................................................
OQ507181.1    ................................................................................
OQ507182.1    ................................................................................
OQ507183.1   GGCGTTCCGATCGTAATGCATTACAACTTAACGGGGAA.....................................
OQ507184.1    ................................................................................
OQ507185.1    ................................................................................
OQ507186.1    ................................................................................
OQ507187.1    ................................................................................
OQ507188.1   GGCCTTCCGGTA....................................................................
OQ507189.1   GGCGTTCCCAGCGTAATGCATAACTGCTTACCGAG.....................................
OQ507190.1   GGCGTTGCG.......................................................................
OQ507191.1    ................................................................................
OQ507192.1    ................................................................................
OQ507193.1   CGCGCTCCTGTCGTGATGTATGAGTACTTAACTTC.....................................
OQ507194.1    ................................................................................
OQ507195.1   GGGGTTCCGGTCGAAACGGATGAGCACTTGATGGGGGGTTTCACTACAAACACTAGCTTTG.............
OQ507196.1    ................................................................................
OQ507197.1   ACCTCTTTTTGTGTAGTACGTGATGAA.....................................................
OQ507198.1    ................................................................................
OQ507199.1   CGCGTTCCGATCGTAATGGATG..........................................................
OQ507200.1    ................................................................................
OQ507201.1    ................................................................................
OQ507202.1    ................................................................................
OQ507203.1    ................................................................................
OQ507204.1    ................................................................................
OQ507205.1   ATCGTTCCAATTATAATGCATA..........................................................
OQ507206.1   GGCGCTCTGGTTGTGATTGATGC.........................................................
OQ507207.1   GGCGTTGTGGGCGTAATGCATGTCTTGTTAACGGGGGGATTCATTGCTAATACTAGGTTGGCTCATG.......
```

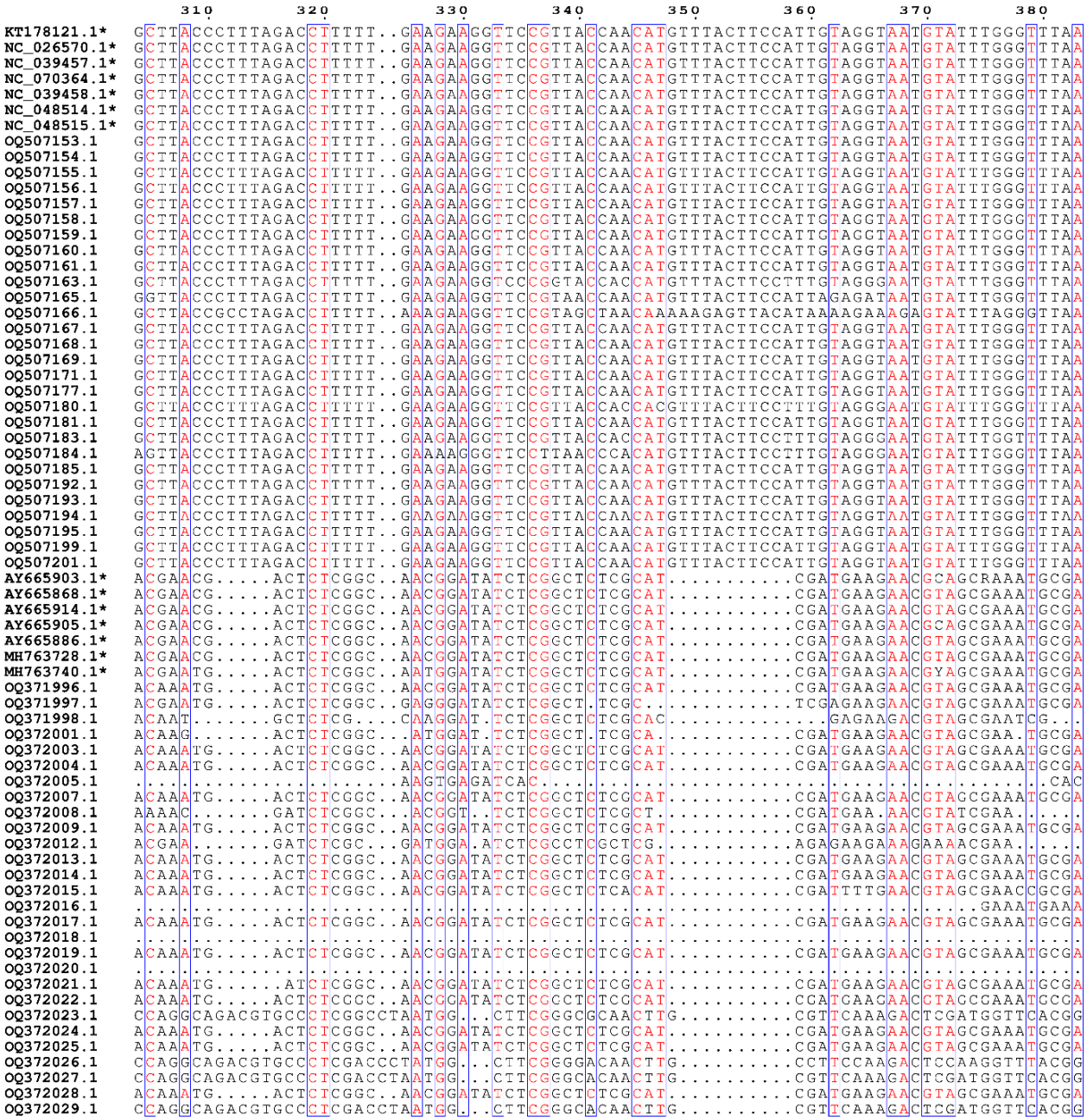
Appendix 2 D:Multiple alignment sequence for ITS2 and *rbcL Physalis* accessions gene sequence as well as reference sequences. (https://espript.ibcp.fr/ESPript/temp/1032964064/0-0-1680466160-esp.pdf)



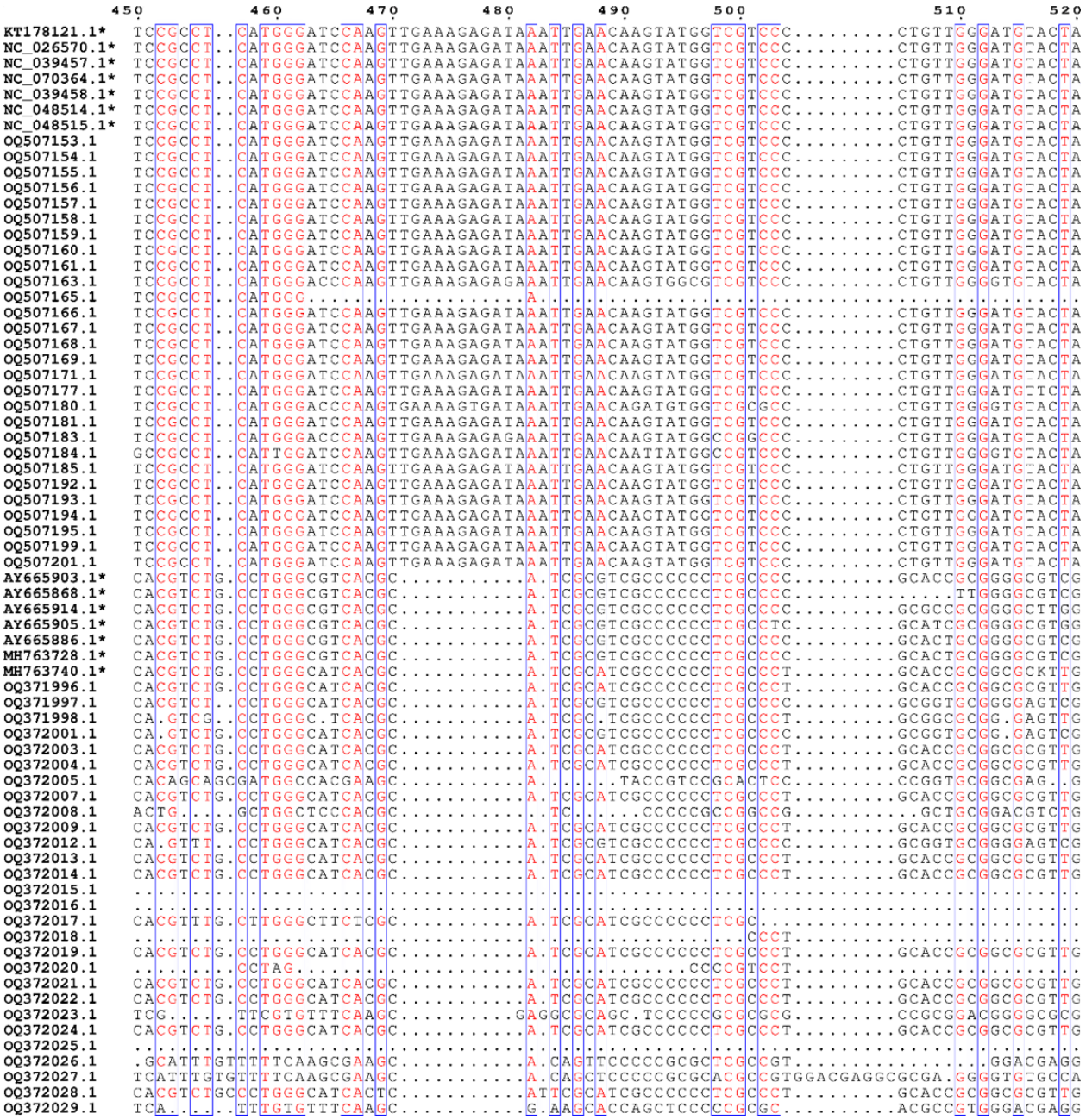

80 90 100 110 120 130 140
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

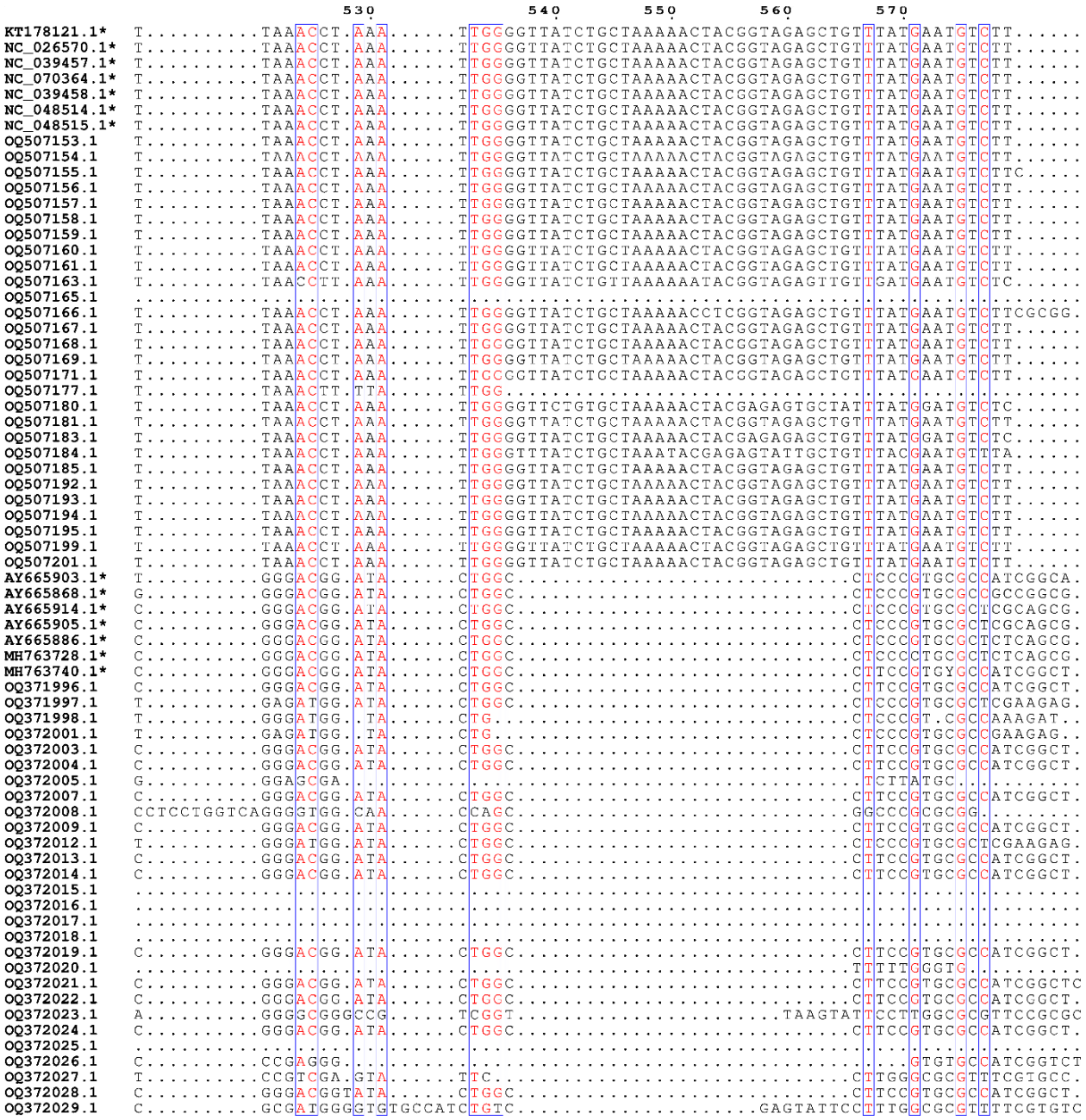

150
160
170
180
190
200
210
220
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

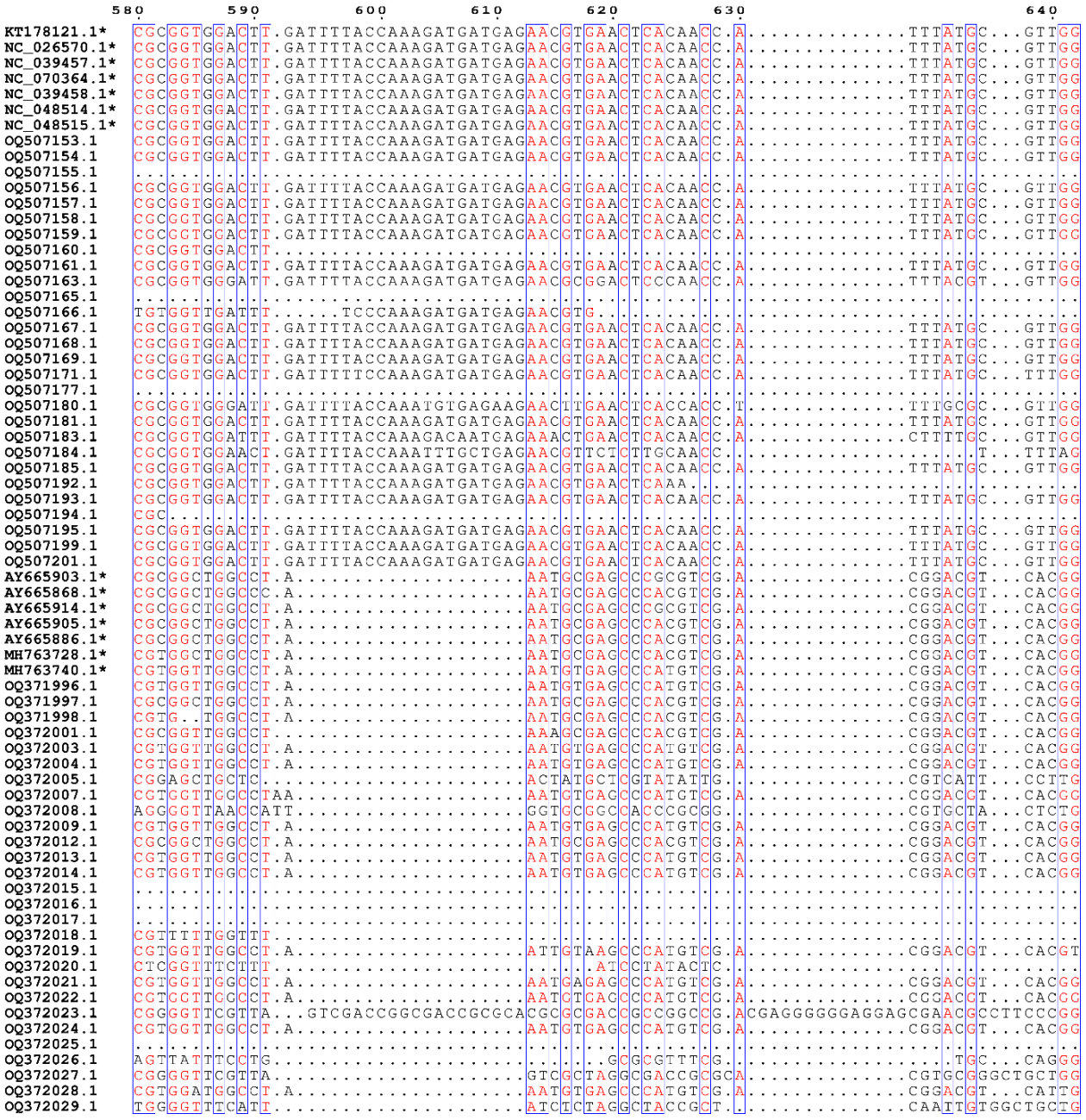

230
240
250
260
270
280
290
300
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

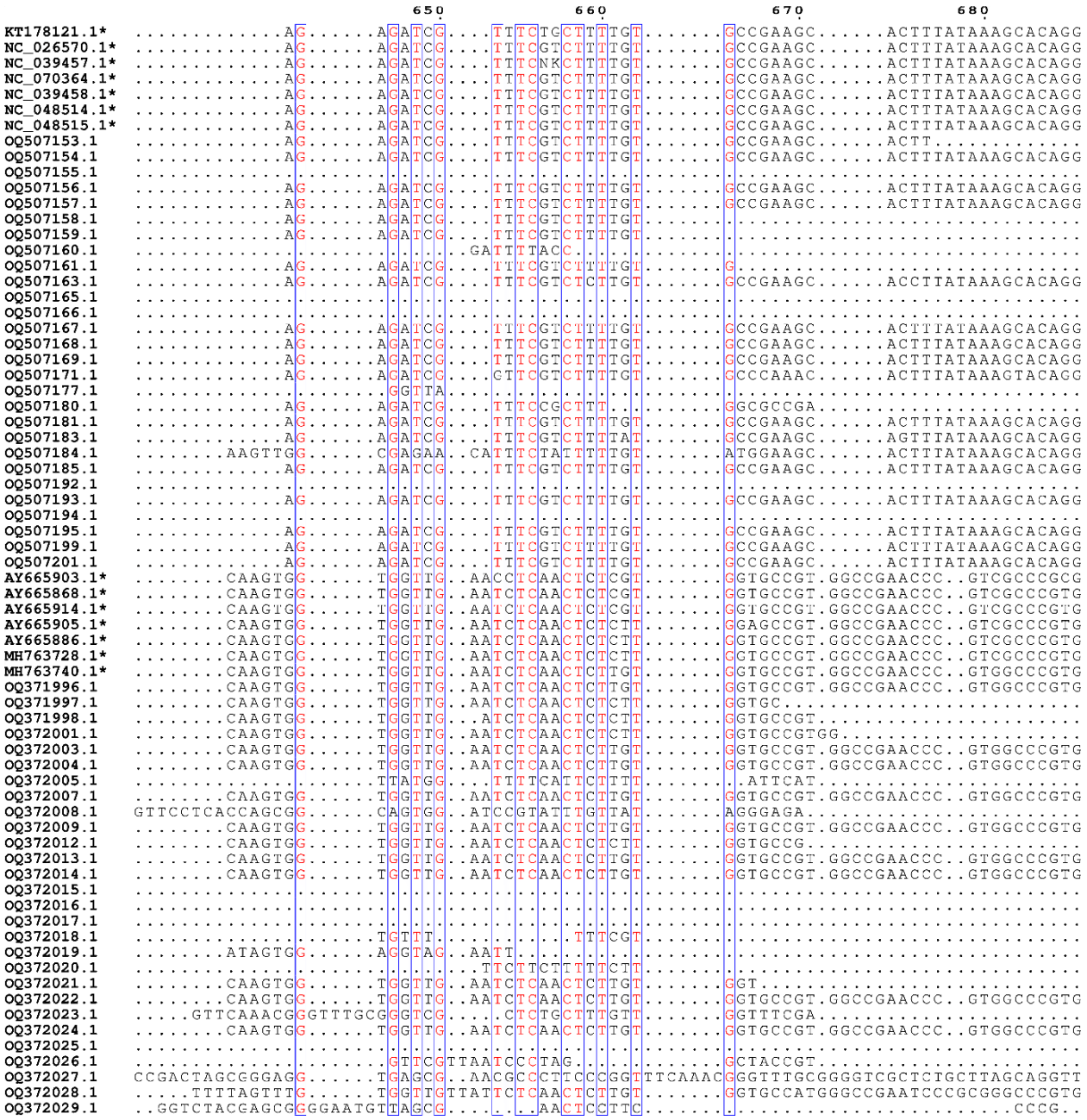

310
320
330
340
350
360
370
380
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1


390
400
410
420
430
440
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1


450
460
470
480
490
500
510
520
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1


530
540
550
560
570
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1


580
590
600
610
620
630
640
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1


650
660
670
680
KT178121.1*
NC_026570.1*
NC_039457.1*
NC_070364.1*
NC_039458.1*
NC_048514.1*
NC_048515.1*
OQ507153.1
OQ507154.1
OQ507155.1
OQ507156.1
OQ507157.1
OQ507158.1
OQ507159.1
OQ507160.1
OQ507161.1
OQ507163.1
OQ507165.1
OQ507166.1
OQ507167.1
OQ507168.1
OQ507169.1
OQ507171.1
OQ507177.1
OQ507180.1
OQ507181.1
OQ507183.1
OQ507184.1
OQ507185.1
OQ507192.1
OQ507193.1
OQ507194.1
OQ507195.1
OQ507199.1
OQ507201.1
AY665903.1*
AY665868.1*
AY665914.1*
AY665905.1*
AY665886.1*
MH763728.1*
MH763740.1*
OQ371996.1
OQ371997.1
OQ371998.1
OQ372001.1
OQ372003.1
OQ372004.1
OQ372005.1
OQ372007.1
OQ372008.1
OQ372009.1
OQ372012.1
OQ372013.1
OQ372014.1
OQ372015.1
OQ372016.1
OQ372017.1
OQ372018.1
OQ372019.1
OQ372020.1
OQ372021.1
OQ372022.1
OQ372023.1
OQ372024.1
OQ372025.1
OQ372026.1
OQ372027.1
OQ372028.1
OQ372029.1

```
                690                 700       710       720       730       740       750       760
KT178121.1*     CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_026570.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_039457.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_070364.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_039458.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_048514.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
NC_048515.1*    CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAGAG
OQ507153.1      ..........................................................................................
OQ507154.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAG.....................................
OQ507155.1      ..........................................................................................
OQ507156.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTG.......................................
OQ507157.1      CT.........GAAA..........................................................................
OQ507158.1      ........................GCCGAAGCACTT......................................................
OQ507159.1      ........................GCCGAAGCACTTTATAAA................................................
OQ507160.1      ...........AAA...........................................................................
OQ507161.1      ..........................................................................................
OQ507163.1      TT.........GGAACAGGTGGAATCAAGGAGCATTACTTGGATGCTTCGGCC.....................................
OQ507165.1      ..........................................................................................
OQ507166.1      ..........................................................................................
OQ507167.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA..........................
OQ507168.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATATTCAAAAGAG
OQ507169.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAG.....................
OQ507171.1      CT.........GAAACAG.......................................................................
OQ507177.1      ..........................................................................................
OQ507180.1      ............................AACATTTTATAAAGGTCAGGAGGAAACAAGGGAAATTGAAGTAT.....
OQ507181.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCAAAA.....................
OQ507183.1      CT.........GAAAGAGGTGAAATCAAAGGGTCTTACTTGAATGCTACTGCAGGTACCCGCGAAGAAATGATCAAAAGAG
OQ507184.1      CT.........GAAATTGGTAAAATCAAAGGGGGTTACTTGAATTTTACTGCAGGTACCGGCGAAGAAATGAGTAAAAGAG
OQ507185.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGGCAAA.....................
OQ507192.1      ..........................................................................................
OQ507193.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACCTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAGAGTT
OQ507194.1      ..........................................................................................
OQ507195.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGGATGCTACTGCAGGTGCACTTGAAGAACTATTCAAAAGAG
OQ507199.1      CT.........GAAACAGGTGAAATTCAACGGCATTACTTGAATGCTACTGCAGTTACATGGGAAGAAATTCTCAAAAGAG
OQ507201.1      CT.........GAAACAGGTGAAATCAAAGGGCATTACTTGGATGCTACTGCAGGTCCAGGGGGAA.....................
AY665903.1*     TC.GGCTGATG.GACCCTTCAT..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
AY665868.1*     TC.GGCTTACA.GACCCTTCCG..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
AY665914.1*     CC.GGCTGACA.GACCCTTCCG..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
AY665905.1*     TC.GGCTGTGA.GACCCTTCCG..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
AY665886.1*     TC.GGCTGCGA.GACCCTTCCG..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
MH763728.1*     TC.GGCTGCGA.GACCCTTCCG..GCGCTTAGGCGCTC..CGACCGCGACCCCAGGTCAGGCGGGATT...................
MH763740.1*     TC.GGCTGACA.GAACCTTCCG..GCGCTTAGGCGCTT..CGACCGCGACCCCAGGTCAGGTGGAATT...................
OQ371996.1      TC.GGCTGACA.GAACCTTCCCG.GCGCTTAGGCGCTT..CGACCGCGACCCCAGGTCAGGTGGAATT...................
OQ371997.1      ..........................................................................................
OQ371998.1      ..........................................................................................
OQ372001.1      ..........................................................................................
OQ372003.1      TC.GGCTGACA.GAACCTTCCGGCGC...TTAGCGCTT..CGA.CGCGACCC.AGGTCAGGTGA.....................
OQ372004.1      TC.GGCTGACA.GAACCTTTCCG.GCGCTTAGGCGCTT..CGACCGCGACCCCAGGTCAGGTGGAAA...................
OQ372005.1      ..........................................................................................
OQ372007.1      TCGGGCTGACA.GAACCTTTCCG.................................................................
OQ372008.1      ..........................................................................................
OQ372009.1      TC.GGCTGACA.GAACCTTCCG..GCGCTTAGGCGCTT..CGACCGCGACCCCAGGTCAGTGGAA.....................
OQ372012.1      ..........................................................................................
OQ372013.1      TC.AGCTGACA.GAACCTTCCG..GCGCTTAGGCGCTT..CGACCGCGACCCCAGGTCAGTGGAA.....................
OQ372014.1      TC.AGCTGACA.GAACCTTCCGGCGC...TTAGCGCTT..CGACCGCGACCCCAGGTCAGTGGAA.....................
OQ372015.1      ..........................................................................................
OQ372016.1      ..........................................................................................
OQ372017.1      ..........................................................................................
OQ372018.1      ..........................................................................................
OQ372019.1      ..........................................................................................
OQ372020.1      ..........................................................................................
OQ372021.1      ..........................................................................................
OQ372022.1      TC.GGCTGACA.GAACCTTTCGGTGCGCTTAGGCGCTT..CGACCGCGACCCCAGGTGCAGGTGGAA.....................
OQ372023.1      ........CAATGATCCTTCCGCAG.........GTTCACCTATCGAAACC.....................................
OQ372024.1      TC.GGCTGACA.GAACCTCCG...GCGCTTAGGCGCT...CGACCGCGACCCAGTCTGGGATTTTCG.....................
OQ372025.1      ..........................................................................................
OQ372026.1      ............GACACCT.....................................................................
OQ372027.1      TC....GACAATGAACCTTCCG.....CATGGTCACCTACGGGAAAC.........................................
OQ372028.1      TCGGGCTGACATAAACCTTCCGGCGTTTTTAGGCCCCTTCCACCCGCGTACCCC..................................
OQ372029.1      ..........................................................................................
```

```
                    770        780        790        800        810        820        830        840
KT178121.1*   CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_026570.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_039457.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_070364.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_039458.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_048514.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
NC_048515.1*  CTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGACTACTTAACGGGGGGATTCACCGCAAATACTAGCTTGG
OQ507153.1    ................................................................................
OQ507154.1    ................................................................................
OQ507155.1    ................................................................................
OQ507156.1    ................................................................................
OQ507157.1    ................................................................................
OQ507158.1    ................................................................................
OQ507159.1    ................................................................................
OQ507160.1    ................................................................................
OQ507161.1    ................................................................................
OQ507163.1    ................................................................................
OQ507165.1    ................................................................................
OQ507166.1    ................................................................................
OQ507167.1    ................................................................................
OQ507168.1   GTGTCTTTGCCAGGGAATTGGGCGTTCCGGTCGTAATATATGACCACTTAACGTT.........................
OQ507169.1    ................................................................................
OQ507171.1    ................................................................................
OQ507177.1    ................................................................................
OQ507180.1    ................................................................................
OQ507181.1    ................................................................................
OQ507183.1   CTGTATTTGTTAGAGAATTGGGCGTTCCGATCGTAATGCATTACAACTTAACGGGGAA.....................
OQ507184.1   CTGTATTTAGTAGAGG................................................................
OQ507185.1    ................................................................................
OQ507192.1    ................................................................................
OQ507193.1   CTATATTTGCTAGAGAATTGCGCGCTCCTGTCGTGATGTATGAGTACTTAACTTC.........................
OQ507194.1    ................................................................................
OQ507195.1   CTTCTTTTGCCACGGAATTGGGGGTTCCGGTCGAAACGGATGAGCACTTGATGGGGGGTTTCACTACAAACACTAGCTTTG
OQ507199.1   CTCTATTCACTAGAGAATTGCGCGTTCCGATCGTAATGGATG.....................................
OQ507201.1    ................................................................................
AY665903.1*   ................................................................................
AY665868.1*   ................................................................................
AY665914.1*   ................................................................................
AY665905.1*   ................................................................................
AY665886.1*   ................................................................................
MH763728.1*   ................................................................................
MH763740.1*   ................................................................................
OQ371996.1    ................................................................................
OQ371997.1    ................................................................................
OQ371998.1    ................................................................................
OQ372001.1    ................................................................................
OQ372003.1    ................................................................................
OQ372004.1    ................................................................................
OQ372005.1    ................................................................................
OQ372007.1    ................................................................................
OQ372008.1    ................................................................................
OQ372009.1    ................................................................................
OQ372012.1    ................................................................................
OQ372013.1    ................................................................................
OQ372014.1    ................................................................................
OQ372015.1    ................................................................................
OQ372016.1    ................................................................................
OQ372017.1    ................................................................................
OQ372018.1    ................................................................................
OQ372019.1    ................................................................................
OQ372020.1    ................................................................................
OQ372021.1    ................................................................................
OQ372022.1    ................................................................................
OQ372023.1    ................................................................................
OQ372024.1    ................................................................................
OQ372025.1    ................................................................................
OQ372026.1    ................................................................................
OQ372027.1    ................................................................................
OQ372028.1    ................................................................................
OQ372029.1    ................................................................................
```

Appendix 3: Tables for ANOVA and Tukey's HSD post hoc analysis of mineral content, phytochemical content and radical scavenging activity of indigenous *Physalis* accessions

Appendix 3 A: Table for one way ANOVA and post hoc analysis of macro-minerals of indigenous *Physalis* accessions.

Accession name (I)	Accession name (J)	ANOVA p-value of Ca content	ANOVA p-value of Na content	ANOVA p-value of K content	ANOVA p-value of Mg content	Tukey's HSD p-value of Ca content	Tukey's HSD p-value of Na content	Tukey's HSD p-value of K content	Tukey's HSD p-value of Mg content
OQ507152.1	OQ372021.1	0.050	0.509	0.380	0.435	0.027	0.973	1.000	0.963
	OQ372022.1					0.976	0.744	1.000	0.865
	OQ372023.1					0.136	0.850	1.000	0.790
	OQ372024.1					0.176	0.960	0.724	1.000
	OQ372025.1					1.000	1.000	0.937	0.964
	OQ372026.1					0.999	1.000	1.000	0.928
	OQ372028.1					1.000	1.000	0.698	1.000
	OQ372029.1					1.000	1.000	1.000	0.943
OQ372021.1	OQ507152.1					0.027	0.973	1.000	0.963
	OQ372022.1					0.193	0.999	0.987	1.000
	OQ372023.1					0.994	1.000	0.994	1.000
	OQ372024.1					0.983	1.000	0.488	0.867
	OQ372025.1					0.044	0.977	0.775	1.000
	OQ372026.1					0.091	0.961	0.997	1.000
	OQ372028.1					0.024	0.994	0.462	0.960
	OQ372029.1					0.074	0.999	0.999	1.000
OQ372022.1	OQ507152.1					0.976	0.744	1.000	0.865
	OQ372021.1					0.193	0.999	0.987	1.000
	OQ372023.1					0.604	1.000	1.000	1.000
	OQ372024.1					0.692	1.000	0.950	0.702
	OQ372025.1					0.995	0.759	0.998	1.000

	OQ372026.1					1.000	0.704	1.000	1.000
	OQ372028.1					0.965	0.854	0.939	0.859
	OQ372029.1					1.000	0.922	1.000	1.000
OQ372023.1	OQ507152.1					0.136	0.850	1.000	0.790
	OQ372021.1					0.994	1.000	0.994	1.000
	OQ372022.1					0.604	1.000	1.000	1.000
	OQ372024.1					1.000	1.000	0.924	0.607
	OQ372025.1					0.205	0.862	0.996	1.000
	OQ372026.1					0.363	0.818	1.000	1.000
	OQ372028.1					0.119	0.930	0.909	0.783
	OQ372029.1					0.310	0.970	1.000	1.000
OQ372024.1	OQ507152.1					0.716	0.960	0.724	1.000
	OQ372021.1					0.983	1.000	0.488	0.867
	OQ372022.1					0.692	1.000	0.950	0.702
	OQ372023.1					1.000	1.000	0.924	0.607
	OQ372025.1					0.259	0.964	1.000	0.869
	OQ372026.1					0.442	0.944	0.900	0.799
	OQ372028.1					0.155	0.989	1.000	1.000
	OQ372029.1					0.382	0.997	0.843	0.826
OQ372025.1	OQ507152.1					1.000	1.000	0.937	0.964
	OQ372021.1					0.044	0.977	0.775	1.000
	OQ372022.1					0.995	0.759	0.998	1.000
	OQ372023.1					0.205	0.862	0.996	1.000
	OQ372024.1					0.259	0.964	1.000	0.869
	OQ372026.1					1.000	1.000	0.992	1.000
	OQ372028.1					1.000	1.000	1.000	0.961
	OQ372029.1					1.000	1.000	0.980	1.000
OQ372026.1	OQ507152.1					0.999	1.000	1.000	0.928
	OQ372021.1					0.091	0.961	0.997	1.000
	OQ372022.1					1.000	0.704	1.000	1.000
	OQ372023.1					0.363	0.818	1.000	1.000
	OQ372024.1					0.442	0.944	0.900	0.799

	OQ372025.1					1.000	1.000	0.992	1.000
	OQ372028.1					0.998	1.000	0.882	0.924
	OQ372029.1					1.000	1.000	1.000	1.000
OQ372028.1	OQ507152.1					1.000	1.000	0.698	1.000
	OQ372021.1					0.024	0.994	0.462	0.960
	OQ372022.1					0.965	0.854	0.939	0.859
	OQ372023.1					0.119	0.930	0.909	0.783
	OQ372024.1					0.155	0.989	1.000	1.000
	OQ372025.1					1.000	1.000	1.000	0.961
	OQ372026.1					0.998	1.000	0.882	0.924
	OQ372029.1					1.000	1.000	0.822	0.939
OQ372029.1	OQ507152.1					1.000	1.000	1.000	0943
	OQ372021.1					0.074	0.999	0.999	1.000
	OQ372022.1					1.000	0.922	1.000	1.000
	OQ372023.1					0.310	0.970	1.000	1.000
	OQ372024.1					0.382	0.997	0.843	0.826
	OQ372025.1					1.000	1.000	0.980	1.000
	OQ372026.1					1.000	1.000	1.000	1.000
	OQ372028.1					1.000	1.000	0.822	0.939

Appendix 3 B: Table for one way ANOVA and Tukey's HSD post hoc analysis of micro-minerals of indigenous *Physalis* accessions

Accession name (I)	Accession name (J)	ANOVA p-value of Fe content	ANOVA P-value of Zn content	ANOVA p-value of Ni content	ANOVA p-value of Cu content	ANOVA p-value of Li content	ANOVA p-value of Mn content	Tukey's HSD p-value of Fe content	Tukey's HSD p-value of Zn content	Tukey's HSD p-value of Ni content	Tukey's HSD p-value of Cu content	Tukey's HSD p-value of Li content	Tukey's HSD p-value of Mn content
OQ507152.1	OQ372021.1	0.691	0.002	0.007	0.021	0.000	0.140	0.967	0.999	1.000	1.000	0.871	1.000
	OQ372022.1							0.997	1.000	0.999	1.000	1.000	0.988
	OQ372023.1							0.998	0.011	0.036	0.999	0.001	0.999
	OQ372024.1							0.974	0.999	1.000	0.305	1.000	0.298
	OQ372025.1							0.987	0.963	0.987	0.050	0.995	0.965
	OQ372026.1							0.934	0.999	0.956	0.572	1.000	0.868
	OQ372028.1							0.462	0.934	0.098	0.594	0.998	0.416
	OQ372029.1							0.858	1.000	0.999	1.000	1.000	0.493
OQ372021.1	OQ507152.1							0.967	0.999	1.000	1.000	0.871	1.000
	OQ372022.1							1.000	1.000	0.994	1.000	0.584	0.975
	OQ372023.1							1.000	0.003	0.024	1.000	0.017	0.997
	OQ372024.1							1.000	1.000	1.000	0.313	0.630	0.248
	OQ372025.1							1.000	0.711	0.957	0.052	0.420	0.939
	OQ372026.1							1.000	1.000	0.897	0.583	0.643	0.813
	OQ372028.1							0.972	0.636	0.066	0.604	0.487	0.354
	OQ372029.1							1.000	1.000	0.994	1.000	0.564	0.425
OQ372022.1	OQ507152.1							0.997	1.000	0.999	1.000	1.000	0.988
	OQ372021.1							1.000	1.000	0.994	1.000	0.584	0.975
	OQ372023.1							1.000	0.003	0.120	1.000	0.000	1.000
	OQ372024.1							1.000	1.000	0.994	0.434	1.000	0.813
	OQ372025.1							1.000	0.744	1.000	0.082	1.000	1.000
	OQ372026.1							1.000	1.000	1.000	0.723	1.000	1.000
	OQ372028.1							0.872	0.671	0.286	0.744	1.000	0.912
	OQ372029.1							0.998	1.000	1.000	1.000	1.000	0.950
OQ372023.1	OQ507152.1							0.998	0.011	0.036	0.999	0.001	0.999
	OQ372021.1							1.000	0.003	0.024	1.000	0.017	0.997
	OQ372022.1							1.000	0.003	0.120	1.000	0.000	1.000
	OQ372024.1							1.000	0.003	0.024	0.630	0.000	0.646

	OQ372025.1							1.000	0.103	0.205	0.149	0.000	1.000
	OQ372026.1							1.000	0.003	0.287	0.885	0.000	0.995
	OQ372028.1							0.865	0.129	1.000	0.899	0.000	0.782
	OQ372029.1							0.997	0.006	0.120	1.000	0.000	0.848
OQ372024.1	OQ507152.1							0.974	0.999	1.000	0.305	1.000	0.298
	OQ372021.1							1.000	1.000	1.000	0.313	0.630	0.248
	OQ372022.1							1.000	1.000	0.994	0.434	1.000	0.813
	OQ372023.1							1.000	0.003	0.024	0.630	0.000	0.646
	OQ372025.1							1.000	0.716	0.957	0.978	1.000	0.891
	OQ372026.1							1.000	1.000	0.897	1.000	1.000	0.975
	OQ372028.1							0.964	0.642	0.066	1.000	1.000	1.000
	OQ372029.1							1.000	1.000	0.994	0.597	1.000	1.000
OQ372025.1	OQ507152.1							0.987	0.963	0.987	0.050	0.995	0.965
	OQ372021.1							1.000	0.711	0.957	0.052	0.420	0.939
	OQ372022.1							1.000	0.744	1.000	0.082	1.000	1.000
	OQ372023.1							1.000	0.103	0.205	0.149	0.000	1.000
	OQ372024.1							1.000	0.716	0.957	0.978	1.000	0.891
	OQ372026.1							1.000	0.699	1.000	0.832	1.000	1.000
	OQ372028.1							0.938	1.000	0.442	0.815	1.000	0.959
	OQ372029.1							1.000	0.887	1.000	0.136	1.000	0.980
OQ372026.1	OQ507152.1							0.934	0.999	0.956	0.572	1.000	0.868
	OQ372021.1							1.000	1.000	0.897	0.583	0.643	0.813
	OQ372022.1							1.000	1.000	1.000	0.723	1.000	1.000
	OQ372023.1							1.000	0.003	0.287	0.885	0.000	0.995
	OQ372024.1							1.000	1.000	0.897	1.000	1.000	0.975
	OQ372025.1							1.000	0.699	1.000	0.832	1.000	1.000
	OQ372028.1							0.989	0.624	0.565	1.000	1.000	0.995
	OQ372029.1							1.000	1.000	1.000	0.864	1.000	0.999
OQ372028.1	OQ507152.1							0.462	0.934	0.098	0.594	0.998	0.416
	OQ372021.1							0.972	0.636	0.066	0.604	0.487	0.354
	OQ372022.1							0.872	0.671	0.286	0.744	1.000	0.912
	OQ372023.1							0.865	0.129	1.000	0.899	0.000	0.782
	OQ372024.1							0.964	0.642	0.066	1.000	1.000	1.000
	OQ372025.1							0.938	1.000	0.442	0.815	1.000	0.959
	OQ372026.1							0.989	0.624	0.565	1.000	1.000	0.995
	OQ372029.1							0.998	0.834	0.286	0.878	1.000	1.000
OQ372029.1	OQ507152.1							0.858	1.000	0.999	1.000	1.000	0.493
	OQ372021.1							1.000	1.000	0.994	1.000	0.564	0.425

	OQ372022.1							0.998	1.000	1.000	1.000	1.000	0.950
	OQ372023.1							0.997	0.006	0.120	1.000	0.000	0.848
	OQ372024.1							1.000	1.000	0.994	0.597	1.000	1.000
	OQ372025.1							1.000	0.887	1.000	0.136	1.000	0.980
	OQ372026.1							1.000	1.000	1.000	0.864	1.000	0.999
	OQ372028.1							0.998	0.834	0.286	0.878	1.000	1.000

Appendix 3 C: Table for ANOVA and Tukey's HSD post hoc analysis of phytochemical content of indigenous *Physalis* accessions.

Accession name (I)	Accession name (J)	ANOVA p-value of phenolic acid content	ANOVA p-value of tannic acid content	ANOVA p-value of flavonoid content	Tukey's HSD p-value of phenolic acid content	Tukey's HSD p-value of tannic acid content	Tukey's HSD p-value of flavonoid content
OQ507152.1	OQ372021.1	0.159	0.000	0.253	0.945	0.970	0.760
	OQ372022.1				0.265	0.521	0.731
	OQ372023.1				0.484	0.850	0.628
	OQ372024.1				0.449	0.991	0.968
	OQ372025.1				1.000	0.018	0.742
	OQ372026.1				0.958	0.021	0.554
	OQ372027.1				1.000	0.051	0.972
	OQ372028.1				0.958	0.011	1.000
	OQ372029.1				1.000	0.111	1.000
OQ372021.1	OQ507152.1				0.945	0.970	0.760
	OQ372022.1				0.931	0.990	1.000
	OQ372023.1				0.994	1.000	1.000
	OQ372024.1				0.991	0.527	1.000
	OQ372025.1				0.995	0.178	1.000
	OQ372026.1				1.000	0.205	1.000
	OQ372027.1				0.993	0.391	1.000
	OQ372028.1				1.000	0.118	0.933
	OQ372029.1				0.997	0.626	0.664
OQ372022.1	OQ507152.1				0265	0.521	0.731
	OQ372021.1				0.931	0.990	1.000
	OQ372023.1				1.000	1.000	1.000
	OQ372024.1				1.000	0.115	1.000
	OQ372025.1				0.466	0.675	1.000
	OQ372026.1				0.913	0.724	1.000
	OQ372027.1				0.442	0.917	1.000
	OQ372028.1				0.912	0.534	0.917
	OQ372029.1				0.496	0.990	0.632
OQ372023.1	OQ507152.1				0.484	0.850	0.628
	OQ372021.1				0.994	1.000	1.000
	OQ372022.1				1.000	1.000	1.000
	OQ372024.1				1.000	0.308	0.998
	OQ372025.1				0.722	0.339	1.000
	OQ372026.1				0.991	0.381	1.000
	OQ372027.1				0.698	0.628	0.997
	OQ372028.1				0.990	0.237	0.850

	OQ372029.1				0.751	0.847	0.526
OQ372024.1	OQ507152.1				0.449	0.991	0.968
	OQ372021.1				0.991	0.527	1.000
	OQ372022.1				1.000	0.115	1.000
	OQ372023.1				1.000	0.308	0.998
	OQ372025.1				0.687	0.002	1.000
	OQ372026.1				0.986	0.003	0.994
	OQ372027.1				0.662	0.006	1.000
	OQ372028.1				0.986	0.001	0.998
	OQ372029.1				0.717	0.015	0.931
OQ372025.1	OQ507152.1				1.000	0.018	0.742
	OQ372021.1				0.995	0.178	1.000
	OQ372022.1				0.466	0.675	1.000
	OQ372023.1				0.722	0.339	1.000
	OQ372024.1				0.687	0.002	1.000
	OQ372026.1				0.997	1.000	1.000
	OQ372027.1				1.000	1.000	1.000
	OQ372028.1				0.997	1.000	0.923
	OQ372029.1				1.000	0.995	0.644
OQ372026.1	OQ507152.1				0.958	0.021	0.554
	OQ372021.1				1.000	0.205	1.000
	OQ372022.1				0.913	0.724	1.000
	OQ372023.1				0.991	0.381	1.000
	OQ372024.1				0.986	0.003	0.994
	OQ372025.1				0.997	1.000	1.000
	OQ372027.1				0.996	1.000	0.993
	OQ372028.1				1.000	1.000	0.793
	OQ372029.1				0.998	0.997	0.455
OQ372027.1	OQ507152.1				1.000	0.051	0.972
	OQ372021.1				0.993	0.391	1.000
	OQ372022.1				0.442	0.917	1.000
	OQ372023.1				0.698	0.628	0.997
	OQ372024.1				0.662	0.006	1.000
	OQ372025.1				1.000	1.000	1.000
	OQ372026.1				0.996	1.000	0.993
	OQ372028.1				0.996	0.999	0.999
	OQ372029.1				1.000	1.000	0.938
OQ372028.1	OQ507152.1				0.958	0.011	1.000
	OQ372021.1				1.000	0.118	0.933
	OQ372022.1				0.912	0.534	0.917
	OQ372023.1				0.990	0.237	0.850
	OQ372024.1				0.986	0.001	0.998
	OQ372025.1				0.997	1.000	0.923
	OQ372026.1				1.000	1.000	0.793
	OQ372027.1				0.996	0.999	0.999

	OQ372029.1				0.998	0.975	1.000
OQ372029.1	OQ507152.1				1.000	0.111	1.000
	OQ372021.1				0.997	0.626	0.664
	OQ372022.1				0.496	0.990	0.632
	OQ372023.1				0.751	0.847	0.526
	OQ372024.1				0.717	0.015	0.913
	OQ372025.1				1.000	0.995	0.664
	OQ372026.1				0.998	0.997	0.455
	OQ372027.1				1.000	1.000	0.938
	OQ372028.1				0.998	0.975	1.000

Appendix 3 D: . Table for one way ANOVA and Tukey's HSD post hoc analysis of DPPH RSA and HRSA of indigenous *Physalis* accessions.

Accession name (I)	Accession name (J)	ANOVA p-value of DPPH RSA	ANOVA p-value of HRSA	Tukey's HSD p-value of DPPH RSA	Tukey's HSD p-value of HRSA
OQ507152.1	OQ372021.1	0.000	0.013	0.000	0.999
	OQ372022.1			0.000	0.579
	OQ372023.1			0.692	0.052
	OQ372024.1			0.004	0.924
	OQ372025.1			0.001	1.000
	OQ372026.1			0.000	1.000
	OQ372027.1			0.000	0.071
	OQ372028.1			0.000	1.000
	OQ372029.1			0.000	0.869
OQ372021.1	OQ507152.1			0.000	0.999
	OQ372022.1			0.452	0.937
	OQ372023.1			0.000	0.203
	OQ372024.1			0.082	1.000
	OQ372025.1			0.277	1.000
	OQ372026.1			1.000	1.000
	OQ372027.1			1.000	0.261
	OQ372028.1			1.000	1.000
	OQ372029.1			1.000	0.998
OQ372022.1	OQ507152.1			0.000	0.579
	OQ372021.1			0.452	0.937
	OQ372023.1			0.027	0.887
	OQ372024.1			0.986	0.999
	OQ372025.1			1.000	0.750
	OQ372026.1			0.387	0.910
	OQ372027.1			0.317	0.937
	OQ372028.1			0.252	0.646
	OQ372029.1			0.777	1.000
OQ372023.1	OQ507152.1			0.692	0.052
	OQ372021.1			0.000	0.203
	OQ372022.1			0.027	0.887
	OQ372024.1			0.198	0.512
	OQ372025.1			0.054	0.091
	OQ372026.1			0.000	0.174
	OQ372027.1			0.000	1.000
	OQ372028.1			0.000	0.065
	OQ372029.1			0.001	0.606

OQ372024.1	OQ507152.1			0.004	0.924
	OQ372021.1			0.082	1.000
	OQ372022.1			0.986	0.999
	OQ372023.1			0.198	0.512
	OQ372025.1			0.999	0.981
	OQ372026.1			0.065	0.999
	OQ372027.1			0.049	0.606
	OQ372028.1			0.036	0.953
	OQ372029.1			0.222	1.000
OQ372025.1	OQ507152.1			0.001	1.000
	OQ372021.1			0.277	1.000
	OQ372022.1			1.000	0.750
	OQ372023.1			0.054	0.091
	OQ372024.1			0.999	0.981
	OQ372026.1			0.229	1.000
	OQ372027.1			0.182	0.122
	OQ372028.1			0.140	1.000
	OQ372029.1			0.574	0.958
OQ372026.1	OQ507152.1			0.000	1.000
	OQ372021.1			1.000	1.000
	OQ372022.1			0.387	0.910
	OQ372023.1			0.000	0.174
	OQ372024.1			0.065	0.999
	OQ372025.1			0.229	1.000
	OQ372027.1			1.000	0.226
	OQ372028.1			1.000	1.000
	OQ372029.1			0.999	0.995
OQ372027.1	OQ507152.1			0.000	0.71
	OQ372021.1			1.000	0.261
	OQ372022.1			0.317	0.937
	OQ372023.1			0.000	1.000
	OQ372024.1			0.049	0.606
	OQ372025.1			0.182	0.122
	OQ372026.1			1.000	0.226
	OQ372028.1			1.000	0.087
	OQ372029.1			0.998	0.699
OQ372028.1	OQ507152.1			0.000	1.000
	OQ372021.1			1.000	1.000
	OQ372022.1			0.252	0.646
	OQ372023.1			0.000	0.065
	OQ372024.1			0.036	0.953
	OQ372025.1			0.140	1.000
	OQ372026.1			1.000	1.000
	OQ372027.1			1.000	0.087
	OQ372029.1			0.992	0.910

OQ372029.1	OQ507152.1			0.000	0.869
	OQ372021.1			1.000	0.998
	OQ372022.1			0.777	1.000
	OQ372023.1			0.001	0.606
	OQ372024.1			0.222	1.000
	OQ372025.1			0.574	0.958
	OQ372026.1			0.999	0.995
	OQ372027.1			0.998	0.699
	OQ372028.1			0.992	0.910