IDENTIFICATION OF TEA VARIETIES COLLECTED FROM NORTHERN OF THAILAND USING DNA BARCODES

PHIJITTRA UMALEE

MASTER OF SCIENCE
IN
BIOLOGICAL SCIENCE

SCHOOL OF SCIENCE
MAE FAH LUANG UNIVERSITY
2022

IDENTIFICATION OF TEA VARIETIES COLLECTED FROM NORTHERN OF THAILAND USING DNA BARCODES

PHIJITTRA UMALEE

THIS THESIS IS A PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
BIOLOGICAL SCIENCE

SCHOOL OF SCIENCE
MAE FAH LUANG UNIVERSITY
2022

IDENTIFICATION OF TEA VARIETIES COLLECTED FROM NORTHERN OF THAILAND USING DNA BARCODES

PHIJITTRA UMALEE

THIS THESIS HAS BEEN APPROVED
TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGICAL SCIENCE
2022

EXAMINATION COMMITTEE

Nakkuntod M. .....CHAIRPERSON

(Asst. Prof. Maliwan Nakkuntod, Ph. D.)

Somrudee .....ADVISOR

(Asst. Prof. Somrudee Nilthong, Ph. D.)

Siam Popluechai .....CO-ADVISOR

(Asst. Prof. Siam Popluechai, Ph. D.)

.....EXAMINER

(Asst. Prof. Eleni Gentekaki, Ph. D.)

ACKNOWLEDGEMENTS

I would like to thank to everyone who helped in my thesis. Special thanks to my advisor, Asst. Prof. Dr. Somrudee Nilthong, and co-advisor, Asst. Prof. Dr. Siam Popluechai who always provided knowledge, ideas, and helping for the past three years. I would also like to grateful to Dr. Eleni Gentekaki, for suggestion about using the program and interpreting the sequence data.

I also could not have undertaken this journey with out chiar person, Asst. Prof. Dr. Maliwan Nakkuntod for advice in improving my thesis.

I would like to thank tea and coffee institute of Mae Fah Luang university for assisting in collection of wild tea plants.

Lastly, I would be remiss in not mentioning my parents and friends who always love and support me whatever I pursue.

Phijittra Umalee

Thesis Tittle	Identification of Tea Varieties Collected from Northern of Thailand Using DNA Barcodes
Author	Phijittra Umalee
Degree	Master of Science (Biological Science)
Advisor	Asst. Prof. Somrudee Nilthong, Ph. D.
Co-Advisor	Asst. Prof. Siam Popluechai, Ph. D.

ABSTRACT

Tea plant (*Camellia sinensis* (L.) O. Kuntze) is an economically important crop in Thailand. Their leaves are frequently used as the primary ingredients in a variety of tea preparation. There are no reports on barcode analysis of wild tea plants in Thailand. Thus, the aimed of this study was to create a DNA barcode of 135 wild tea plants using the *matK*, and *psbA- trnH* genes. Both DNA barcodes were 100% successfully amplified and sequenced. The *matK* gene showed a good result in identification efficiency, accurately demonstrating 100% identity to the expected species, *C. sinensis* in 64 tea plants. Some tea plants displayed 100% identification with the other *Camellia* species, whereas the remaining tea plants displayed 99.10% to 99.89% identity with *C. sinensis*. In *psbA-trnH* just only one tea plant was detected 100% identity to *C. sinensis* and the remaining showed 97.24%-99.80%. Based on the blast results, all of the tea plants in this study were identified as *C. sinensis*. According to sequence alignment, *matK* detected 12 SNPs in 72 tea plants, which is more than *psbA-trnH* did (7 SNPS in 72 tea plants). The *psbA-trnH* showed the highest genetic distance (0.0131), which might be caused by indels. Phylogenetic trees clearly separated all wild tea plants into numerous groups. This demonstrated that tea plant with the same SNP position and

pattern of SNPs are likely to be classified into the same cluster, though this was not entirely consistent with the morphological tree. As a results, DNA barcoding is a practical technique for identifying plant species, and the *matK* gene was thought to be a barcode for wild tea plants.

Keywords: *C. sinensis*, DNA Barcoding, *matK*, *psbA-trnH*, Tea Plant

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

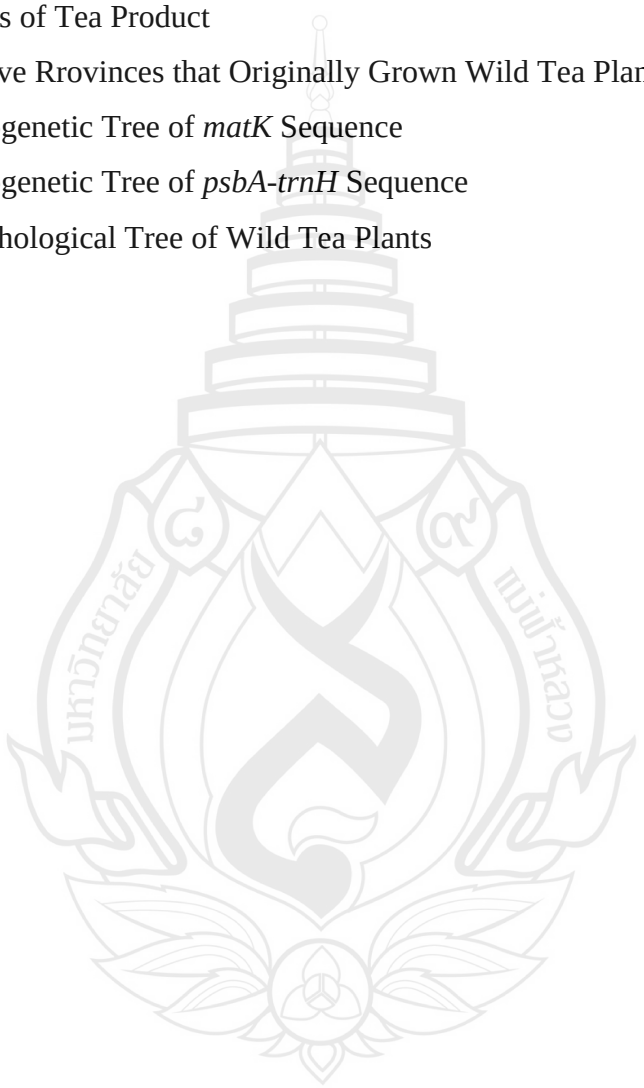
LIST OF TABLES

LIST OF TABLES (continued)

Table **Page**

LIST OF FIGURES

Figure **Page**



ABBREVIATION AND SYMBOL

A	Adenine base
Asst. Prof.	Assistant Professor
C	Cytocine base
cm.	Centimetre
Dr	Doctor
et al	et alii
Fre.	Frequency
G	Guanine base
p. (pp.)	page (pages)
T	Thymine base

CHAPTER 1

INTRODUCTION

Tea (*Camellia sinensis* (L.) O. Kuntze) has played an important role in human life since the ancient time. Tea is the most popular non-alcoholic beverage throughout the world (Hilal & Engelhardt, 2007). Especially in the Asia region, tea has been widely consumed because of its pleasant taste, aroma, and health-promoting such as delaying aging (Afzal et al., 2015), lowering blood lipids, lowering blood glucose level (Tsuneki et al., 2004), preventing high blood pressure (Chacko et al., 2010) and diabetes (Tang et al., 2013). Tea plant has been cultivated for thousands of years and their leaves have been applied for medicinal purposes until today (Khan & Mukhtar, 2013). China is the first country to utilize leaves of tea plant as medicine for curing, and later as an infusion beverage for more than 3,000 years (Eden, 1958). Leaves of tea plant are rich in bioactive compounds such as amino acids, vitamins, carbohydrates, caffeine, and purine alkaloids, especially polyphenols (Ye, 2012), which are used as the source for the most popular beverages such as black, green, oolong, and white tea. Of those tea products, black tea is the most widely consumed worldwide about 78%, which is mainly consumed in the Western countries, especially India. Green tea is the second most consumed drink in the world about, 20%, which is usually consumed in Asian countries such as Japan, Taiwan, and China. Oolong tea is only 2% worldwide consumption (Khan & Mukhtar, 2007)

In Thailand, tea plants are mainly grown in the northern part of the country, in Chiang Rai, Chiang Mai, Mae Hong Son, Phrea, Nan, Lampang, and Tak provinces. Tea plants are an economically important crop. Thai tea market has grown gradually in the past few years. Thailand is the no.1 of exporter tea products in ASEAN and no. 8 in the world in 2021 (Thansettakij, 2022). In 2021, Thailand exported 10.6 million tons of instant tea to the world market, worth 38.8 million US dollars, increasing 19% from 2020 (Thansettakij, 2022). The major export markets of Thailand are ASEAN (19.1

million USD), United state (6.8 million USD), Japan (3.7 million USD), China (3.6 million USD), and Hong Kong (1.5 million USD). Moreover, green tea leaves and black tea were exported 3.5 thousand to the world market worth 24.9 million USD, which has increase by 4.9% since 2020. The main export markets are ASEAN (14.8 million USD), United states (3.6 million USD), and China (2.7 million USD) (Thansettakij, 2022). The ratio of tea in the market can be divided into two parts in which 30% of the production is commercialized in the domestic market, while the remaining 70% is exported (Theppakorn et al., 2014). In beverage markets, tea is also earning a significant share of this market. Green tea, oolong tea, and black tea are the most popular tea infusion beverages in the world market. In addition, Thai tea made from black tea is very popular in foreigners and Thai people. All tea types use tea leaves as the main ingredient. Leaves, buds, and delicate stems of two widely cultivated tea plants, namely China tea (*Camellia sinensis* var *sinensis)* and Assam tea (*Camellia sinensis* var. *assamica*) are commonly used as an important material for making various tea products (Meng et al., 2018). Assam tea is more popular than China tea which is used for producing various beverages such as green tea, black tea, and Thai tea. In addition, leaves of Assam tea are also used for making chewing tea or 'miang' in Thai language. Chewing tea is made by steamed and fermented tea leave, which is eaten with side dishes such as salt, oil, or garlic as a snack (Khanongnuch et al., 2017) and often served for visitors and also in traditional religious ceremonies especially in Northern part of Thailand. Nowadays, people are more interested in high-quality products with a great stories and drinks that give them specific health benefits. Therefore, high quality and purity of tea varieties are needed for making tea products. Generally, different types of tea can give the difference in taste or aroma as it has a different chemical composition. Based on the morphological features, each tea variety has very similar appearances such as all shrubs or tree, colour and shape of leave, flowers and leaves arrangement. It is easy to contaminate with other tea varieties, wild plants, or herbs. The traditional techniques are used for identifying plant species such as the organoleptic test, microscopy test, and microscopy test. Leave, young shoot, flower, fruit, or seed must be present for use in these techniques. Chemical analysis is also used to identify plant species by using chemical composition, which is specific in each plant species. Of those techniques, there will be some limitations caused by environmental factors,

plant age, or genetic relationships between varieties (Lee et al., 2016). Additionally, tea plant is predominantly cross-pollinated and highly heterogeneous plants, which made more difficult to identify (Lee et al., 2019). The misidentification can also occur from the phenotypic plasticity, which is the ability of a solitary organism to switch its morphology to stress environmental conditions (Schlichting, 1986) and the genetic variation that is used to identify varieties may lead to wrong identifications. However, the chemical composition can be affected by many reasons such as geographic location, season variations, and horticultural practice (Rubel Mozumder et al., 2020). All traditional and chemical techniques mostly rely on the expertise of researchers and require a long time to be investigated and high risk of wrongly identified (Chen et al., 2014a). Using DNA-based techniques to identify plants is an alternative technique for identifying or discriminating organisms. DNA barcoding has been proposed as an effective tool using DNA barcode for identification because it does not depend on external factors such as morphological traits, climate condition, weather or environment. Because of its slowly evolving nature and relatively low rate of nucleotide substitution, the COI gene found in the mitochondrial genome is not suited for plants. The high rate of replacement, DNA barcodes like *matK, rbcL, rpoC1, atpF-atpH, psbA-trnH,* and *psbK-psbl* have been substituted for nuclear or chloroplast genome in plants. The two types of DNA substitution mutation are transition (the interchanging of two-ring purines or one-ring pyrimidines) and transversion (interchanging of purine for pyrimidine bases).

1.1 Objectives of This Study

1.1.1 To evaluate the effectiveness of the four potential DNA barcodes (*rbcL, matK, psbA-trnH,* and *rps7-trnV*) utilized in this study to distinguish between wild tea plants.

1.1.2 To examine the level of genetic relationship between wild tea plants in Tea germplasm.

1.1.3 To identify potential genetic resources for future tea breeding program.

1.2 Scope of Research

In this investigation, 135 wild tea plants collected from tea germplasm at Boon Rawd Farm in the province of Chiang Rai were used. Originally, tea plants were grown in Chiang Rai, Chiang Mai, Nan, Phrea, Lampang, and Mae Hong Son provinces, Thailand. The morphology (color of young tips, leaf apex, leaf margins, leaf bases, and leaf shape) and DNA barcodes of tea plants were analysed. All tea plants were amplified and sequenced. The raw sequences were performed alignment for finding Single Nucleotide Polymorphisms (SNPs), genetic distance, and phylogenetic tree.

CHAPTER 2

LITERATURE REVIEW

Tea plant (*Camellia sinensis* (L.) O. Kuntze) is an evergreen flowering shrub belonging to the family Theaceae, in genus *Camellia* section *Thea*. In *Camellia* genus contains more than 250 species of Asian evergreen shrubs and trees, which have economically various crucial plants including tea plants, oil camellia and many ornamental flowering species (Zhao et al., 2017). The tea plants are an important crop for human life for a long time. Their leaves or buds are used to produce flourishing infusion tea (Agarwal et al., 2017). The original region of tea was near the Irrawaddy River in Burma and spread into South-Eastern China, westward into upper Burma and Assam (North-Eastern India) (Dattner et al., 2003). The other report suggested that the origin of tea plants was in the mountainous regions of Southwestern China and Southeast Asia (Hayat et al., 2015). Therefore, tea plant is considered to be around the meeting points of the land of north Burma and southwest China (Li & Zhu, 2016). Nowadays, tea plant is widely cultivated in many countries around the world, which includes tropical and subtropical regions such as Asian, African, and Latin American countries (Tang et al., 2020). The commercial tea cultivars are mainly classified as three taxon, *Camellia sinensis* var. *sinensis* (L.) O. Kuntze, *Camellia sinensis* var. *assamica* (Masters) Wight and *Camellia sinensis ssp*. *lasiocalyx* (Planchon *ex* Watt), which are known as China, Assam, and Cambod types, respectively. China type is mainly cultivated in South China and in some Southeast Asian countries. Assam type is cultivated in India and other tea growing countries such as Sri-Lanka, Indonesia, and Bangladesh. Cambodia type is mostly cultivated in Indo-China region (Meegahakumbura et al., 2018). Only China and Assam types are used for producing various tea products. Usually, China type is used to produce green tea while black tea is produced from Assam type. For tea plant horticulture, tea plants are cultivated by keeping height only 0.6-1.0 m, which is easy for leave plucking. In the forest, tea plants

are grown as a green shrub and reaching a height of 15 m. For the propagation, tea plants can be propagated by seed, grafting, or plant tissue culture, which need 4-12 years for a plant to bear seed and 3 years later for harvesting (Li & Zhu, 2016). The maximum yield for collecting tea leave is between 30 to 50 years.

As we all known that tea is consumed as a healthy beverage because of its medicinal purposes. In tea leaves are rich of bioactive compound such as phenolic compound (phenolic acids, flavonoids, and tannins), alkaloids (methylxanthines) and nutrients (amino acids, carbohydrates, proteins, chlorophyll, volatile compounds, fluoride, mineral and several undefined compounds), which are responsible for many medicinal benefits (Graham, 1992; Sharma et al., 2021). The phenolic compounds are predominantly found in tea leave up to 30% of their dry weight, especially flavan-3-ols or catechins. The catechins are colorless, astringent, and water-soluble, namely catechins, epicatechin (EC), gallocatechin (GC), epicatechin gallate (ECG), epigallocatechin (EGC) and epigallocatechin gallate (EGCG), which can occur in abundant amounts in the tea beverages (Wu & Wei, 2009). Especially, EGCG and ECG are significant for biological functions such as antioxidants, anti-rancidity and antimicrobial activities (Wu & Wei, 2009). Additionally, it contains many types of amino acids, such as aspartic acid, serine, histidine, arginine, γ-aminobutyric acid (GABA), threonine, tyrosine, valine, methionine, leucine, phenylalanine, lysine and theanine (D. Chen et al., 2020). The L-theanine and GABA mainly found in white tea (Afzal et al., 2015). The L-theanine has been involved in many health benefits such as neuroprotection, mental relaxation, hypolipidemic, regulation of systemic blood pressure, and suppression of the stimulatory action of caffeine (Vuong et al., 2011). While GABA is the most significant inhibitory neurotransmitter in the human central nervous system (Da Silva Pinto, 2013). Each variety or manufacturing type of tea plants will give the difference in taste, aroma, health effects, especially chemical composition such as polyphenols (mainly flavonoids and catechins). Tea products can be divided into many types according to different manufacturing types including non-fermented tea, semi-fermented tea, completely-fermented tea and post-fermented (KC et al., 2020). Green tea, white tea, oolong tea, black tea, and pu-erh tea belong to those four types of manufacturing, respectively. The polyphenol compound can be found in green tea more than black and oolong teas because polyphenol compound in black and oolong

teas will transform into theaflavins and thearubigins in the fermentation process (Yang et al., 2002).

However, the constituents in tea leaves depend on the species, season, age of the leaf (plucking position), climate, and horticultural practices (Lin et al., 1996). The quality of tea leaves is an important for maintaining the medicinal benefits (polyphenol content) and taste. Identification techniques with accuracy power require to prevent misidentification.

2.1 Basic Varieties of Tea Products

2.1.1 White Tea

White tea is made from new growth buds and young leaves with tiny and silvery hairs (Alcázar et al., 2007). Being the least processed tea, white tea has a high percentage of polyphenol contents because polyphenol compounds can be found in high amounts in the bud and gradually decrease with age of leaves (Soni et al., 2015). Therefore, white tea is considered as a high therapeutic value. For making white tea, buds and leaves are plucked before they are open, then withered or steamed as quickly as possible to inactivate the oxidation process, and drying (Kosińska & Andlauer, 2014). White tea will give fresh, subtle, and comparatively sweet. White tea is very famous among the health conscious people especially in America and Europe.

2.1.2 Green Tea

Green tea is made from fresh tea leaves. Tea leaves are inhibited from fermentation, which make this tea to retain most of its nutrients and antioxidant properties. From all types of tea, green tea is the richest of catechins (Kosińska & Andlauer, 2014). For green tea process, tea leaves are immediately steamed or fired as soon as possible to inactivate the polyphenol oxidase and then, dried for preservation (Soni et al., 2015). Differences in heat, rolling, and drying can affect on the flavours of green tea. The process of making green tea is very similar to that of white tea.

2.1.3 Oolong Tea

Oolong tea, tea leaves are partially fermented and still retain natural polyphenol. To make oolong tea, after plucking, tea leaves are withered under the sunlight for 30 to 60 minutes. Then, tea leaves will be transferred for withering without the sunlight for 6 to 8 hours with or without gentle rolling. Finally, withering tea leaves were dried to inactivate the enzyme in the fermentation process. The colour of oolong tea is half between green and black tea.

2.1.4 Black Tea

Black tea, tea leaves are completely fermented. Theaflavins and thearubigins are the major polyphenolic compounds in black tea, which occurs in the fermentation process (Koch, 2019). In the fermentation process, catechins are converted into theaflavins and thearubigins by polyphenol oxidase (Nagini et al., 2013). To produce black tea, tea leaves are withered without sunlight for 18 hours. Then rolling, in this step, tea leaves are crushed or macerated to break the cell structure to promote the enzymatic oxidation known as fermentation process. Finally, tea leaves are fired to halt enzyme activation. Black tea leaves are a dark red hue and a sweet aroma of malt sugar.

2.1.5 Pu-erh Tea

Pu-erh tea is post-fermented, using a unique microbial fermented. Processing of making pu-erh tea is similar to green tea, using sun-dried leaves of tea plant as a raw material. Fresh tea leaves are spread out for 8 hours to partially dry after harvesting. Tea leaves are then fixed by heating to inactivate the endogenous polyphenol oxidase (Liang et al., 2005). Tea leaves will be rolling in a short of time. Before the post-fermented step, tea leaves are partially dried to give moisture content under the sun for 3 to 5 hours. In the post-fermented process, sun-dried leaves are piled up for a week for a series of oxidation and degradation of tea chemical constituents (Shao et al., 1995). It is most often commercialized into many forms of compressed brick, cake log, nest, or gourd shape after autoclaving and drying (Ahmed et al., 2010). Pu-erh tea is very popular in Southeast Asia because of its health benefits (Mok et al., 2008) as well as flavor and taste (Ahmed et al., 2010).

Note (1) White tea, (2) Green tea, (3) Oolong tea, (4)Black tea, and (4) Pu-erh tea

Source Silk Road Teas (2022)

Figure 2.1 Five Types of Tea Product

2.2 Process of Tea Manufacturing

2.2.1 Plucking

Plucking is an essential process. Normally, the fresh green tea leaves are plucked by hand. The rapidly growing shoot tips down to about the second or third unfolded leaf is plucked and used

2.2.2 Withering

After plucking, tea leaves will undergo withering. In the withering step, the plucked tea leaves are allowed for withering to remove moisture content and preparing for the rolling step. This step can occur under the sun or on a rack in a heated room. The aim of the withering is to improve the flavor in green, oolong, and black tea

2.2.3 Rolling

The rolling step is performed to obtain the different kinds of shape of tea leaves. Tea leaves will go through maceration on a rolling table, which can both disrupt the leaf cells and give the leaves a fine twisted shape. The aim of this step is to release tea oil and allow the access of enzymes to polyphenol, which is important for the oxidation process (Kosińska & Andlauer, 2014).

2.2.4 Oxidation/Fermentation

In this step, the rolled leaves are spread out on the tray and left in a room with controlled temperature, humidity, and aeration for 2-4 hours (Kosińska & Andlauer, 2014). During this time, polyphenol oxidase catalyzes the polyphenols into tannis and complex quinones.

2.2.5 Fixation

This step can be done by steaming, pan-fired, and sunning. The aim of the fixation step is to inactivate the endogenous enzymes presented in the fresh tea leaves.

2.2.6 Drying

Drying is the last step in the manufacturing of tea. Tea leaves undergo a moisture reduction to a final moisture content of approximately 3-4% during the drying process to facilitate handling and transportation. Thereafter, tea is sorted and packed.

2.3 Tea Plant

2.3.1 China Tea (*Camellia sinensis* var. *sinensis)*

The China tea plant is a big shrub, approximately 2-6 m tall shrub, which has many straight or slender stems arising from the base of the plant. Mature leaves are small (3-6 cm long) relatively erect, dark green with smooth, and matt surface (De Costa et al., 2007). Leave margins are bluntly serrated in shape like a saw blade. Young shoots of China tea plant are garnet- brown to purple. Normally, China tea plants grow faster than Assam tea plants and can resist low temperatures and variable environments.

2.3.2 Assam Tea (*Camellia sinensis* var. *assamica)*

Assam tea plant is a small tree, approximately high 6-18 m. Mature leaves are big (15-20 cm long), light green with a glossy surface and persistently hairy on the midrib below (De Costa et al., 2007). The leave shape is elliptic, elliptical-oblong to narrowly elliptical-oblong, or elliptic-obovate. Leaves are almost acuminate on the apex and cuneate to acute or rarely rounded at base. Leave margins are denticulate to bluntly and distantly undulate-serrate or serrulate with incurved and black tipped teeth along the margins. Young shoots of Assam tea are light green to green. Leaves of Assam tea are bigger than China tea. This tea is drought tolerant and adapts well to the environment, which grows well in the forest.

2.3.3 Cambod Tea (*Camellia assamica* ssp. *Lasiocalyx* (Planchon *ex* Watt))

Cambod tea plant is a shrub or under tree about 6 m tall with many elected branches. Leave are broadly elliptic, erect, coriaceous, glossy and non-punctate above, glabrous on both surfaces and glossy. Leave are shortly and bluntly acuminate at the apex. Leave margins are denticulate with incurred and black-tipped teeth along the margins. Young leaves are oxblood.

2.4 DNA Barcoding

DNA barcoding was initially proposed by Hebert and coworkers from the University of Guelph, Ontario, Canada in 2003. DNA barcoding is increasingly becoming a more popular tool for species identification because of its rapid, accuracy, and affordable cost. The concept of DNA barcoding is using short standard of nucleotide sequences of one or more specific regions from either the nuclear or chloroplast genome to identify species called "DNA barcode" (Parveen et al., 2017). It has been widely used for identifying and classifying animals, plants or fungal. The mitochondrial cytochrome *C oxidase1 (COI)* gene was suggested to be a standard DNA barcode for animal identification (Techen et al., 2014). However, the mitochondrial genes in plants are slowly evolving with a low mutation rate. Therefore, mitochondrial cytochrome *C oxidase1* cannot be used in plants (Sun et al., 2016). Then, the search of plant DNA barcodes was shifted to chloroplast and nuclear genomes because of high mutation rates. For plants, several gene regions have been used as DNA barcodes, such as Internal transcribed spacer (ITS), ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL)*, Maturase K (*matK)*, the intergenic spacer *atpF-atpH* (*atpF-atpH*), the chloroplast *psbK-psbI* intergenic region (*psbK-psbI)*, and the *trnH–psbA* intergenic spacer region (*trnH-psbA)*. Four DNA barcodes (*rbcL*, *matK*, *trnH-psbA*, and ITS) have been proposed and agreed as a primary standard DNA barcode for plants (CBOL Plant Working Group, 2009). Two genes, *rbcL* and *matK* genes from the chloroplast genome have been proposed as a core barcodes (CBOL Plant Working Group, 2009). Meanwhile, the Internal Transcribed Spacer (ITS and ITS2) regions and chloroplast intergenic spacer (*trnH-psbA)* have been proposed as supplementary DNA barcodes for plants. The ITS and ITS2 regions have also been applied for fungal kingdom diagnostics (Bellemain et al., 2010). DNA barcodes have become a useful tool for biodiversity investigation, monitoring, molecular phylogeny, and evolution because of more effective compared to traditional techniques. Traditional techniques are mainly based on experience of the observers such as morphological characters or chemical profiling tests, which require significant training to be perform accurately, time-consuming, and laborious. The DNA barcoding technique is accurate, rapid, and

automated species identification, which non-expert or experience can perform well (Hebert et al., 2003). Furthermore, this technique is not restricted by morphology characteristics (organ, tissue difference, or growth stage), physiological conditions and allows species authentication without specialist taxonomic knowledge (Chen et al., 2014a). DNA barcoding techniques are normally used for species identification, discovering species, or authentication from other species. For instance of using DNA barcoding, five DNA barcodes (*matK, rbcL, ITS2, psbA-trnH* and *ycf5*) were applied to select the best DNA barcode for plant identification in *Araliaceae*. The results showed that *ITS2* is the best and suitable barcode in the family *Araliaceae*, whereas *ycf5* poorly showed result in PCR amplification (Z. Liu et al., 2012). Similarly, three candidates of DNA barcodes (*ITS2, trnF-L,* and *rbcL*) were used to identify *Isatis indigotica* Fort. and its adulterants (*I. minima* and *I. oblongata*), which the ITS2 sequence demonstrated to efficiently distinguish I. *indigotica* from its adulterants (Chen et al., 2014b). For poisonous plants, DNA barcoding helped differentiate the poisonous species from the edible species, *Melientha suavis* Pierre and *Sauropus androgynus* (L.) Merr. (Thongkhao et al., 2020)

Not just identification purpose, DNA barcodes can serve in many other ways such as supporting food safety and purity of labeling on botanical products, which disclosing cryptic species (Wallace et al., 2012). Six DNA barcode regions (ITS, *trnH-psbA, matK, rbcL, trnL-trnF,* and *psbB-psbF*) were tested for differentiate *Dumasia* species. The results showed that ITS alone or in combination with any of the cpDNA markers had the highest success rate of accuracy species identification (>99.6%). Phylogenetic tree showed that *Dumasia yunnanensis* was separated into two group, one with *Dumasia cordifolia* and another with *Dumasia forestii.* Two groups of *Dumasia* species have no morphologically distinguish. As a result, these two *Dumasia* species may be cryptic species of *Dumasia yunnanensis* (Jiang et al., 2020). In forensic, DNA barcoding is helping for linking the biological evidences into the crime scenes (Yoon, 1993). Park et al. (2017) suggested that DNA barcoding technique is an effective tool for identifying plant evidents. The combination of *rbcL* and *psbA-trnH* markers were the best markers for botany species identification in forensics (Ferri et al., 2015). DNA barcoding can be used to address ecology, evolution, and conservation issues such as

the ecological rules controlling the assembly of species in plant communities (Kress et al., 2009).

2.5 DNA Barcode that Commonly Used in DNA Barcoding

2.5.1 *matK*

The maturase *K* gene (*matK*) is located in the group II intron between the 5′ and 3′ exons of *trnK* at the large single copy region of the chloroplast genome that encodes for enzyme maturase in most plants. Maturase is an enzyme that catalyzes non autocatalytic intron removal from premature RNAs, the splicing of type-II introns from RNA transcripts (Kang et al., 2017). This gene can increase a copy of DNA target easily using the PCR technique. The *matK* gene is one of the most rapidly evolving genes when compares with other genes because it has high substitution rates within species. Moreover, it has been proposed as highly conserved in plant systematics. Using the *matK* gene, the identification of medicinal plants in the Tanzanian market, *matK* gene showed the highest species-level discrimination (Veldman et al., 2020). The *matK* gene has been reported as a candidate barcode for *Lavandula* species identification (Soares et al., 2018). In another report, the *matK* gene had sequence divergence and sequence variation (SNP) of DNA sequence between T. *edulis* and its adulterants much higher than the *rbcL* and *psbA-trnH* genes. Furthermore, *matK* could correctly identify and distinguish T. *edulis* from its adulterants (Ma et al., 2014). In tea plant, *matK* gene was used for studying genetic diversity and variation of some tea clones of Darjeeling and Dooars. A total of nine substitution sites were detected in the tea samples. However, the *matK* gene has been successfully applied as a single locus to study intraspecific variation but not being 100% conserved with the same species as *Camellia* (Reha et al., 2021).

2.5.2 *psbA-trnH*

The chloroplast intergenic *psbA-trnH* spacer gene (*psbA-trnH*) is a non-coding region and located at the intergenic spacer region in the chloroplast, which is the region between the *psbA* and histidine transfer RNA (*trnH*) genes. The gene *trnH* controls the formation of histidine protein, while the gene *psbA* produces proteins D1 and D2 that

belong to the reaction centre of photosystem II (Mulo et al., 2012). The chloroplast intergenic spacer *psbA-trnH* gene is one of the most successfully amplified with an optimum ratio of length and number of informative positions (Kress et al., 2005). This non-coding region not only shows high sequence divergence, but also high rates of insertion/deletion (Kress & Erickson, 2007). Usually, the usefulness of *psbA-trnH* spacer has been used as a phylogenetic marker at lower taxonomic levels and suitable for DNA barcoding studies. In previously study, they studied the organization of chloroplast *psbA-trnH* spacer in angiosperms of the Family *Umbelliferae*. The *psbA-trnH* gene was used to clarify the origin and the phylogenetic position of the species belonging to the genus *Ilex* (Aquifoliaceae). However, *psbA-trnH* gene showed slow evolve within *Ilex*, but a major gap present in this region was useful for the phylogenetic study of the genus. The *psbA-trnH* gene was used to identify three kinds of *Plumeria* flowers, which were successfully amplified and sequenced (Zhao et al., 2018). Similarity with *Araliaceae* plants, *psbA-trnH* suggested as another candidate barcode with correct discrimination of 100% and 79.39% at genus and species levels, respectively (Z. Liu et al., 2012). For the poison plants, *psbA-trnH* region could distinguish *lllicium verum* (llliciaceae) from other congeneric toxic adulterants (*I. lanceolatum*, *Illicium henryi*, *Illicium micranthum* and *Illicium simonsii*) compared with other regions such as *matK*, *rbcL*, and ITS2 (M. Liu et al., 2012). In *Apocynaceae*, *psbA-trnH* barcode could distinguish 95% and 100% of the samples at the species and genus levels (Lv et al., 2020).

2.5.3 *rbcL*

The ribulose-bisphosphate carboxylase large subunit gene (*rbcL*) is a coding region and located in the large single copy region of the chloroplast genome, which its has a full length of approximately 1,400 bp. This gene encodes for a large subunit of ribulose 1,5 bisphosphate carboxylase/oxygenase (RUBISCO) enzyme (Gielly & Taberlet, 1994). Rubisco enzyme consists of two kinds of protein subunits including the large chain and the small chain. The large chain gene (*rbcL*) has been used to the analysis of phylogenetic trees in plant taxonomy. This enzyme is responsible for carbon dioxide fixation during photosynthesis in the Calvin cycle (Jusuf et al., 2019). The *rbcL* gene has relatively easy amplification and comparability, which has been widely used

for species identification. The *rbcL* gene was used to confirm the Chinese 'cooling' beverage (Liangcha) from its adulterants or substitutes (Li et al., 2012). Li et al. (2012) reported that the *rbcL* gene was the most easily amplified (100%), sequenced and analyzed (97.7%). For tea plants, *rbcL* gene was successfully amplified and recovered in commercial tea product identification (Stoeckle et al., 2011). The *rbcL* gene was also proposed as a barcode for sequencing success, quality, discrimination power, and phylogenetic tree in the genus *Chlorophytum* (Patil et al., 2016). Another report also suggest that *rbcL* gene successfully authenticated and recognised 20 plant species in Mediterranean oolitic sand dunes west of Alexandria, Egypt, which was the first time of authentication of these plants (Fouad et al., 2021). Even though, *rbcL* gene has high universality, but low discrimination power (Jusuf et al., 2019). Therefore, CBOL plant working group suggested using a combination of *rbcL* and *matK* to increase the ability of discrimination power.

2.5.4 ITS1 and ITS2

Nuclear ribosomal internal transcribed spacer region (ITS/ITS2) is the nuclear genome and located in the ribosomal RNA operon, which has a length of around 450 to 750 bp (Beeck et al., 2014). The ITS and ITS2 have been proposed as a supplement barcode with *psbA-trnH* for various land plant groups. Moreover, these genes have also been used as DNA barcodes for the identification of fungal kingdom and the other biological groups such as algae, protists and animals (Wang et al., 2015). The ITS gene has showed highly polymorphic non-coding region, which will be able to discriminate sequences into species level in the fungal kingdom (Fajarningsih, 2016). In plants, ITS gene is widely used because of its high resolution of interspecific and intraspecific relationships. However, ITS gene also has limitations such as fungal contamination with DNA samples, incomplete concerted evolution, and difficulties in amplification and sequencing (Hollingsworth et al., 2011). Parveen et al. (2017) reported that ITS can discriminate more than 94.9% in Indian orchids, while the chloroplast, *matK* can provide a discrimination rate of only 85.7%, highest among the chloroplast loci, *rbcL*, *rpoB* and *rpoC1*. The ITS2 gene was used to verify black tea and green tea. The results showed that ITS2 gene was not suitable as DNA barcode for tea because it showed potential double peaks and noise in the raw sequence data, which has been discarded

from the experiments (De Castro et al., 2018). Similarity report, ITS2 gene was also used to identify varieties and origins of Taiwan’s domestic and imported made teas. The ITS2 can identify varieties of tea but it cannot identify the original by using phylogenetic tree. However, SNP results could be used to trace the origins of tea samples (Lee et al., 2016).

CHAPTER 3

METHODOLOGY

3.1 Plant Materials

Fresh young leaves of 135 wild tea plants were collected from the Singha Park Tea germplasm, Chiang Rai, Thailand. Wild tea plants were originally grown from 5 provinces including Chiang Mai (75 tea plants), Chiang Rai (26 tea plants), Nan (14 tea plants), Lampang (11 tea plants), Mae Hong Son (5 tea plants), and Phrea (4 tea plants) as show in Table 3.1.

Source Amazing Thailand (2022)

Figure 3.1 Map of Five Provinces that Originally Grown Wild Tea Plants

Table 3.1 List of 135 Wild Tea Plants Used in this Study

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCR1	Chiang Khian	Chiang Rai	1/3	N : 19.3652 E : 99.59532	398
TCR 2	Chiang Khian	Chiang Rai	1/4	N : 19.3652 E : 99.59532	398
TCR 6	Chiang Khian	Chiang Rai	2/6	N : 19.3656 E : 99.59534	402
TCR 8	Chiang Khian	Chiang Rai	3/4	N : 19.3652 E : 100.00236	425
TCR 13	Chiang Khian	Chiang Rai	5/1	N : 19.36517 E : 100.00225	414
TCR 18	Pasang	Chiang Rai	1/3	N : 10.11007 E : 99.46194	473
TCR 20	Pasang	Chiang Rai	2/2	N : 20.11008 E : 99.46196	472
TCR 28	Pasang	Chiang Rai	6/1	N : 20.10587 E : 99.46266	469
TCR 30	Pasang	Chiang Rai	7/7	N : 20.10587 E : 99.46266	469
TCR 33	Maekorn	Chiang Rai	2/2	N : 19.50299 E : 99.39577	481
TCR 39	Maekorn	Chiang Rai	6/7	N : 19.51419 E : 99.41131	497
TCR 41	Maekorn	Chiang Rai	7/4	N : 19.51409 E : 99.40115	495
TCR 42	Mae Chedi	Chiang Rai	5/2	N : 19.0754 E : 99.194	1,110
TCR 44	Mae Chedi	Chiang Rai	7/1	N : 19.08474 E : 99.20516	1,105
TCR 46	Mae Chedi	Chiang Rai	7/3	N : 19.08474 E : 99.20516	1,105
TCR 47	Mae Fah Luang	Chiang Rai	1/1	N : 20.165067 E : 99.44272	660
TCR 50	Mae Fah Luang	Chiang Rai	3/4	N : 20.16077 E : 99.44282	649
TCR 56	Mae Fah Luang	Chiang Rai	5/6	N : 20.16076 E : 99.44283	639

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCR 57	Mae Fah Luang	Chiang Rai	6/1	N : 20.16002 E : 99.44307	670
TCR 63	Wa wi	Chiang Rai	5/1	N : 19.52508 E : 99.34497	1,780
TCR 64.1	Wa wi	Chiang Rai	5/5	N : 19.52508 E : 99.34497	1,780
TCR 69	Wa wi	Chiang Rai	7/6	N : 19.52519 E : 99.3453	1,696
TCR 70	Wa wi	Chiang Rai	7/7	N : 19.52519 E : 99.3453	1,696
TCR 74	Huai Chomphu	Chiang Rai	1/3	N : 19.52574 E : 99.34571	1,667
TCR 84	Huai Chomphu	Chiang Rai	6/4	N : 19.54034 E : 99.36521	1,286
TCR 88	Huai Chomphu	Chiang Rai	7/4	N : 19.54187 E : 99.3801	1,183
TCM 90	Chang Khian	Chiang Mai	1/5	N : 18.8377 E : 98.89634	1,330
TCM 98	Sanpakia	Chiang Mai	1/2	N : 19.31414 E : 98.83266	1,572
TCM 99	Sanpakia	Chiang Mai	1/3	N : 19.31414 E : 98.83266	1,572
TCM 100	Sanpakia	Chiang Mai	1/5	N : 19.31414 E : 98.83266	1,572
TCM 102	Sanpakia	Chiang Mai	2/5	N : 19.31414 E : 98.83266	1,572
TCM 104	Sanpakia	Chiang Mai	3/2	N : 18.8377 E : 98.89634	1,572
TCM106	Sanpakia	Chiang Mai	3/4	N : 18.8377 E : 98.89634	1,572
TCM 107	Sanpakia	Chiang Mai	4/1	N : 18.8377 E : 98.89634	1,572
TCM108	Sanpakia	Chiang Mai	5/3	N : 18.8377 E : 98.89634	1,572
TCM 110	Sanpakia	Chiang Mai	6/1	N : 18.8377 E : 98.89634	1,572

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCM 113	Pang Hang	Chiang Mai	2/2	N : 19.33602 E : 98.88393	856
TCM 117	Pang Hang	Chiang Mai	5/3	N : 19.3371 E : 98.8845	900
TCM 118	Pang Hang	Chiang Mai	6/6	N : 19.3371 E : 98.8845	900
TCM 123	Pang Hong	Chiang Mai	2/6	N : 19.34513 E : 98.86799	1,084
TCM 125	Pang Hong	Chiang Mai	3/6	N : 19.34513 E : 98.86799	1,084
TCM 126	Pang Hong	Chiang Mai	6/6	N : 19.34513 E : 98.86799	1,084
TCM 128.1	Pang Hong	Chiang Mai	7/7	N : 19.34513 E : 98.86799	1,084
TCM 129	Mae Rim	Chiang Mai	1/1	N : 18.87289 E : 98.825	780
TCM 133	Mae Rim	Chiang Mai	3/4	N : 18.87051 E : 98.82574	794
TCM 139	Mae Rim	Chiang Mai	6/5	N : 18.92202 E : 98.81563	1,047
TCM 141	Muser Pak Tang	Chiang Mai	1/2	N : 17.57394 E : 98.52184	1,330
TCM 142	Muser Pak Tang	Chiang Mai	1/5	N : 17.57394 E : 98.52184	1,330
TCM 143	Muser Pak Tang	Chiang Mai	1/7	N : 17.57394 E : 98.52184	1,330
TCM 148	Muser Pak Tang	Chiang Mai	4/3	N : 17.56856 E : 98.52464	1,288
TCM 150	Muser Pak Tang	Chiang Mai	5/3	N : 17.56856 E : 98.52464	1,288
TCM 154	Mae Maea	Chiang Mai	2/5	N : 19.32041 E : 98.90014	727
TCM 156	Mae Maea	Chiang Mai	3/4	N : 19.32041 E : 98.90014	727
TCM 157	Mae Maea	Chiang Mai	4/1	N : 19.31965 E : 98.89422	756

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCM 158	Mae Maea	Chiang Mai	4/3	N : 19.31965 E : 98.89422	756
TCM 160	Mae Maea	Chiang Mai	5/1	N : 19.31965 E : 98.89422	756
TCM 162	Mae Maea	Chiang Mai	5/3	N : 19.31965 E : 98.89422	756
TCM 163	Mae Maea	Chiang Mai	7/2	N : 19.31975 E : 98.8903	733
TCM 165	Mae Maea	Chiang Mai	7/5	N : 19.31975 E : 98.8903	733
TCM 166	Ra Ming	Chiang Mai	1/4	N : 19.27247 E : 98.92135	918
TCM 167	Ra Ming	Chiang Mai	2/5	N : 19.27247 E : 98.92135	918
TCM 170	Ra Ming	Chiang Mai	3/1	N : 19.27247 E : 98.92135	918
TCM 172	Ra Ming	Chiang Mai	5/6	N : 19.27247 E : 98.92135	918
TCM 175	Ra Ming	Chiang Mai	7/4	N : 19.27247 E : 98.92135	918
TCM 176	Huai Tad	Chiang Mai	1/3	N : 19.15908 E : 98.90065	389
TCM 177	Huai Tad	Chiang Mai	2/1	N : 19.15908 E : 98.90065	389
TCM178	Huai Tad	Chaing Mai	2/2	N: 19.15908 E: 98.90065	389
TCM 180	Huai Tad	Chiang Mai	3/2	N : 19.15908 E : 98.90065	389
TCM 182	Huai Tad	Chiang Mai	7/7	N : 19.15908 E : 98.90065	389
TCM 183	Pha Deng	Chiang Mai	1/5	N : 19.11708 E : 98.72549	900
TCM 184	Pha Deng	Chiang Mai	2/3	N : 19.11708 E : 98.72549	900
TCM 185	Pha Deng	Chiang Mai	3/2	N : 19.11708 E : 98.72549	900

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCM 186	Pha Deng	Chiang Mai	6/3	N : 19.11708 E : 98.72549	900
TCM 189	Pang Manao	Chiang Mai	1/4	N : 19.12659 E : 98.70964	847
TCM 190	Pang Manao	Chiang Mai	2/7	N : 19.12659 E : 98.70964	847
TCM 193	Pang Manao	Chiang Mai	3/6	N : 19.12659 E : 98.70964	847
TCM 194	Pang Manao	Chiang Mai	3/7	N : 19.12659 E : 98.70964	847
TCM 196	Pang Manao	Chiang Mai	4/4	N : 19.12659 E : 98.70964	847
TCM 202	Pang Manao	Chiang Mai	6/7	N : 19.12659 E : 98.70964	847
TCM 203	Pang Manao	Chiang Mai	7/1	N : 19.12659 E : 98.70964	847
TCM 204	Tha Pha	Chiang Mai	1/1	N : 19.12638 E : 98.75706	662
TCM 205	Tha Pha	Chiang Mai	1/4	N : 19.12638 E : 98.75706	662
TCM 207	Tha Pha	Chiang Mai	2/4	N : 19.12638 E : 98.75706	662
TCM 208	Tha Pha	Chiang Mai	3/1	N : 19.12638 E : 98.75706	662
TCM 212	Tha Pha	Chiang Mai	5/5	N : 19.12092 E : 98.70709	832
TCM 215	Tha Pha	Chiang Mai	7/5	N : 19.12092 E : 98.70709	832
TCM 216	Kiw Tam	Chiang Mai	1/1	N : 19.00139 E : 99.33125	1,157
TCM 217	Kiw Tam	Chiang Mai	1/2	N : 19.00139 E : 99.33125	1,157
TCM 219	Kiw Tam	Chiang Mai	2/2	N : 19.00139 E : 99.33125	1,157
TCM 222	Pang Pong	Chiang Mai	1/5	N : 18.97011 E : 99.32008	828

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TCM 225	Pang Pong	Chiang Mai	7/7	N : 18.97011 E : 99.32008	828
TCM 227	Kam Phang Hin	Chiang Mai	1/7	N : 18.9489 E : 99.35128	1,083
TCM 228	Kam Phang Hin	Chiang Mai	2/1	N : 18.9489 E : 99.35128	1,083
TCM 229	Kam Phang Hin	Chiang Mai	2/2	N : 18.9489 E : 99.35128	1,083
TCM 231	Kam Phang Hin	Chiang Mai	4/7	N : 18.9489 E : 99.35128	1,083
TCM 235	Mae Ton	Chiang Mai	5/1	N : 18.93543 E : 99.37676	1,436
TCM 236	Mae Ton	Chiang Mai	6/2	N : 18.93543 E : 99.37676	1,436
TCM 237	Mae Ton	Chiang Mai	7/7	N : 18.93543 E : 99.37676	1,436
TCM 238	Pang Hi	Chiang Mai	3/6	N : 18.92282 E : 99.33241	950
TCM 239	Pang Hi	Chiang Mai	4/3	N : 18.92282 E : 99.33241	950
TCM 241	Ban Dong	Chiang Mai	5/6	N : 18.92399 E : 99.35216	1,157
TN 242	Reuang	Nan	1/3	N : 18.47007 E : 100.39187	282
TN247	Reuang	Nan	5/6	N : 18.47345 E : 100.3751	373
TN 249	Reuang	Nan	7/3	N : 18.47307 E : 100.3827	384
TN 250	Reuang	Nan	7/6	N : 18.47307 E : 100.3827	384
TN 251	Reuang	Nan	7/7	N : 18.47307 E : 100.3827	384
TN 252	Sakad A	Nan	1/3	N : 19.16021 E : 101.00441	1,033
TN 253	Sakad A	Nan	2/1	N : 19.16003 E : 101.00459	1,059

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TN 254	Sakad A	Nan	3/1	N : 19.15588 E : 101.00374	1,053
TN 256	Sakad A	Nan	3/5	N : 19.15588 E : 101.00374	1,053
TN 262	Sakad A	Nan	6/1	N : 19.15497 E : 101.00009	958
TN 268	Sakad B	Nan	1/1	N : 19.16036 E : 101.01097	1,148
TN 273	Sakad B	Nan	4/2	N : 19.16049 E : 101.0104	1,141
TN 278	Sakad B	Nan	7/1	N : 19.16092 E : 101.01003	1,083
TN 280	Sakad B	Nan	7/5	N : 19.16092 E : 101.01003	1,083
PPD 281	Pa Daeng	Prae	1/3	N : 18.02288 E : 101.19042	391
PPD 282	Pa Daeng	Prae	1/5	N : 18.02288 E : 101.19042	391
PPD 285	Pa Daeng	Prae	3/3	N : 18.02287 E : 100.19028	440
PPD 290	Pa Daeng	Prae	4/5	N : 18.02282 E : 101.19027	433
TMS 295	Mai Hung	Mae Hong Son	1/1	N : 19.62125 E : 98.20842	1,105
TMS 296	Mai Hung	Mae Hong Son	1/2	N : 19.62125 E : 98.20842	1,105
TMS 297	Mai Hung	Mae Hong Son	3/1	N : 19.62125 E : 98.20842	1,105
TMS 302	Mai Hung	Mae Hong Son	5/3	N : 19.62125 E : 98.20842	1,105
TMS 306	Mae La Na	Mae Hong Son	1/1	N : 19.59016 E : 98.21456	777
TLP 309	Chae Son	Lampang	1/7	N : 18.49574 E : 99.2315	998

Table 3.1 (continued)

Code	Origin			Latitude and longitude	Altitude (m)
	Locality	Province	variety code		
TLP 310	Chae Son	Lampang	2/7	N : 18.5006 E : 99.23322	1,039
TLP 311	Chae Son	Lampang	3/7	N : 18.50078 E : 99.2335	1,099
TLP 312	Chae Son	Lampang	4/7	N : 18.50067 E : 99.23381	1,081
TLP 313	Ban Rong	Lampang	1	N : 19.02328 E : 99.4693	833
TLP 314	Ban Rong	Lampang	2	N : 19.02306 E : 99.46481	870
TLP 315	Ban Rong	Lampang	3	N : 19.0229 E : 99.46486	866
TLP 316	Ban Rong	Lampang	4	N : 19.02283 E : 99.46484	865
TLP 317	Ban Rong	Lampang	5/7	N : 19.02269 E : 99.46473	861
TLP 318	Ban Rong	Lampang	7/1	N : 19.02582 E : 99.46272	843
TLP 319	Ban Rong	Lampang	7/2	N : 19.02582 E : 99.46272	843

3.2 DNA Extraction

Fresh young shoots or leaves of each tea plant were used to isolate DNA using a DNAsecure plant kit (Tiagen Biotech, China) according to the manufacturer's protocol. Genomic DNA was resuspended in 200 µl of TE buffer. The quality and quantity of DNA were examined using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, USA) and 0.8% agarose gel electrophoresis with 1x TAE buffer. Agarose gel was visualized under a UV light by Gel documentation (Biorad, USA). The extracted DNA was diluted to 50 ng/µl and stored at -20 °C.

3.3 PCR Amplification

PCR technique was performed to amplify four candidate DNA barcode regions (*rcbL, psbA-trnH, matK* and *rps7-trnV*). PCR amplification was carried out in a volume of 30 µl, containing 3 µl of DNA template (50 ng/µl), 3 µl dNTPs (2mM), 6 µl 5x Phusion HF buffer, 1.5 µl of each forward and reverse primer (10 mM), 0.3 µl of Phusion DNA polymerase (Thermo Fisher Scientific, USA) and the rest was filled with MQ water. The details of the primers shown in Table 3.2. The PCR reactions were performed using a Mastercycler X50 (Eppendorf, USA) according to the following cycling conditions: 94 °C for 5 min; 35 cycles: 94°C for 1min; 55 °C (*rbcL*); 48°C (*matK*); 55 °C (*psbA–trnH*); 56 °C (*rps7-trnV*) for 1 min; 72 °C for 1 min; 72 °C for 10 min. The amplified PCR products were visualized on 1.5% gel electrophoresis in 1X TAE buffer and 100 bp ladder was used to estimate the fragment sizes of PCR products. Agarose gel was visualized by Gel documentation (Biorad, USA).

Table 3.2 Details Of The Primer Used

Gene	Primer name	Primer sequence (5' - 3")	Expected PCR product size (bp)	Reference
psbA - trnH	fwd PA	F: GTT ATG CAT GAA CGT AAT GCTC	500	Newmaster and Ragupathy (2009)
	rev TH	R: CGC GCA TGG TGG ATT CAC AAT CC		
matK	390F	F: CGA TCT ATT CAT TCA ATA TTT C	900	Cuenoud et al .(2002)
	1326R	R: TCT AGC ACA CGA AAG TCG AAG T		
rbcL	1F	F: ATG TCA CCA CAA ACA GAA AC	800	Fay et al. (1997)
	724R	R: TCG CAT GTA CCT GCA GTA GC		
rps7-trnV	CS_rps7_1	F: GTG CTA TGG CTC GAA TCC GT	200	De Castro et al. (2017)
	CS_trnV(GAC)_1	R: ACC ACG TCA AGG TGA ACA TC		

3.4 Sequence Analysis and Phylogenetic Tree

The amplified PCR products of each wild tea plant were bidirectionally sequenced from 5' and 3' by the Solgent Company in South Korea. The forward and reverse of raw sequences were trimmed the low quality parts at both ends to increase quality of the entire sequences. The obtained sequences were assembled to create consensus sequences using Bioedit software (Hall, 1999). The assembled sequences were employed in the BLASTN software for species identification by checking the similarity. The multiple alignments were carried out in MEGA X using MUSCLE, Multiple Sequence Comparison by Log- Expectation. The alignment sequences were calculated the genetic distances between wild tea plants using the Kimura-2-parameter (K-2-P) model with transition-transversion. Phylogenetic trees of *psbA-trnH* and *matK* sequences were constructed using Neighbour-joining (NJ) with 1000 bootstraps in the MEGA X software. The sequences of the non-coding regions *psbA-trnH* and coding region *matK* were deposited in the GenBank database with the accession number provided in Table 3.1.

3.5 Morphological Analysis

The morphological characters of each wild plants were documented the format following the tea descriptions developed by Ellis et al. (2009) and Whiting (2017). Five characters of tea leaf, colour of young shoot, leaf apex, leaf base, leaf margin, and shape, were recorded. The five morphological characters are described in the Table 3.3 -3.7. All morphological characters of tea plants were used to construct the morphological tree using neigbor joining (NJ) method in the Past 4 software.

Table 3.3 Colour of Young Tip

Character of young tip	Description	Figure
Color	Green	
	Red	
	Purple	

Table 3.4 Five Types of Leaf Apex

Character of leaf apex	Description	Figure
Acuminate	The margin between the apex is concave, curving toward the centre of the leaf	
Acute	The leaf apex ends in a short sharp point.	
Emarginate	With the apex distinctly indented or one side of the leaf apex extends further than the other.	
Rounded	The margin forms a smooth arc across the apex	
Obtuse	The leaf apex is broadly rounder or blunt.	

Table 3.5 Four Types of Leaf Base

Character of leaf base	Description	Figure
Acuminate	The margin between the base has no significant curvature	
Acute	Base is pointed, showing angle less than a 90 degree angle.	
Rounded	The margin forms a smooth arc across the base	
Truncate	The base terminates abruptly as if cut	

Table 3.6 Seven Types of Leaf Margin

Character of leaf margin	Description	Figure
Serrate	Teeth pointed with their axes inclined to the trend of the leaf margin.	
Doubly serrate	Edges have a saw like teeth that have even smaller teeth within the larger ones.	

Table 3.6 (continued)

Character of leaf margin	Description	Figure
Barely serrate (entire)	Margin is smooth, without teeth, or almost.	
Serrulate	Edges have smaller teeth, more evenly spaced serrations.	

Table 3.6 (continued)

Character of leaf margin	Description	Figure
Crenate	Edges have a saw like teeth , but smoothly rounded.	
Undulate-serrate	Very wavy margins	

Table 3.6 (continued)

Character of leaf margin	Description	Figure
Sinuate-serrate	Margins are slightly wavy	

Table 3.7 Six types of leaf shape

Character of leaf shape	Description	Figure
Elliptical	Leaves widest in the middle, tapering on both ends	

Tabel 3.7 (continued)

Character of leaf shape	Description	Figure
Ovate	Leaf is broadest below the middle and about 2 times as long as the width.	
Lanceolate	Leaf is 3 times or more longer than width and broadest below the middle.	

Table 3.7 (continued)

Character of leaf shape	Description	Figure
Oblong	Widest part of the leaf is the middle part of the long axis where the opposite margins are roughly parallel.	
Obovate	Widest part of the leaf is on an axis in the apical 2/5 of the leaf.	

Table 3.7 (continued)

Character of leaf shape	Description	Figure
Oblanceolate	The middle is 3 times longer than wide and broadest above.	

CHAPTER 4

RESULTS

4.1 DNA Sequence Analysis

Four candidate DNA barcode regions, including *psbA-trnH*, *matK*, *rbcL*, and *rps7-trnV* were amplified in 135 wild tea plants. Only two candidate DNA barcode regions, *matK* and *psbA-trnH*, performed very well in all wild tea plants.

The approximately length of PCR products of *matK* and *psbA-trnH* regions were 900-1000 bp and 400-500 bp, respectively. The sequences of two DNA barcode regions were trimmed at both ends forward and reverse stands to increase the quality of the whole sequences and assembled together. After blasting and editing the sequences, the final sequences of the *matK* region ranged from 795 bp to 919 bp with an average of 888 bp, while *psbA-trnH* region ranged from 398 bp to 513 bp with an average of 493 bp. The longest sequence of the *matK* belongs to TCM185 (919 bp) while the shortest sequence belongs to TCR47 (795 bp). In *psbA-trnH* region, the longest sequence belongs to TCM178 (513 bp) and TCM182 (513 bp) while the shortest sequence belongs to TCR2 (398 bp). All the sequences of this study were deposited in the GenBank database with accession numbers provided in Table 4.1.

Table 4.1 Code and Details of the Submit *matK* and *psbA- trnH* Sequences Retrieved from NCBI

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TCR1	OL944024	906	OM617596	505
TCR2	OL944025	828	OM617597	398
TCR6	OL944026	825	OM617598	512
TCR8	OL944027	832	OM617599	506
TCR13	OM677530	840	OM617600	504
TCR18	OL944028	804	OM617601	505
TCR20	OL944029	807	OM617602	506
TCR28	OL944030	807	OM617603	509
TCR30	OL944031	803	OM617604	504
TCR33	OL944032	804	OM617605	505
TCR39	OL944034	806	OM617606	508
TCR41	OL944035	803	OM617607	491
TCR42	OL944036	796	OM617608	509
TCR44	OL944037	802	OM617609	509
TCR46	OL944038	806	OM617610	505
TCR47	OL944039	795	OM617611	467
TCR50	OL944040	805	OM617612	503
TCR56	OL944041	797	OM617613	476
TCR57	OL944042	808	OM617614	498
TCR63	OL944043	813	OM617615	512
TCR64.1	OL944044	802	OM617616	502
TCR69	OL944045	811	OM617617	489

Table 4.1 (continued)

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TCR70	OL944046	807	OM617618	509
TCR74	OL944047	800	OM617619	504
TCR84	OL944048	796	OM617620	496
TCR88	OL944049	885	OM617621	492
TCM90	OM677531	912	OM617622	493
TCM98	OL944050	905	OM617623	499
TCM99	OM677532	913	OM617624	497
TCM100	OM677533	913	OM617625	502
TCM102	OL944051	913	OM617626	444
TCM104	OM677534	873	OM617627	489
TCM106	OL944052	891	OM617628	498
TCM107	OL944053	903	OM617629	498
TCM108	OL944054	906	OM617630	494
TCM110	OL944055	912	OM617631	497
TCM113	OM677535	915	OM617632	498
TCM117	OL944056	914	OM617633	502
TCM118	OM677536	918	OM617634	492
TCM123	OM677537	914	OM617635	490
TCM125	OM677538	917	OM617636	488
TCM126	OM677539	906	OM617637	488
TCM128.1	OM677540	914	OM617638	489
TCM129	OM677541	918	OM617639	492
TCM133	OM677542	918	OM617640	489
TCM139	OL944057	906	OM617641	493
TCM141	OM677543	917	OM617642	500

Table 4.1 (continued)

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TCM142	OM677544	915	OM617643	498
TCM143	OM677545	917	OM617644	492
TCM148	OM677546	915	OM617645	492
TCM150	OM677547	913	OM617646	490
TCM154	OM677548	914	OM617647	487
TCM156	OM677549	846	OM617648	494
TCM157	OM677550	912	OM617649	491
TCM158	OM677551	913	OM617650	492
TCM160	OL944058	912	OM617651	492
TCM162	OL944059	912	OM617652	492
TCM163	OL944060	912	OM617653	489
TCM165	OL944061	912	OM617654	485
TCM166	OL944062	917	OM617655	482
TCM167	OM677552	857	OM617656	491
TCM170	OM677553	914	OM617657	492
TCM172	OM677554	912	OM201147	503
TCM175	OM677555	887	OM201148	499
TCM176	OL944063	802	OM201149	498
TCM177	OL944064	895	OM201150	500
TCM178	OM677556	911	OM201151	513
TCM180	OL944065	911	OM201152	498
TCM182	OM677557	910	OM201153	513
TCM183	OL944066	912	OM201154	507
TCM184	OL944067	917	OM201155	508
TCM185	OM677558	919	OM201156	495

Table 4.1 (continued)

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TCM186	OM677559	915	OM617658	491
TCM189	OM677560	914	OM617659	491
TCM190	OM677561	914	OM617660	492
TCM193	OL944068	912	OM617661	492
TCM194	OM677562	888	OM617662	483
TCM196	OM677563	914	OM617663	491
TCM202	OM677564	915	OM617664	491
TCM203	OM677565	917	OM617665	492
TCM204	OM677566	915	OM617666	491
TCM205	OM677567	911	OM617667	489
TCM207	OM677568	918	OM617668	493
TCM208	OM677569	918	OM617669	488
TCM212	OM677570	897	OM617670	491
TCM215	OM677571	851	OM617671	477
TCM216	OM677572	882	OM617672	493
TCM217	OL944069	912	OM617673	492
TCM219	OM677573	915	OM617674	489
TCM222	OM677574	916	OM617675	492
TCM225	OM677575	918	OM617676	491
TCM227	OM677576	915	OM617677	493
TCM228	OM677577	891	OM617678	491
TCM229	OM677578	918	OM617679	493
TCM231	OM677579	851	OM617680	489
TCM235	OM677580	915	OM617681	491

Table 4.1 (continued)

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TCM236	OM677581	915	OM617682	490
TCM237	OM677582	918	OM617683	488
TCM238	OM677583	871	OM617684	491
TCM239	OM677584	888	OM617685	492
TCM241	OM677585	877	OM617686	492
TN242	OM677586	915	OM617687	493
TN247	OM677587	870	OM617688	493
TN249	OM677588	918	OM617689	490
TN250	OM677589	917	OM617690	492
TN251	OM677590	915	OM617691	497
TN252	OM677591	867	OM617692	490
TN253	OM677592	874	OM617693	492
TN254	OL944070	913	OM617694	492
TN256	OM677593	882	OM617695	496
TN262	OM677594	888	OM617696	494
TN268	OL944071	916	OM617697	493
TN273	OL944072	917	OM617698	494
TN278	OM677595	870	OM617699	479
TN280	OM677596	888	OM617700	490
PPD281	OM677597	875	OM617701	489
PPD282	OM677598	915	OM617702	479
PPD285	OM677599	866	OM617703	490
PPD290	OL944073	913	OM617704	491
TMS295	OM677600	915	OM617705	491

Table 4.1 (continued)

Code	*matK*		*psbA-trnH*	
	NCBI accession	**Length (bp)**	**NCBI accession**	**Length (bp)**
TMS296	OM677601	874	OM617706	493
TMS297	OM677602	914	OM617707	486
TMS302	OM677603	914	OM617708	497
TMS306	OM677604	915	OM617709	487
TLP309	OM677605	855	OM617710	487
TLP310	OM677606	858	OM617711	493
TLP311	OM677607	863	OM617712	487
TLP312	OM677608	849	OM617713	489
TLP313	OM677609	906	OM617714	490
TLP314	OM677610	875	OM617715	490
TLP315	OM677611	845	OM617716	493
TLP316	OL944074	911	OM617717	488
TLP317	OM677612	847	OM617718	488
TLP318	OM677613	876	OM617719	492
TLP319	OM677614	877	OM617720	490

The nucleotide sequences of the *matK* and *psbA-trnH* regions were aligned using the MUSCLE algorithm in MEGA software. In *matK*, twelve single nucleotide polymorphisms (SNPs) were detected at the aligned positions 103 (C to T), 110 (T to G), 156 (C to T), 210 (T to G), 223 (G to A), 236 (A to T), 351 (T to C), 361 (G to A) ,635 (C to T), 676 (T to G), 697 (G to A), and 749 (G to T), which was found in 72 of 135 tea plants (53.3%). The SNPs position in each tea plant are showed in Table 4.2. The highest SNPs (6 SNPs) were detected at aligned positions 210, 223, 361, 676, 697, and 749 in eight tea plants, TCM175, TCM194, TCM215, TCM217, TCM239, TN254, TN262, and TN280. Five SNPs were detected in 57 tea plants, which divided into two

patterns, at aligned positions 103, 223, 676, 697, and 749 (10 tea plants) and 210, 223, 676, 697, and 749 (47 tea plants). Among the four SNPs, only TCR2 was detected SNPs at aligned positions 236, 351, 635, and 697, which had three nucleotide difference from the other 71 tea plants. Thus, three nucleotides can be used to discriminate TCR2 from the rest of the tea plants. Moreover, one SNP was detected in six tea plants at aligned position 156 (C to T) in TCR42, 697 (G to A) in TCM170, TCM172, TCM193, TCM196, and TCM238, and 110 (T to G) in TCM156. The highest SNP frequency of the *matK* sequence was detected in TCM215 (0.71%), followed by TCM175, TCM194, TCM239, TN262, and TN280 (0.68%). The *matK* sequence showed five unique variable positions in three tea plants, including TCR2, TCR42, and TCM156 that differed from the rest of the sequences. Insertion/deletion was not found in *matK* sequences. For the genetic variation, the genetic distances between all *matK* sequences varied from 0.0000 to 0.0103, with an average of 0.0036. The highest variation in genetic distance occurred between TCR2 and TCM175, TCM194, TCM215, TCM217, TCM239, TN254, TN262, and TN280. As a result showed that tea plants with a high genetic distance have a high number of SNP.

For nucleotide sequences of the *psbA-trnH*, a total of seven SNPs were detected at aligned positions 38 (C to T), 66 (C to T), 305 (A to C), 318 (C to T), 330 (G to T), 463 (A to G) and 468 (T to G) in 72 of 135 tea plants (52.9%) the same as *matK*. Most tea plants found only one or two SNPs at aligned position 318 or 330 or both positions (Table 4.3). Forty-two aligned sequences were found one SNP at aligned positions 38 (1 tea plant), 66 (1 tea plant), 318 (34 tea plants), 330 (5 tea plants), 463 (1 tea plant), and 468 (1 tea plant). Two SNPs were detected at aligned positions 318 and 330 in 29 tea plants. Four SNPs at aligned position 66, 305, 330, and 463 were detected in 72 tea plants but only TCR2 had three nucleotide different from the other. The SNPs result of TCR2 in *matK* and *psbA-trnH* showed the same results meaning TCR2 differently from the other tea plants. The highest SNP frequency of *psbA-trnH* was detected in only TCR2 (1.01%). Furthermore, the *psbA-trnH* sequence showed five unique variable positions in three tea plants in TCR2, TCM185, and TCM219 that differed from the rest of the tea plant. Insertions/deletions (Indels) were detected in 19 of the 135 sequences (14.07%). Five nucleotide deletions (ATAAA) were found at aligned positions 354-358 in TCR1, TCR46, TCR74, TCM90, TCM99 and TLP311.

Nucleotide insertions were found at aligned position 170 with the base “T” in TCM229, TN273, and TN278, at 241-247 with base “TTATGGG” in TCM98, TCM102, TCM106, TCM107, TCM117, TCM141, TCM142, TCM178, TCM182, and TMS302 and at 449 - 452 with base “ATTT” in TCR1, TCR46, TCR74, TCM90, and TLP311. For genetic variation, the genetic distance between all *psbA-trnH* sequences ranged from 0.0000 to 0.0131, with an average of 0.0023. The highest genetic distance value was found in TCR2 and TCM194. From *psbA-trnH* result, this region exhibits high sequence divergence and rate of insertion/deletion, when compared with *matK* region. Based on SNP results, *matK* region shows more variation in SNPs than *psbA-trnH* region. TCR2 showed the highest genetic distance from the rest of the tea plants in both *matK* and *psbA- trnH* regions, which matches with the SNP results. However, the genetic variation among the tea plants with different origins was not detected. All the sequences of tea plant were identified and compared with the sequence references of *C. sinensis* from NCBI for ensuring bases and position of SNP. The reference sequences showed nucleotide base similarly with the sequence that had no SNPs. So, we can assume the sequence that showed the different base from other sequences was a SNP. As a result, the *matK* region is a good candidate for DNA barcoding of wild tea plants.

Table 4.2 Single Nucleotide Polymorphisms of the *matK* Sequences in Wild Tea Plants

Code	*matK*												Total bp	Total SNP	% SNP fre.
	103	110	156	210	223	236	351	361	635	676	697	749			
TCR2	C	T	C	T	G	T	C	G	T	T	A	G	828	4	0.48
TCR6	C	T	C	G	A	A	T	G	C	G	A	T	825	5	0.61
TCR8	C	T	C	G	A	A	T	G	C	G	A	T	832	5	0.60
TCR13	C	T	C	G	A	A	T	G	C	G	A	T	840	5	0.60
TCR18	C	T	C	G	A	A	T	G	C	G	A	T	804	5	0.62
TCR28	C	T	C	G	A	A	T	G	C	G	A	T	807	5	0.62
TCR30	C	T	C	G	A	A	T	G	C	G	A	T	803	5	0.62
TCR33	C	T	C	G	A	A	T	G	C	G	A	T	804	5	0.62
TCR39	C	T	C	G	A	A	T	G	C	G	A	T	806	5	0.62
TCR41	C	T	C	G	A	A	T	G	C	G	A	T	803	5	0.62
TCR42	C	T	T	T	G	A	T	G	C	T	G	G	796	1	0.13
TCR44	C	T	C	G	A	A	T	G	C	G	A	T	802	5	0.62
TCR47	C	T	C	G	A	A	T	G	C	G	A	T	795	5	0.63
TCR56	C	T	C	G	A	A	T	G	C	G	A	T	797	5	0.63
TCR57	C	T	C	G	A	A	T	G	C	G	A	T	808	5	0.62
TCR63	C	T	C	G	A	A	T	G	C	G	A	T	813	5	0.62
TCR64.1	C	T	C	G	A	A	T	G	C	G	A	T	802	5	0.62
TCR69	C	T	C	G	A	A	T	G	C	G	A	T	811	5	0.62
TCR70	C	T	C	G	A	A	T	G	C	G	A	T	807	5	0.62
TCR84	C	T	C	G	A	A	T	G	C	G	A	T	796	5	0.63
TCM98	T	T	C	T	A	A	T	G	C	G	A	T	905	5	0.55
TCM102	T	T	C	T	A	A	T	G	C	G	A	T	913	5	0.55
TCM104	C	T	C	G	A	A	T	G	C	G	A	T	873	5	0.57

Table 4.2 (continued)

Code	*matK*												Total bp	Total SNP	% SNP fre.
	103	110	156	210	223	236	351	361	635	676	697	749			
TCM106	T	T	C	T	A	A	T	G	C	G	A	T	891	5	0.56
TCM107	T	T	C	T	A	A	T	G	C	G	A	T	903	5	0.55
TCM108	C	T	C	G	A	A	T	G	C	G	A	T	906	5	0.55
TCM110	C	T	C	G	A	A	T	G	C	G	A	T	912	5	0.55
TCM117	T	T	C	T	A	A	T	G	C	G	A	T	914	5	0.55
TCM141	T	T	C	T	A	A	T	G	C	G	A	T	917	5	0.55
TCM142	T	T	C	T	A	A	T	G	C	G	A	T	915	5	0.55
TCM150	C	T	C	G	A	A	T	G	C	G	A	T	913	5	0.55
TCM156	C	G	C	T	G	A	T	G	C	T	G	G	846	1	0.12
TCM162	C	T	C	G	A	A	T	G	C	G	A	T	912	5	0.55
TCM163	C	T	C	G	A	A	T	G	C	G	A	T	912	5	0.55
TCM166	C	T	C	G	A	A	T	G	C	G	A	T	917	5	0.55
TCM167	C	T	C	G	A	A	T	G	C	G	A	T	857	5	0.58
TCM170	C	T	C	T	G	A	T	G	C	T	A	G	914	1	0.11
TCM172	C	T	C	T	G	A	T	G	C	T	A	G	912	1	0.11
TCM175	C	T	C	G	A	A	T	A	C	G	A	T	887	6	0.68
TCM178	T	T	C	T	A	A	T	G	C	G	A	T	911	5	0.55
TCM182	T	T	C	T	A	A	T	G	C	G	A	T	910	5	0.55
TCM183	C	T	C	G	A	A	T	G	C	G	A	T	912	5	0.55
TCM189	C	T	C	G	A	A	T	G	C	G	A	T	914	5	0.55
TCM193	C	T	C	T	G	A	T	G	C	T	A	G	912	1	0.11
TCM194	C	T	C	G	A	A	T	A	C	G	A	T	888	6	0.68
TCM196	C	T	C	T	G	A	T	G	C	T	A	G	914	1	0.11
TCM203	C	T	C	G	A	A	T	G	C	G	A	T	917	5	0.55

Table 4.2 (continued)

Code	*matK*												Total bp	Total SNP	% SNP fre.
	103	110	156	210	223	236	351	361	635	676	697	749			
TCM215	C	T	C	G	A	A	T	A	C	G	A	T	851	6	0.71
TCM217	C	T	C	G	A	A	T	A	C	G	A	T	912	6	0.66
TCM227	C	T	C	G	A	A	T	G	C	G	A	T	915	5	0.55
TCM236	C	T	C	G	A	A	T	G	C	G	A	T	915	5	0.55
TCM237	C	T	C	G	A	A	T	G	C	G	A	T	918	5	0.54
TCM238	C	T	C	T	G	A	T	G	C	T	A	G	871	1	0.11
TCM239	C	T	C	G	A	A	T	A	C	G	A	T	888	6	0.68
TCM241	C	T	C	G	A	A	T	G	C	G	A	T	877	5	0.57
TN252	C	T	C	G	A	A	T	G	C	G	A	T	867	5	0.58
TN253	C	T	C	G	A	A	T	G	C	G	A	T	874	5	0.57
TN254	C	T	C	G	A	A	T	A	C	G	A	T	913	6	0.66
TN262	C	T	C	G	A	A	T	A	C	G	A	T	888	6	0.68
TN280	C	T	C	G	A	A	T	A	C	G	A	T	888	6	0.68
PPD281	C	T	C	G	A	A	T	G	C	G	A	T	875	5	0.57
TMS296	C	T	C	G	A	A	T	G	C	G	A	T	874	5	0.57
TMS302	T	T	C	T	A	A	T	G	C	G	A	T	914	5	0.55
TMS306	C	T	C	G	A	A	T	G	C	G	A	T	915	5	0.55
TLP309	C	T	C	G	A	A	T	G	C	G	A	T	855	5	0.58
TLP312	C	T	C	G	A	A	T	G	C	G	A	T	849	5	0.59
TLP314	C	T	C	G	A	A	T	G	C	G	A	T	875	5	0.57
TLP315	C	T	C	G	A	A	T	G	C	G	A	T	845	5	0.59
TLP316	C	T	C	G	A	A	T	G	C	G	A	T	911	5	0.55
TLP317	C	T	C	G	A	A	T	G	C	G	A	T	847	5	0.59
TLP318	C	T	C	G	A	A	T	G	C	G	A	T	876	5	0.57

Table 4.2 (continued)

Code	*matK*												Total bp	Total SNP	% SNP fre.
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**			
TLP319	C	T	C	G	A	A	T	G	C	G	A	T	877	5	0.57

Table 4.3 Single Nucleotide Polymorphisms of the *psbA-trnH* Sequences in Wild Tea Plants

Code	*psbA - trnH*							Total bp	Total SNP	% SNP fre.
	38	**66**	**305**	**318**	**330**	**463**	**468**			
TCR2	C	T	C	C	T	G	T	398	4	1.01
TCR6	C	C	A	T	T	A	T	512	2	0.39
TCR8	C	C	A	T	T	A	T	506	2	0.40
TCR13	C	C	A	T	T	A	T	504	2	0.40
TCR18	C	C	A	T	G	A	T	505	1	0.20
TCR28	C	C	A	T	T	A	T	509	2	0.39
TCR30	C	C	A	T	T	A	T	504	2	0.40
TCR33	C	C	A	T	T	A	T	505	2	0.40
TCR39	C	C	A	T	T	A	T	508	2	0.39
TCR41	C	C	A	T	T	A	T	491	2	0.41
TCR44	C	C	A	T	T	A	T	509	2	0.39
TCR47	C	C	A	T	T	A	T	467	2	0.43
TCR56	C	C	A	T	T	A	T	476	2	0.42
TCR57	C	C	A	T	T	A	T	498	2	0.40
TCR63	C	C	A	T	T	A	T	512	2	0.39
TCR64.1	C	C	A	T	T	A	T	502	2	0.40
TCR69	C	C	A	T	T	A	T	489	2	0.41

Table 4.3 (continued)

Code	*psbA - trnH*							Total bp	Total SNP	% SNP fre.
	38	66	305	318	330	463	468			
TCR70	C	C	A	T	T	A	T	509	2	0.39
TCR84	C	C	A	T	T	A	T	496	2	0.40
TCM98	C	C	A	T	G	A	T	499	1	0.20
TCM102	C	C	A	T	G	A	T	444	1	0.23
TCM104	C	C	A	T	G	A	T	489	1	0.20
TCM106	C	C	A	T	G	A	T	498	1	0.20
TCM107	C	C	A	T	G	A	T	498	1	0.20
TCM108	C	C	A	T	G	A	T	494	1	0.20
TCM110	C	C	A	T	G	A	T	497	1	0.20
TCM117	C	C	A	T	G	A	T	502	1	0.20
TCM141	C	C	A	T	G	A	T	500	1	0.20
TCM142	C	C	A	T	G	A	T	498	1	0.20
TCM150	C	C	A	T	G	A	T	490	1	0.20
TCM162	C	C	A	T	G	A	T	492	1	0.39
TCM163	C	C	A	T	G	A	T	489	1	0.43
TCM166	C	C	A	T	G	A	T	482	1	0.42
TCM167	C	C	A	T	G	A	T	491	1	0.40
TCM170	C	C	A	C	T	A	T	492	1	0.39
TCM172	C	C	A	C	T	A	T	503	1	0.40
TCM175	C	C	A	T	G	A	T	499	1	0.41
TCM178	C	C	A	T	G	A	T	513	1	0.39
TCM182	C	C	A	T	G	A	T	513	1	0.40
TCM183	C	C	A	T	G	A	T	507	1	0.20
TCM185	C	C	A	C	G	A	G	495	1	0.23

Table 4.3 (continued)

Code	*psbA - trnH*							Total bp	Total SNP	% SNP fre.
	38	66	305	318	330	463	468			
TCM189	C	C	A	T	G	A	T	491	1	0.20
TCM193	C	C	A	C	T	A	T	492	1	0.20
TCM194	C	C	A	T	G	A	T	483	1	0.20
TCM196	C	C	A	C	T	A	T	491	1	0.20
TCM203	C	C	A	T	G	A	T	492	1	0.20
TCM215	C	C	A	T	G	A	T	477	1	0.20
TCM217	C	C	A	T	G	A	T	492	1	0.20
TCM219	T	C	A	C	G	A	T	489	1	0.20
TCM227	C	G	A	T	G	A	T	493	1	0.20
TCM236	C	G	A	T	G	A	T	490	1	0.39
TCM237	C	G	A	T	G	A	T	488	1	0.43
TCM238	C	G	A	C	T	A	T	491	1	0.42
TCM239	C	G	A	T	G	A	T	492	1	0.40
TCM241	C	G	A	T	G	A	T	492	1	0.20
TN252	C	G	A	T	T	A	T	490	2	0.41
TN253	C	G	A	T	T	A	T	492	2	0.41
TN254	C	G	A	T	G	A	T	492	1	0.20
TN262	C	G	A	T	G	A	T	494	1	0.20
TN280	C	G	A	T	G	A	T	490	1	0.20
PPD281	C	G	A	T	T	A	T	489	2	0.41
TMS296	C	G	A	T	T	A	T	493	2	0.41
TMS302	C	G	A	T	G	A	T	497	1	0.20
TMS306	C	G	A	T	G	A	T	487	1	0.21
TLP309	C	G	A	T	T	A	T	487	2	0.41

Table 4.3 (continued)

Code	*psbA - trnH*							Total bp	Total SNP	% SNP fre.
	38	66	305	318	330	463	468			
TLP312	C	G	A	T	T	A	T	489	2	0.41
TLP314	C	G	A	T	T	A	T	490	2	0.41
TLP315	C	G	A	T	T	A	T	493	2	0.41
TLP316	C	G	A	T	T	A	T	488	2	0.41
TLP317	C	G	A	T	T	A	T	488	2	0.41
TLP318	C	G	A	T	T	A	T	492	2	0.41
TLP319	C	G	A	T	T	A	T	490	2	0.41

4.2 Comparing Sequence Variation between *matK* and *psbA-trnH* Barcode Regions

From the SNP result, *matK* had 12 SNPs, seven patterns of SNPs, while *psbA-trnH* had 7 SNPs, six patterns of SNPs. SNP patterns of both *matK* and *psbA-trnH* regions were compared together. The result showed that TCR2 had the different SNP positions from the other tea plants in both *matK* and *psbA-trnH* regions. One group of tea plant, including TCM170, TCM172, TCM193, TCM196, and TCM238 showed the same SNP positions in both *matK* (at aligned position 697) and *psbA-trnH* (at aligned position 330). Of those tea plants were all originally grown in Chiang Mai province. Tea plants with five SNPs that showed in *matK* were detected one or two SNPs in *psbA-trnH*, which were at aligned position 318 or 318 and 330. A group of six tea plants with six SNPs at aligned position 210, 223, 361, 676, 697, and 749 in *matK* region was detected only one SNP in *psbA-trnH* region at aligned position 318 the same as SNP position in *matK*. Moreover, there are some tea plants were detected SNP positions in only one DNA barcode region (*matK* or *psbA-trnH*). For example, TCR42 and TCM156 were only found in *matK* region, while TCM185 and TCM219 were only found SNPs

in *psbA-trnH* region. As a SNPs result, tea plants that find the SNPs pattern in *matK* region may not find the same SNP pattern in *psbA-trnH* region (Table 4.4).

Table 4.4 Compared SNP Patterns between *matK* and *psbA-trnH* Regions

Code	*matK*													*psbA-trnH*					
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TCR1																			
TCR2						T	C		T		A			T	C		T	G	
TCR6				G	A					G	A	T				T	T		
TCR8				G	A					G	A	T				T	T		
TCR13				G	A					G	A	T				T	T		
TCR18				G	A					G	A	T				T			
TCR20																			
TCR28				G	A					G	A	T				T	T		
TCR30				G	A					G	A	T				T	T		
TCR33				G	A					G	A	T				T	T		
TCR39				G	A					G	A	T				T	T		
TCR41				G	A					G	A	T				T	T		
TCR42			T																
TCR44				G	A					G	A	T				T	T		
TCR46																			
TCR47				G	A					G	A	T				T	T		
TCR50																			
TCR56				G	A					G	A	T				T	T		
TCR57				G	A					G	A	T				T	T		
TCR63				G	A					G	A	T				T	T		
TCR64.1				G	A					G	A	T				T	T		

Table 4.4 (continued)

Code	*matK*												*psbA-trnH*						
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TCR69				G	A					G	A	T				T	T		
TCR70				G	A					G	A	T				T	T		
TCR74																			
TCR84				G	A					G	A	T				T	T		
TCR88																			
TCM90																			
TCM98	T				A					G	A	T				T			
TCM99																			
TCM100																			
TCM102	T				A					G	A	T				T			
TCM104				G	A					G	A	T				T			
TCM106	T				A					G	A	T				T			
TCM107	T				A					G	A	T				T			
TCM108				G	A					G	A	T				T			
TCM110				G	A					G	A	T				T			
TCM113																			
TCM117	T				A					G	A	T				T			
TCM118																			
TCM123																			
TCM125																			
TCM126																			
TCM128.1																			
TCM129																			

Table 4.4 (continued)

Code	*matK*												*psbA-trnH*						
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TCM133																			
TCM139																			
TCM141	T				A					G	A	T				T			
TCM142	T				A					G	A	T				T			
TCM143																			
TCM148																			
TCM150				G	A					G	A	T				T			
TCM154																			
TCM156		G																	
TCM157																			
TCM158																			
TCM160																			
TCM162				G	A					G	A	T				T			
TCM163				G	A					G	A	T				T			
TCM165																			
TCM166				G	A					G	A	T				T			
TCM167				G	A					G	A	T				T			
TCM170											A						T		
TCM172											A						T		
TCM175				G	A			A		G	A	T				T			
TCM176																			
TCM177																			
TCM178	T				A					G	A	T				T			

Table 4.4 (continued)

Code	*matK*												*psbA-trnH*						
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TCM180																			
TCM182	T				A					G	A	T				T			
TCM183				G	A					G	A	T				T			
TCM184																			
TCM185																			G
TCM186																			
TCM189				G	A					G	A	T				T			
TCM190																			
TCM193											A						T		
TCM194				G	A			A		G	A	T				T			
TCM196											A						T		
TCM202																			
TCM203				G	A					G	A	T				T			
TCM204																			
TCM205																			
TCM207																			
TCM208																			
TCM212																			
TCM215				G	A			A		G	A	T				T			
TCM216																			
TCM217				G	A			A		G	A	T				T			
TCM219													T						
TCM222																			

Table 4.4 (continued)

Code	*matK*												*psbA-trnH*						
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TCM225																			
TCM227				G	A					G	A	T				T			
TCM228																			
TCM229																			
TCM231																			
TCM235																			
TCM236				G	A					G	A	T				T			
TCM237				G	A					G	A	T				T			
TCM238											A						T		
TCM239				G	A			A		G	A	T				T			
TCM241				G	A					G	A	T				T			
TN242																			
TN247																			
TN249																			
TN250																			
TN251																			
TN252				G	A					G	A	T				T	T		
TN253				G	A					G	A	T				T	T		
TN254				G	A			A		G	A	T				T			
TN256																			
TN262				G	A			A		G	A	T				T			
TN268																			
TN273																			

Table 4.4 (continued)

Code	*matK*												*psbA-trnH*						
	103	**110**	**156**	**210**	**223**	**236**	**351**	**361**	**635**	**676**	**697**	**749**	**38**	**66**	**305**	**318**	**330**	**463**	**468**
TN278																			
TN280				G	A			A		G	A	T				T			
PPD281				G	A					G	A	T				T	T		
PPD282																			
PPD285																			
PPD290																			
TMS295																			
TMS296				G	A					G	A	T				T	T		
TMS297																			
TMS302	T				A					G	A	T				T			
TMS306				G	A					G	A	T				T			
TLP309				G	A					G	A	T				T	T		
TLP310																			
TLP311																			
TLP312				G	A					G	A	T				T	T		
TLP313																			
TLP314				G	A					G	A	T				T	T		
TLP315				G	A					G	A	T				T	T		
TLP316				G	A					G	A	T				T	T		
TLP317				G	A					G	A	T				T	T		
TLP318				G	A					G	A	T				T	T		
TLP319				G	A					G	A	T				T	T		

Note Empty Column Means Tea Plant with no SNPs

4.3 Blast Analysis

4.3.1 *matK* Sequence

Blast analysis showed 64 tea samples out of 135 to be 100% identical with *C. sinensis* (Table 4.5). The rest of the tea plants did not show 100% identical with *C. sinensis*. Only one tea plant (TCM156) showed the highest similarity with *C. sinensis*, 99.76%. Six tea plants were 100% identical to the other species of *Camellia* and followed by *C. sinensis*, including TCM170, TCM172, TCM193, TCM196, and TCM238 (99.89% with C. *sinensis*) and TCM110 (99.45% with C. *sinensis*). Most tea plants showed the percentage of similarity with C. *sinensis* or other species *Camellia* species lower than 100%. There were 12 tea plants showed a higher percentage of similarity with the other species of *Camellia* higher than C. *sinensis* such as *C. mingii*, *C. japonica*, *C. gymnogyna*, and *C. taliensis*, which were 99.64% (TCR2), 99.55% (TCM106), 99.89% (TCM98, TCM102, TCM107, TCM117, TCM141, TCM142, TCM178, TCM182, TN280, and TMS302). While the percentage similarity of all 12 tea plants with *C. sinensis* was 99.45%, except TCM106 was 99.10%. Additionally, 52 tea plants did not show or had a lower percentage of similarity with C. *sinensis* than 99.10%. However, all 52 tea plants showed a percentage similarity of *matK* sequences to several congeneric species such as *C. crassicolumna*, *C. taliensis*, *C. sasanqua*, or *C. japonica* ranged from 99.78%-100%. The lowest percentage of similarities found in TCM106, showed 99.10% with *C. sinensis*, and 99.55% *C. taliensis* as shown in Table 4.5-4.9.

Table 4.5 The *matK* Sequences Showed 100% Identical with *C. sinensis*

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR1	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCR20	100	MN138454.1	100	*C. ptilophylla*	MN138454.1
TCR42	100	MN138437.1	99.87	*C. ptilophylla*	NC_038198.1

Table 4.5 (continued)

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR46	100	MK393398.1	100	*C. ptilophylla*	NC_038198.1
TCR50	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCR74	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCR88	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM90	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM99	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM100	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM113	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM118	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM123	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM125	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM126	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM128.1	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM129	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM133	100	LC488797.1	100	*C. ptilophylla*	NC_038198.1
TCM139	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM143	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM148	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM154	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM157	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM158	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM160	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1
TCM165	100	LC488797.1	100	*C. grandibracteata*	KJ197936.1

Table 4.5 (continued)

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM176	100	MK393398.1	100	*C. grandibracteata*	KJ197936.1
TCM177	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM180	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM184	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM185	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM186	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM190	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM202	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM204	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM205	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM207	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM208	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM212	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM216	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM219	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM222	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM225	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM228	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM229	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TCM231	100	MG946957.1	100	*C. grandibracteata*	KJ197936.1
TCM235	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN242	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN247	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN249	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1

Table 4.5 (continued)

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TN250	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN251	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN256	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN268	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN273	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TN278	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
PPD282	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
PPD285	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
PPD290	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TMS295	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TMS297	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1
TLP310	100	MG946957.1	100	*C. grandibracteata*	KJ197936.1
TLP311	100	MG946957.1	100	*C. grandibracteata*	KJ197936.1
TLP313	100	AF380077.1	100	*C. grandibracteata*	KJ197936.1

Table 4.6 The *matK* Sequences Showed 100% Identical with the Other *Camellia* Species, but not Showed the Percentage of Similarity with *C. sinensis*

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR6	-	-	100	*C. crassicolumna*	KJ197926.1
TCR8	-	-	100	*C. crassicolumna*	KJ197926.1
TCR13	-	-	100	*C. crassicolumna*	KJ197926.1
TCR18	-	-	100	*C. crassicolumna*	KJ197926.1
TCR28	-	-	100	*C. crassicolumna*	KJ197926.1
TCR30	-	-	100	*C. crassicolumna*	KJ197926.1
TCR33	-	-	100	*C. crassicolumna*	KJ197926.1
TCR039	-	-	100	*C. crassicolumna*	KJ197926.1
TCR041	-	-	100	*C. crassicolumna*	KJ197926.1
TCR44	-	-	100	*C. crassicolumna*	KJ197926.1
TCR47	-	-	100	*C. crassicolumna*	KJ197926.1
TCR056	-	-	100	*C. crassicolumna*	KJ197926.1
TCR057	-	-	100	*C. crassicolumna*	KJ197926.1
TCR063	-	-	100	*C. crassicolumna*	KJ197926.1
TCR064.1	-	-	100	*C. crassicolumna*	KJ197926.1
TCR069	-	-	100	*C. crassicolumna*	KJ197926.1
TCR070	-	-	100	*C. crassicolumna*	KJ197926.1
TCR84	-	-	100	*C. crassicolumna*	KJ197926.1
TCM104	-	-	100	*C. crassicolumna*	KJ197926.1
TCM108	-	-	100	*C. crassicolumna*	KJ197926.1
TCM150	-	-	100	*C. crassicolumna*	KJ197926.1
TCM162	-	-	100	*C. crassicolumna*	KJ197926.1
TCM163	-	-	100	*C. crassicolumna*	KJ197926.1

Table 4.6 (continued)

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM166	-	-	100	*C. crassicolumna*	KJ197926.1
TCM167	-	-	99.89	*C. crassicolumna*	KJ197926.1
TCM175	-	-	99.89	*C. crassicolumna*	KJ197926.1
TCM183	-	-	100	*C. crassicolumna*	KJ197926.1
TCM189	-	-	100	*C. crassicolumna*	KJ197926.1
TCM194	-	-	99.89	*C. crassicolumna*	KJ197926.1
TCM203	-	-	100	*C. crassicolumna*	KJ197926.1
TCM215	-	-	99.88	*C. crassicolumna*	KJ197926.1
TCM217	-	-	99.78	*C. crassicolumna*	KJ197926.1
TCM227	-	-	100	*C. crassicolumna*	KJ197926.1
TCM236	-	-	100	*C. crassicolumna*	KJ197926.1
TCM237	-	-	100	*C. crassicolumna*	KJ197926.1
TCM239	-	-	99.89	*C. crassicolumna*	KJ197926.1
TCM241	-	-	100	*C. crassicolumna*	KJ197926.1
TN252	-	-	100	*C. crassicolumna*	KJ197926.1
TN253	-	-	100	*C. crassicolumna*	KJ197926.1
TN254	-	-	99.78	*C. crassicolumna*	KJ197926.1
TN262	-	-	99.88	*C. crassicolumna*	KJ197926.1
PPD281	-	-	100	*C. crassicolumna*	KJ197926.1
TMS296	-	-	100	*C. crassicolumna*	KJ197926.1
TMS306	-	-	100	*C. crassicolumna*	KJ197926.1
TMS309	-	-	100	*C. crassicolumna*	KJ197926.1
TLP312	-	-	100	*C. crassicolumna*	KJ197926.1
TLP314	-	-	100	*C. crassicolumna*	KJ197926.1

Table 4.6 (continued)

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TLP315	-	-	100	*C. crassicolumna*	KJ197926.1
TLP316	-	-	99.89	*C. crassicolumna*	KJ197926.1
TLP317	-	-	100	*C. crassicolumna*	KJ197926.1
TLP318	-	-	100	*C. crassicolumna*	KJ197926.1
TLP319	-	-	100	*C. crassicolumna*	KJ197926.1

Table 4.7 The *matK* Sequences Showed 100% Identical with the Other *Camellia* Species, Followed by the Percentage of Similarity with *C. sinensis*

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM110	99.45	MZ379786.1	100	*C. crassicolumna*	KJ197926.1
TCM170	99.89	MZ379786.1	100	*C. cucphuongensis*	KY010405.1
TCM172	99.89	MZ379786.1	100	*C. cucphuongensis*	KY010405.1
TCM193	99.89	MZ379786.1	100	*C. mingii*	NC_046699.1
TCM196	99.89	MZ379786.1	100	*C. mingii*	NC_046699.1
TCM238	99.89	MZ379786.1	100	*C. mingii*	NC_046699.1

Table 4.8 The *matK* Sequences Showed the Percentage of Similarity with the Other *Camellia* Higher than *C. sinensis*

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR2	99.52	MZ379786.1	99.64	*C. mingii,*	NC_046699.1
TCM98	99.45	MZ379786.1	99.89	*C. gymnogyna*	NC_039626.1
TCM102	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM106	99.10	MZ379786.1	99.55	*C. taliensis*	KJ197934.1
TCM107	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM117	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM141	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM142	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM178	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TCM182	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1
TN280	99.32	MZ379786.1	99.89	*C. crassicolumna*	KJ197926.1
TMS302	99.45	MZ379786.1	99.89	*C. taliensis*	KJ197934.1

Table 4.9 The *matK* sequences showed the percentage of similarity with the other *Camellia* species equal with *C. sinensis*

Code	The percentage of similarity in *matK (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM156	99.76	MZ379786.1	99.76	*C. ptilophylla*	MG946957.1

4.3.2 *psbA-trnH* sequence

Blast analysis revealed that only one tea sample (TCR50) out of 135 has a 100 % identical with *C. sinensis.* The rest of tea plants did not show 100% identical with *C. sinensis*, which having the percentage of similarity with C. *sinensis* ranged from 97.24%-99.80%. Eight tea plants that showed the highest percentage of similarity with *C. sinensis* was 99.79% (TCM98, TCM106, TCM107, TCM117, TCM141, and TMS302), 99.16% (TCM142), and 99.10% (TCM102). Sixty tea plants showed higher percentage of similarity to *psbA-trnH* sequences of the other species of *Camellia* such as *C. gymnogyna, C. weiningensis,* or *C. pubicosta* than *C. sinensis*, which showed the highest the percentage of similarity at 99.80%. Furthermore, 66 tea plants showed the same value of percentage similarity to *C. sinensis* and other species *Camellia*. The lowest percentage of similarities found in TCR2, was 97.49% with *C.lungzhouensis*, 97.49% with *C. ptilophylla*, and 97.24% *C. sinensis* as shown in Table 4.10-4.13

Table 4.10 The *psbA- trnH* Sequences Showed the Percentage of Similarity with Other Species of *Camellia* Higher than *C. sinensis*

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR2	97.24	KX121719.1	97.49	*C. lungzhouensis*	KX121752.1
TCR6	99.61	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCR8	99.41	MH042531.1	99.60	*C. gymnogyna*	NC_039626.1
TCR13	99.01	MH042531.1	99.21	*C. gymnogyna*	NC_039626.1
TCR18	99.21	MH042531.1	99.41	*C. gymnogyna*	NC_039626.1
TCR28	99.41	MH042531.1	99.61	*C. gymnogyna*	NC_039626.1
TCR30	99.20	MH042531.1	99.40	*C. gymnogyna*	NC_039626.1
TCR33	99.60	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCR39	99.01	MH042531.1	99.21	*C. gymnogyna*	NC_039626.1
TCR41	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1

Table 4.10 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR44	99.41	MH042531.1	99.60	*C. gymnogyna*	NC_039626.1
TCR47	99.14	MH042531.1	99.36	*C. gymnogyna*	NC_039626.1
TCR56	98.95	MH042531.1	99.16	*C. gymnogyna*	NC_039626.1
TCR57	99.20	MH042531.1	99.40	*C. gymnogyna*	NC_039626.1
TCR63	99.41	MH042531.1	99.61	*C. gymnogyna*	NC_039626.1
TCR64.1	99.20	MH042531.1	99.40	*C. gymnogyna*	NC_039626.1
TCR69	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TCR70	99.01	MH042531.1	99.21	*C. gymnogyna*	NC_039626.1
TCR84	99.19	MH042531.1	99.40	*C. gymnogyna*	NC_039626.1
TCM104	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM108	99.60	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM110	99.60	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM150	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM162	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TCM163	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM166	99.59	MH042531.1	99.79	*C. gymnogyna*	NC_039626.1
TCM167	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM175	99.60	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM178	98.05	MH042531.1	98.25	*C. gymnogyna*	NC_039626.1
TCM182	98.05	MH042531.1	98.25	*C. gymnogyna*	NC_039626.1
TCM183	99.21	MH042531.1	99.41	*C. gymnogyna*	NC_039626.1
TCM189	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM194	98.96	MH042531.1	99.17	*C. gymnogyna*	NC_039626.1
TCM203	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1

Table 4.10 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM215	99.58	MH042531.1	99.79	*C. gymnogyna*	NC_039626.1
TCM217	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM227	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM229	98.79	MW148820.1	98.99	*C. pubicosta*	KJ806277.1
TCM236	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM237	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM239	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TCM241	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TN252	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TN253	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TN254	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TN262	99.60	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TN273	98.79	MH042531.1	98.99	*C. pubicosta*	KJ806277.1
TN278	98.54	MW148820.1	98.75	*C. pubicosta*	KJ806277.1
TN280	98.57	MH042531.1	98.78	*C. gymnogyna*	NC_039626.1
PPD281	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TMS296	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TMS306	99.59	MH042531.1	99.79	*C. gymnogyna*	NC_039626.1
TLP309	99.38	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TLP312	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TLP314	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TLP315	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1
TLP316	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TLP317	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1

Table 4.10 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TLP318	99.59	MH042531.1	99.80	*C. gymnogyna*	NC_039626.1
TLP319	99.39	MH042531.1	99.59	*C. gymnogyna*	NC_039626.1

Table 4.11 The *psbA- trnH* Sequences Showed the Percentage of Similarity with the Other Species of *Camellia* are Aqual with C. *sinensis*

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR1	98.03	MH042531.1	98.03	*C. weiningensis*	MK820035.1
TCR20	99.41	MH042531.1	99.41	*C. weiningensis*	MK820035.1
TCR42	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCR46	98.03	MH042531.1	98.03	*C. weiningensis*	MK820035.1
TCR74	98.22	MH042531.1	98.22	*C. weiningensis*	MK820035.1
TCR88	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM90	97.79	MH042531.1	97.79	*C. weiningensis*	MK820035.1
TCM99	98.80	MH042531.1	98.80	*C. weiningensis,*	MK820035.1
TCM100	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM113	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM118	98.78	MH042531.1	98.78	*C. weiningensis*	MK820035.1
TCM123	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM125	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM126	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM128.1	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1

Table 4.11 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM129	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM133	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM139	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM143	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM148	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM154	99.59	MH042531.1	99.59	*C. weiningensis*	MK820035.1
TCM156	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM157	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM158	99.59	MH042531.1	99.59	*C. weiningensis*	MK820035.1
TCM160	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM165	99.79	MH042531.1	99.79	*C. weiningensis*	MK820035.1
TCM170	99.59	MH042531.1	99.59	*C. suaveolens*	MT123282.1
TCM172	99.40	MH042531.1	99.40	*C. semiserrata*	MT317096.1
TCM176	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM177	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM180	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM184	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM185	98.95	MH042531.1	98.95	*C. weiningensis*	MK820035.1
TCM186	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM190	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM193	99.39	MH042531.1	99.39	*C. suaveolens*	MT123282.1
TCM196	99.39	MH042531.1	99.39	*C. suaveolens*	MT123282.1
TCM202	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM204	99.59	MH042531.1	99.59	*C. weiningensis*	MK820035.1

Table 4.11 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM205	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM207	99.39	MH042531.1	99.39	*C. weiningensis*	MK820035.1
TCM208	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM212	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM216	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM219	99.39	MH042531.1	99.39	*C. weiningensis*	MK820035.1
TCM222	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM225	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM228	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM231	99.59	MH042531.1	99.59	*C. weiningensis*	MK820035.1
TCM235	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TCM238	99.39	MH042531.1	99.39	*C. suaveolens*	MT123282.1
TN242	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN247	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN249	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN250	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN251	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN256	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TN268	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
PPD282	99.79	MH042531.1	99.79	*C. weiningensis*	MK820035.1
PPD285	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
PPD290	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TMS295	99.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1
TMS297	99.79	MH042531.1	99.79	*C. weiningensis*	MK820035.1

Table 4.11 (continued)

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TLP310	99.59	MH042531.1	99.59	*C. weiningensis*	MK820035.1
TLP311	97.97	MH042531.1	97.97	*C. weiningensis*	MK820035.1
TLP313	98.80	MH042531.1	99.80	*C. weiningensis*	MK820035.1

Table 4.12 The *psbA-trnH* Sequences Showed the Percentage of Similarity with *C. sinensis* Higher than the Other Species of *Camellia*

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCM98	99.79	AY741452.1	98.40	*C. gymnogyna*	NC_039626.1
TCM102	99.10	AY741452.1	97.52	*C. gymnogyna*	NC_039626.1
TCM106	99.79	AY741452.1	98.39	*C. gymnogyna*	NC_039626.1
TCM107	99.79	AY741452.1	98.39	*C. gymnogyna*	NC_039626.1
TCM117	99.79	AY741452.1	98.41	*C. gymnogyna*	NC_039626.1
TCM141	99.79	AY741452.1	98.40	*C. gymnogyna*	NC_039626.1
TCM142	99.16	AY741452.1	97.79	*C. gymnogyna*	NC_039626.1
TMS302	99.79	AY741452.1	98.39	*C. gymnogyna*	NC_039626.1

Table 4.13 The *psbA-trnH* Sequences Showed 100% Identical with *C. sinensis*

Code	The percentage of similarity in *psbA-trnH (%)*				
	with *Camellia sinensis*	NCBI accession	with other *Camellia*	Species	NCBI accession
TCR50	100	MH042531.1	100	*C. weiningensis*	MK820035.1

4.4 Comparing of Blast Results between *matK* and *psbA-trnH* Sequence

In *matK* sequence, 64 sequences of tea plants were 100% identical to *C. sinensis*. Compared with the same tea plants in *psbA-trnH* sequence, only one tea sequence (TCR50) was 100% identical to *C. sinensis* in both DNA barcodes. Fifty-two sequences showed a percentage of similarity with *C. sinensis* ranged from 99.39%-99.80% while 11 sequences showed a percentage of similarity below 99% including TCR1 and TCR46 (98.03%), TCR74 (98.22%), TCM90 (97.79%), TCM99 (98.80%), TCM118 (98.78%), TCM185 (98.95%), TCM229 and TN273 (98.79%), TN278 (98.54%), and TLP311 (97.97%). However, 52 sequences of *matK* did not show or had a lower percentage of similarity with *C. sinensis*, instead of that, it was 100% identical to other *Camellia* species. The 52 sequences of *matK* were compared with *psbA-trnH* and 50 sequences showed a percentage of similarity with *C. sinensis* was between 99.01% -99.60%, which had a percentage of similarity with C. *sinensis* higher than *matK* sequence. There were two sequences showed percentage of similarity below 99% including TCR56 (98.95%) and TCM194 (98.96%). The *matK* sequence, six tea plants including TCM110, TCM170, TCM172, TCM193, TCM196, and TCM238 were 100% identical to other *Camellia* species, while the percentage of similarity with *C. sinensis* was only 99.45%-99.89%. Compared *psbA-trnH* sequence of the same tea plants with the *matK* sequence, five sequences showed the percentage of similarity with C. *sinensis* as same as the other *Camellia* species, including TCM170 (99.59%), TCM172 (99.40%), TCM193 (99.39%), TCM196 (99.39%), and TCM238 (99.39%), while only TCM110 showed a percentage of similarity with the other *Camellia* species (99.80%)

higher than C. *sinensis* (99.60%). In addition, some tea sequences were not 100% identical to C. *sinensis* or other *Camellia* species. In *matK*, twelve tea plants showed the percentage of similarity with *C. sinensis* and the other *Camellia* species ranged from 99.10%-99.52% and 99.64%-99.89%, respectively. Compared 12 sequences of *matK* with the same sequence of *psbA-trnH*, eight tea plants showed the percentage of similarity ranging from 99.10%-99.79% with C. *sinensis*, close to *matK* sequence. There were four sequences of *psbA-trnH* showed a percentage of similarity below 99% including TCR2 (97.24%), TCM178 (98.05%), TCM182 (98.05%), and TN280 (98.57%), which in *matK* showed the percentage of similarity 99.52% (TCR2), 99.45% (TCM170 and TCM182), and 99.32% (TN280), respectively. In *matK*, only TCM156 did not show 100% identical to C. *sinensis* and other *Camellia* species. While in *psbA-trnH*, TCM156 showed the percentage of similarity with C. *sinensis* as same as other *Camellia* species (99.80%).

4.5 Phylogenetic Tree Analysis

The phylogenetic tree was constructed by using *matK* and *psbA-trnH* DNA sequences. The Neighbour-Joining (NJ) tree of *matK* sequences divided all wild tea plants into five groups with ~62% bootstrap value (Figure 4.1). In the first group consisted of 65 tea plants and most tea plants did not have SNP, except TCR42 at aligned position 156 (C to T), and TCM156 at aligned position 110 (T to G). The second group consisted of six tea plants including TCR2, TCM170, TCM172, TCM193, TCM196, and TCM238. Five tea plants had one SNP (G to A) at aligned position 697 (G to A), whereas TCR2 had four SNPs at aligned position 236 (A to T), 351 (T to C), 635 (C to T), and 697 (G to A). The third group, consisted of 10 tea plants with five SNPs at aligned position 103 (C to T), 223 (G to A), 676 (T to G), 697 (G to A), and 749 (G to T). For the fourth group consisted of 46 tea plants with five SNPs at aligned position 210 (T to G), 223 (G to A), 676 (T to G), 697 (G to A), and 749 (G to T) also, but this group has one different SNP position from the third group. For the fifth group consisted of eight tea plants with six SNPs at aligned position 210 (T to G), 223 (G to A), 361 (G to A), 676 (T to G), 697 (G to A), and 749 (G to T). The last three SNP

groups were on the same branch of tree due to there had the same SNP position at aligned position 223 (G to A), 676 (T to G), 697 (G to A), and 749 (G to T). While, the first and second group were on the same branch.

The NJ phylogenetic tree based on *psbA-trnH* sequence of 135 tea plants was constructed and divided wild tea plants into four groups (as shown in Figure 4.2). Sixty-three tea plants were categorized in the first group and most tea plants did not show any SNPs. One tea plant (TCM185) had SNP at aligned position 468 (T to G) while 19 out of 63 were detected insertion or deletion. The insertion was found at positions 170 (3 plants), 241-247 (10 plants), and 449- 452 (5 plants), while deletion was detected at positions 354-358 (6 plant). The second group consisted of 35 tea plants with one SNPs at aligned positions 318 (C to T). The third group consisted of five tea plants with one SNP at aligned position 330 (G to T). The fourth group consisted of 29 tea plants with two SNPs at aligned position 318 (C to T) and 330 (G to T). Additionally, two tea plants, TCR2 and TCM219 did not cluster in any group. TCR2 had four SNPs at aligned positions 66 (C to T), 305 (A to C), 330 (G to T), and 463 (A to G) while TCM219 had only one SNP at aligned position 38 (C to T). The sequence divergence value scale of *matK* and *psbA-trnH* regions were 0.0001, explaining the length of the branch representing the number of nucleotide substitution (number of substitution per 100 nucleotide site). The tea plants with the same position of indels, or having the similarity of genetic distance located close to each other. Moreover, the results showed that some tea plants having both SNPs and indels were mostly found in the same group.

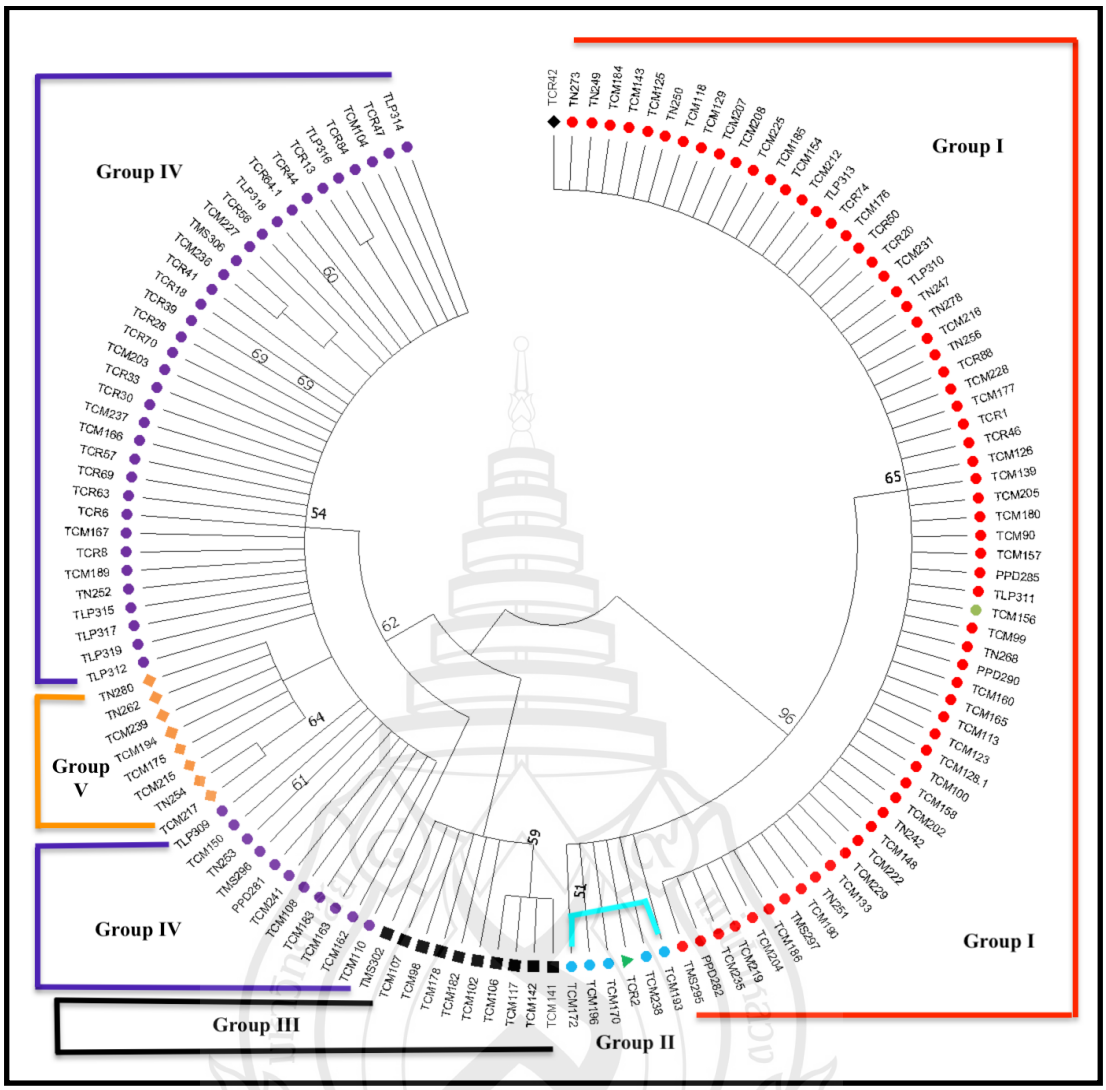


Figure 4.1 The Phylogenetic Tree of *matK* Sequence

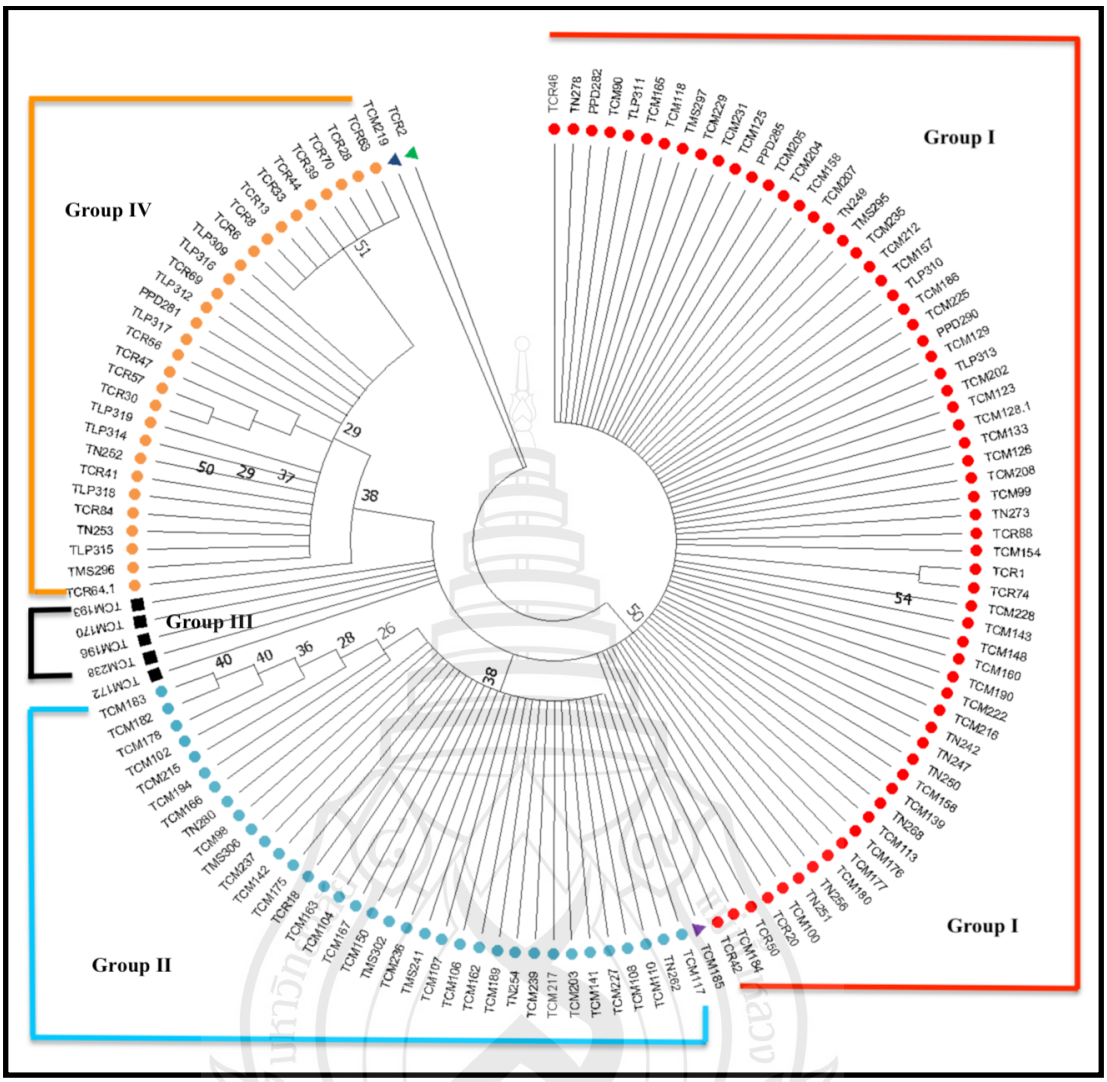


Figure 4.2 The Phylogenetic Tree of *psbA-trnH* Sequence

4.6 Morphological Analysis

Five characteristic traits of leave in this test including color of young shoot, apex, base, margin, and shape were characterized. Most tea plants have green young shoot. Only TCR1, TCR50, TCR74 and TCM193 have red and purple of young shoot, respectively. Five types of leaf apex were observed including acuminate (78.0%), acute (13.2%), emarginate (7.9%), rounded (0.9%), and obtuse (1.8%). For margin, seven types were observed including serrate (70.2%), doubly serrate (5.3%), barely serrate (entire) (14.0%), serrulate (7.9%), undulate-serrate (4.4%), crenate (1.8%), and sinuate-serrate (1.8%), which mainly found serrate margin. In some tea plants, leaf margin is barely serrate or almost entire which mostly found in tea plants in originally Chiang Mai (13 out of 19). For leaf base, only four types were observed including acuminate (25.4%), acute (58.8%), rounded (14.9%), and truncate (0.9%). The acute and acuminate base are mainly found while truncate and rounded found only one (TCM193) and 18 (TCM98, TCM102, TCM103, TCM106, TCM108, TCM143, TCM163, TCM170, TCM172, TCM176, TCM194, TCM196, TCM215, TCM219, TN249, TN262, TN280, and TMS302) from the rest of the tea plants, respectively. For shape of tea leaf, six types were observed including elliptical (51.8%), ovate (36.0%), lanceolate (11.4%), oblanceolate (6.1%), oblong (2.6%) and obovate (0.9%). Elliptical and ovate are found the most in tea plants. The morphology of tea plants as shown in Table 4.14. The varieties and regions do not effect on the morphological characters of wild tea plants. Only 114 from 135 tea plants were observed for morphological test because 21 tea plants dead when we recollected.

Table 4.14 Morphological traits of wild tea leaf

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TCR1	red	acuminate	serrate	acute	elliptical-ovate	11.5	4.5
TCR2	No data						
TCR6	green	acuminate	serrate	acute	elliptical	13.4	4.6
TCR8	No data						
TCR13	green	acuminate	serrate	acute	elliptical	12.0	5.1
TCR18	green	acuminate	serrate (really sharp)	acute	elliptical-lanceolate	13.2	4.7
TCR20	green	acuminate	serrate	acute	elliptical	21.5	6.5
TCR28	No data						
TCR30	green	acute	serrate	acute	elliptical-ovate	16.0	6.0
TCR33	No data						
TCR39	green	acuminate	serrate-undulate	acute	elliptical	11.7	5.5
TCR41	green	acuminate	serrulate-undulate	acute	elliptical-ovate	18.8	6.5
TCR42	green	acuminate	serrulate	acute	elliptical	10.5	4.7
TCR44	green	acuminate	serrate (sharp)	acute	elliptical-oblanceolate	14.5	5.0
TCR46	green	acuminate	serrate (sharp)-undulate	acuminate	elliptical	13.8	4.5
TCR47	green	acuminate	serrate	acute	ovate	16.3	6.2
TCR50	red	acuminate	crenate	acute	elliptical	18.4	5.5
TCR56	green	acuminate	serrate	acute	elliptical-ovate	16.5	6.7
TCR57	No data						
TCR63	green	acuminate	serrate-undulate	acute	ovate	15.3	6.0

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TCR64.1				No data			
TCR69				No data			
TCR70				No data			
TCR74	red	acuminate	serrate	acute	ovate	15.3	6.0
TCR84				No data			
TCR88				No data			
TCM90	green	acuminate	serrate	acuminate	elliptical	11.0	4.6
TCM98	No data	acute	serrate	rounded	elliptical	11.3	4.7
TCM99	No data	acuminate	serrate	acuminate	elliptical	13.5	4.3
TCM100	green	acuminate	serrulate	acute	ovate	13.6	5.8
TCM102	green	acuminate	serrate	rounded	elliptical	14.3	5.8
TCM104	green	acuminate	serrate	acute	ovate	10.6	4.0
TCM106	No data	emarginate	barely serrate	rounded	elliptical	10.7	4.5
TCM107	No data	acuminate	serrate	acuminate	ovate	12.5	5.1
TCM108	No data	acuminate	serrate	rounded	ovate	12.7	4.5
TCM110	No data	acuminate	serrate	acute	elliptical	12.0	4.5
TCM113	green	acuminate	serrate	acuminate	ovate	10.9	4.0
TCM117	green	acuminate	serrate	acute	elliptical	16.2	6.3
TCM118	No data	acuminate	serrate	acuminate	lanceolate	15.6	5.4
TCM123	green	acuminate	serrate	acute	elliptical	15.0	5.6
TCM125	No data	acuminate	serrulate	acute	elliptical	10.7	4.5
TCM126	green	acuminate	serrate	acute	oblanceolate	12.0	5.1

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TCM128.1	No data	emarigate	serrate	acute	ovate	6.4	5.0
TCM129	green	acuminate	sinuate	acute	elliptical	12.3	5.7
TCM133	No data	acute	serrate	acute	oblanceolate	8.6	3.3
TCM139	No data	emarginate-acuminate	serrate	acute	oblanceolate	9.2	3.4
TCM141	green6	acuminate	barely serrate	acuminate	lanceolate	8.1	2.7
TCM142	green	acute	barely serrate	acute	elliptical	11.0	4.2
TCM143	No data	acuminate	serrate	rounded	ovate	11.1	4.9
TCM148	green	acuminate	serrulate	acute	ovate	11.5	4.7
TCM150	green	acute	doubly serrate	acute	ovate	11.2	4.8
TCM154	green	acuminate	serrate	acuminate	elliptical	13.0	4.8
TCM156	green	acuminate	barely serrate	acuminate	lanceolate	8.5	2.8
TCM157	green	acuminate	barely serrate	acuminate	ovate	8.5	3.1
TCM158	No data	acuminate	serrulate	acuminate	ovate	13.5	4.8
TCM160	green	acuminate	serrate	acuminate	lanceolate	8.9	3.2
TCM162	green	acuminate	doubly serrate	acuminate	elliptical	11.0	3.7
TCM163	green	acuminate	serrate	rounded	elliptical	8.1	3.1
TCM165	No data	emarginate	crenate	acuminate	lanceolate	8.6	2.9

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TCM166	No data	acuminate	doubly serrate	acute	elliptical	9.1	3.3
TCM167	No data	acuminate	serrulate	acuminate	elliptical	6.8	2.8
TCM170	No data	rounded	barely serrate	rounded	ovate	8.9	4.3
TCM172	No data	acute	barely serrate	rounded	ovate	7.7	3.8
TCM175	No data	acuminate	serrate	acuminate	elliptical	8.5	3.4
TCM176	green	acute	doubly serrate	rounded	elliptical	14.3	5.8
TCM177	No data	acuminate	serrate	acute	elliptical	10.2	3.6
TCM178	No data	acuminate	serrate	acute	ovate	9.2	3.9
TCM180	green	acuminate	serrate	acute	elliptical	9.9	3.8
TCM182	No data	emarginate	serrate	acute	elliptical	9.4	3.5
TCM183	green	acuminate	serrate	acute	elliptical	8.5	2.8
TCM184	No data	acuminate	serrate	acuminate	ovate	11.1	4.0
TCM185	green	acuminate	serrate	acuminate	lanceolate	9.3	3.3
TCM186	No data	acuminate	serrate	acuminate	oblanceolate	11.7	4.0
TCM189	green	acuminate	serrate	acuminate	oblanceolate	8.2	2.8
TCM190	green	acuminate	serrate	acute	elliptical	9.1	3.5
TCM193	purple	acuminate	barely serrate	truncate	ovate	6.0	2.9
TCM194	green	acuminate	serrate	rounded	elliptical	7.9	3.5
TCM196	No data	obtuse	barely serrate	rounded	ovate-oblong	9.2	4.6
TCM202	No data	acute	serrate	acute	ovate	10.4	4.8
TCM203	green	acuminate	serrate	acute	lanceolate	8.3	3.0

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TCM204	green	acuminate	serrate	acuminate	ovate	7.7	7.7
TCM205	green	acuminate	barely serrate	acuminate	lanceolate	7.5	2.6
TCM207	No data	acuminate	serrate	acute	elliptical	12.8	4.4
TCM208	No data	acute	serrulate	acute	elliptical	10.7	3.9
TCM212	No data	emarginate	serrate	acuminate	ovate	8.6	3.7
TCM215	green	acute	serrate	rounded	ovate	6.5	2.9
TCM216	No data	acute	barely serrate	acute	elliptical	10.3	3.7
TCM217	No data	acuminate	serrate	acuminate	elliptical	9.5	3.7
TCM219	green	emarginate	barely serrate	rounded	elliptical	10.5	4.5
TCM222	green	acuminate	serrate	acute	elliptical	7.5	2.6
TCM225	green	acuminate	serrulate	acute	lanceolate	7.7	2.5
TCM227	No data	acuminate	barely serrate	acuminate	obovate	15.1	6.5
TCM228	No data	acuminate	serrate	acuminate	elliptical	13.6	5.2
TCM229	green	acuminate	serrate	acute	elliptical	9.2	3.9
TCM231	No data	acuminate	doubly serrate	acute	ovate	10.3	4.1
TCM235	No data	emarginate	serrate	acute	elliptical	11.9	4.6
TCM236	green	acuminate	serrate	acute	elliptical	9.9	3.6
TCM237	No data	acuminate	serrate	acute	ovate	13.5	5.0
TCM238	No data	obtuse	serrate	acute	ovate	11.6	4.9
TCM239	green	acuminate	sinuate	acute	ovate	7.0	3.3
TCM241	No data	acute	serrate	acute	ovate	8.5	4.1

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TN242	No data	acuminate	serrate	acute	elliptical	12.0	4.8
TN247	green	acuminate	serrate	acute	Elliptical-lanceolate	12.2	5.1
TN249	No data	acumiante	serrate	rounded	elliptical	11.3	4.5
TN250	green	acuminate	Serrate-undulate	acute	elliptical	11.0	4.8
TN251	No data	acute	serrate	acuminate	ovate	8.7	3.9
TN252	No data	acuminate	serrate	acute	ovate	9.3	4.1
TN253	green	emarginate	serrate	acute	elliptical	11.7	4.8
TN254	green	acuminate	serrate	acuminate	elliptical	12.2	4.6
TN256	No data	acuminate	barely serrate	acute	elliptical	12.1	4.5
TN262	green	acuminate	serrate	rounded	ovate	9.6	4.3
TN268	green	acuminate	serrate	acute	ovate	9.4	4.0
TN273	green	acuminate	barely serrate	acute	elliptical	11.0	3.8
TN278	green	acuminate	serrate	acute	lanceolate	12.3	4.3
TN280	green	acuminate	serrate	rounded	ovate	9.4	3.5
PPD281	No data	acuminate	serrate	acute	elliptical-oblong	11.4	4.2
PPD282	No data	acuminate	serrate	acuminate	lanceolate	11.6	4.7
PPD285	No data	acuminate	serrate	acute	ovate	9.6	3.6
PPD290	No data	acute	serrate	acute	ovate	9.5	4.1
TMS295	green	acuminate	serrate	acute	elliptical	10.9	4.3
TMS296	No data	acuminate	serrate	acute	oblanceolate	16.8	6.3
TMS297	No data	acuminate	serrate	acute	ovate	11.0	4.2

Table 4.14 (continued)

Code	Morphology of leaf						
	Color of young tips	**Apexes**	**Margins**	**Bases**	**Shapes**	**Length (cm)**	**Width (cm)**
TMS302	green	acute	serrate	rounded	elliptical-oblong	10.8	4.3
TMS306	green	acuminate	doubly serrate	acute	elliptical	6.7	2.9
TLP309	No data						
TLP310	No data						
TLP311	No data						
TLP312	No data						
TLP313	No data						
TLP314	No data						
TLP315	No data						
TLP316	No data						
TLP317	No data						
TLP318	No data						
TLP319	No data						

4.7 Morphological Tree Analysis

The morphological tree was constructed based on the morphological traits using Past 4 program as show in Figure 4.3. All tea plants were categorised into four major groups. There were 59 tea plants in the first group, which was separated into two cluster. Cluster A included many tea plants with elliptical-shaped leaves, while the cluster B showed ovate shape. Additionally, the Group I contained three out of four tea plants with red young tips. In this group and other groups, the remaining tea plants tips were green color. Group II consisted of 40 tea plants, which the most common morphological traits were barely serrate and doubly serrate. Similar morphological traits, including

elliptical shaped and ovate shape, were seen in both group I and II. Only 15 tea plants were clustered in the group III, and nearly all of them had acuminate bases and lanceolate shapes. Lanceolate shape was not quite found in the other groups. Group IV consisted of 21 tea plants with had no morphological traits data because it had already died, when we went back to collect it. Two morphological traits, acuminate apex and serrate margin, were found in almost all tea plants. In each group, the variations in length and width of leaves were comparable. As a result, each group in the morphological tree showed morphological traits that were quite comparable.

The morphological tree and the phylogenetic tree were compared. All tea plants were divided into the same number of groups in the morphological tree as there are in the phylogenetic tree. In NJ phylogenetic tree, all tea plants were divided into different groups based on the number and position of SNPs in the *matK* and *psbA - trnH* genes. In contract, the morphological tree divided groups of tea plants into those that mixed with SNPs and those that did not. Some tea plants were still grouped or placed nearby with the same tea plants as in the NJ phylogenetic tree. For example, in *matK*, eight tea plants (TCM175, TCM194, TCM215, TCM217, TCM239, TN254, TN262, and TN280), which were grouped together in the NJ phylogenetic tree, were divided into two different groups in the morphological tree (TCM175, TCM194, TCM215, TCM217, TN262, and TN280 found in the first group, and TCM239 and TN254 found in the second group). In *psbA-trnH*, five tea plants, TCM170, TCM172, TCM193, TCM196, and TCM238 were grouped together in the NJ phylogenetic tree while the morphological tree, five tea plants were divided into three different groups, TCM196 and TCM238 being found in the first group, TCM172 and TCM193 being found in the second group, and TCM170 being found in the third group. However, some tea plants were grouped with other new tea plants. In addition, tea plants that divided in the NJ phylogenetic tree by SNPs position also showed difference in morphological traits because the environmental factors that surround tea plants may have an impact on tea plants even in a group of tea plant from the same region. In response to environmental stress, phenotypic plasticity mechanisms can lead to morphological variation.

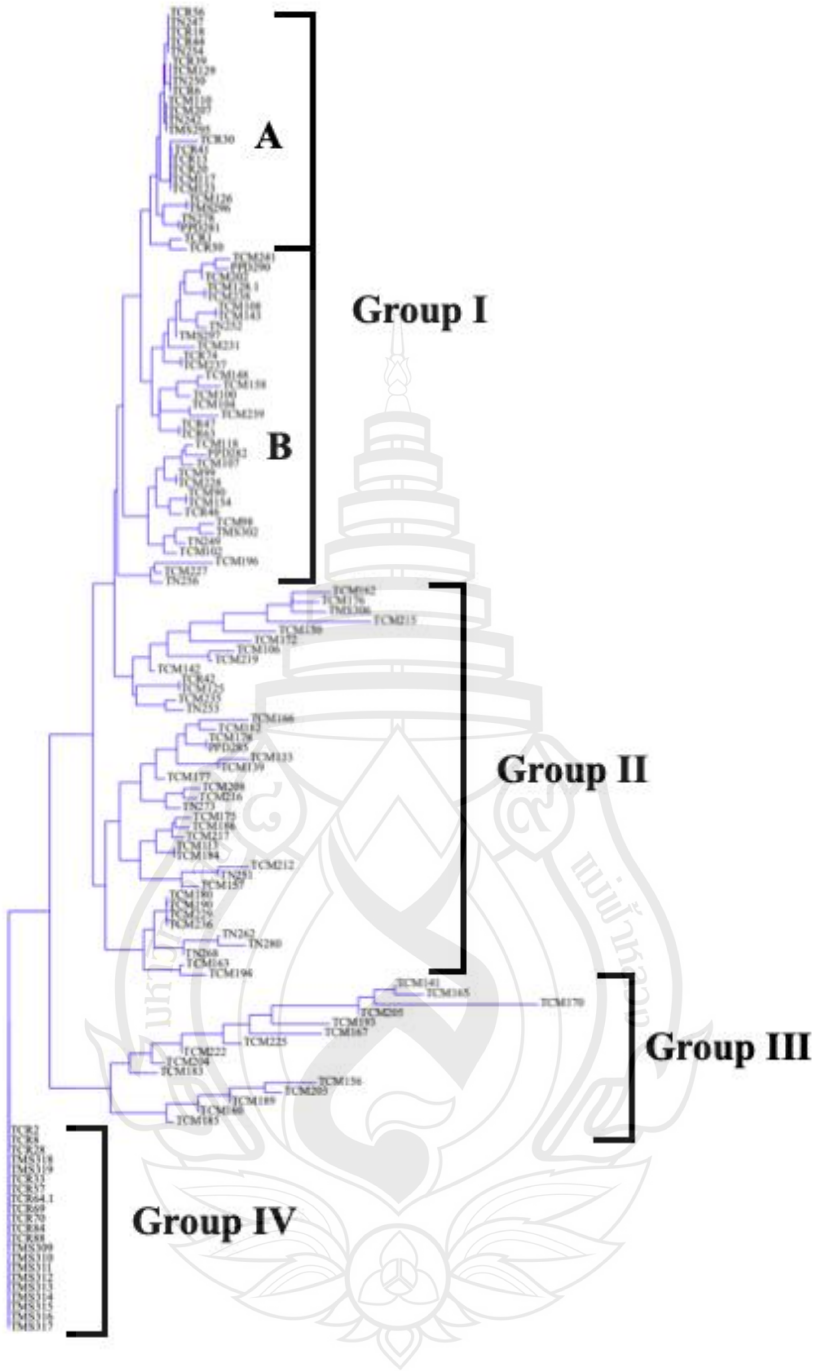


Figure 4.3 The Morphological Tree of Wild Tea Plants

CHAPTER 5

DISCUSSION

5.1 PCR Amplification and Sequencing Analysis

Tea (*Camellia sinensis* (L.) O. Kuntze) has been widely consumed for 3000 years since ancient times. Many kinds of tea, such as black tea, white tea, oolong tea, or herbal tea are producing from the leaves of tea plants in the genus *Camellia*. Tea is considered as one of the three beverages most popular in the world (Yan et al., 2020). In many countries, green tea and black tea are the two most popular types of tea to drink, especially black tea prepared from Assam tea or wild tea varieties. The morphological resemblance between tea varieties makes it difficult to distinguish using traditional identification technique such as organoleptic or microscopic methods. It must be necessary to use an accurate identification procedure.

This study aimed to determine the best DNA barcode for identifying wild tea plants. DNA barcodes may be useful in identifying a variety of plant species, including vitally important medicinal species, poison species, rare species, and processed products, such as processed products of *Phaleria macrocarpa* (Scheff.) (Rahman & Ali, 2020); poisonous plants in China (Xie et al., 2014); and medicinal plants in Saint Katherine protectorate (Hashim et al., 2021). DNA barcoding is increasingly being used not only to identify species, but also to detect substitution species,

adulterant species of economically crops, herbal medicine (Chen et al., 2010), and trade for wild or harvested plants and animal species (Wasser et al., 2007). This method is highly important because sometimes traditional methods of identification based on morphological traits may not be enough and fail to distinguish between species (Duminil & Di Michele, 2009). The criteria for choosing an appropriate DNA barcode region require (1) sufficient nucleotide sequence variability to discriminate between closely related species, (2) a comprehensive reference database, and (3) the capacity to obtain the barcode sequences from the samples (Kress & Erickson, 2008).

All tea plants tested positive for PCR amplification and sequencing in *matK* with a 100% success rate. *matK* has been previously been shown to be effective in PCR amplification and DNA sequencing in studies of tea plants such as elite tea clones of Darjeeling and Doors (*Camellia sinensis*) (Reha et al., 2021) or commercial teas (Stoeckle et al., 2011). But according to several earlier studies, *matK* failed or was challenging to amplify in some plant species, such as angiosperms (Yu et al., 2011); *Myristica fragrans* Houtt.(Swetha et al., 2017); Eurasian yews (Liu et al., 2010); durian (Cahyaningsih et al., 2021); aquatic freshwater plants (Hoveka et al., 2016). This could be due to a high rate of primer site substitution, which makes amplification challenging. According to the CBOL Plant Working Group. (2009) *matK* showed the highest discrimination efficiency for distinguish the species using only single barcode loci. As a result, it was suggested as one of *the* core barcodes for plants. Our blast analysis using *matK* showed discriminatory power between 99% and 100% with *C. sinensis*, which 64 of 135 tea plants showed 100% identity. Similar results with Reha et al. (2021) elite tea clones of Darjeeling and Dooars (*Camellia sinensis*) showed 24 of 29 tea samples showed 100% with *C. sinensis* and the remaining showed between 99.29% and 99.89%. In other plant species studies, such as on Dipterocarps in Sumatra, Indonesia, *matK* shown significant discriminatory power in comparison to other plant species (Carneiro de Melo Moura et al., 2019). In a related work by Worthy et al. (2022), *matK* showed discriminatory power of 63% at the species level and 78% at the genus level of montane forest orchids; *Aloe vera* showed similarity with *Aloe shadensis* at 99.76% (Algarni, 2022). Also, *matK* showed the highest discriminatory power in *Fritillariae cirrhosae* bulbus (Q. Chen et al., 2020), *Dioscorea* species (Sun et al., 2012), and Vicia species

(Raveendar et al., 2017). For alignment analysis, 12 SNPs were detected in 72 tea plants. Compared with other tea plant varieties, Reha et al. (2021) found that *matK* can detect 14 SNPs in tea varieties from tea clones of Darjeeling and Dooars. In other plant species, *matK* can detect 21 SNPs in *Tulipa edulis* (Ma et al., 2014). The type of substitution base at SNP positions in *matK* showed more transition than tranversion DNA substitution. 7 SNPs in total, aligned at positions 103, 156, 223, 351, 361, 635, and 697, showed transition bases while 5 SNPs at aligned positions 110, 210, 236, 676, and 749 showed transversion base from 12 SNPs.

In *psbA-trnH* gene, all tea plants showed 100% PCR amplification and sequencing efficiency the same as the *matK* gene. The *psbA-trnH* barcode is recommended as easily amplified and sequenced with an average length of 465 bp across the 30 taxa (Kress et al., 2005); angiosperm, gymnosperm, ferns, mosses, and liverworts (Kress et al., 2005); and *Plumeria* flower (Zhao et al., 2018). Our blast analysis for *psbA-trnH* revealed discriminatory power ranging from 97% to 100%. Only one tea plant demonstrated complete identity (100%) with *C. sinensis* while the other tea plants showed a degree of similarity to *C. sinensis* varying between 97.24% to 99.80%. The *psbA-trnH* demonstrated 95% and 100% at the species and genus levels of the samples in the Apocynaceae family (Lv et al., 2020). In poison medicinal plants, results showed 100% identification efficiency at the genus level (Liu et al., 2019). Similar results were found in *Rhododendron* (Ericaceae), *psbA-trnH* showed 100% species identification efficiency in 24 samples from 13 different species (Chen et al., 2012). Furthermore, identification effectiveness of medicinal plants in the fabaceae family was 100% at the genus level, with the finding indicated that 95 of 104 species were appropriately classified (Gao et al., 2013). In sequence analysis of *psbA-trnH*, seven SNPs (72 tea plants) and four indels (19 tea plants) were detected. According to research on the other plant species, *psbA-trnH* can identify 25 variable sites in Vincetoxicum (Apocynaceae-Asclepiadodeae) and 60 SNPs and 9 indwells in Myristica fragrant Houtt. (Swetha et al., 2017). Also, *Tylophora japonica* and *T. floribunda* showed large indels, which had 103 bp and 114 bp deletion, respectively. The other species of *Tylophora* and all species of *Vincetoxicum* had a 121 bp deletion (Yamashiro et al., 2004). Compared to the other plants, 60 SNPs and 9 indels were

found in the *Myristica fragrant* Houtt. species, which showed a high sequence variation between M. fragrant and its adulterant, *M. malabarica* (Swetha et al., 2017). In *Vincetoxicum* (*Apocynaceae-Asclepiadoideae*), 25 variable sites were detected. In the phylogenetic relationship of artemisia species, 11, 2, and 1 SNPs were detected in *Artemisia*, *Dracunculus*, and *Serphidium*, respectively (Haghighi et al., 2014). The results of DNA substitution in *psbA-trnH* showed that 4 SNPs at aligned positions 38, 66, 318, and 463 showed transition substitution base, whereas 3 SNPs at aligned positions 305, 330, and 468 showed transversion substitution base from 7 SNPs. Moreover, *psbA-trnH* also has quite serious problems with alignment for this locus causing the high frequencies of insertions or deletions (Chase et al., 2007). Insertions/deletions and short sequence repeats typically occur more frequently than single base pair substitutions in *psbA-trnH* (Bruni et al., 2010). Additionally, the *psbA-trnH* region showed a high number of insertions/deletions that could be problematic for the correct identification. So, the alignment of *psbA-trnH* sequence can be highly ambiguous because of its complicated molecular evolution, considerable length variation (Chang et al., 2006), and high rate of insertion/deletion in large families of angiosperms (Chase et al., 2007). In addition, indels in this locus are said to be non-informative for species discrimination because their occurrence does not correlate with a species (Swetha et al., 2017).

5.2 The Phylogenetic Tree

According to the phylogenetic analysis, Neighbour-joining (NJ) was performed with both DNA barcode regions. Tea plants were divided into many different groups according to their number and position of SNP. Because SNP profiles are useful for differentiating closely related species, they have been widely used in molecular diagnostics (Merrill et al., 2016).

In *matK*, all tea plants were divided into five groups according to the SNP positions of each tea plant, which were supported by a bootstrap value of 96%. Similar to our result with Reha et al. (2021), all the nucleotide sequences were grouped together except Thurbo 3 (11125), Thurbo 9 (11126), which clustered together and showed two

unique variable sites that differed from the rest of the sequences. *Momordica* (Cucurbitaceae) species were studied by Ramesh et al. (2022). M. *dioica, M. sahyadrica ssp. anamalayana,* and *M. sahyadrica* ssp. *sahyadrica* shared certain SNPs and were grouped together in the phylogenetic tree.

In *psbA-trnH,* tea plants, like *matK,* can be divided into four groups based on SNP *positions*, with all SNP groups supported by a bootstrap value of 50%. In *artemisia* species, the number of SNP positions was not affected by groups of samples in the phylogenetic tree. In the *Plumeria* flower, three species of flower can be clustered into three groups in the NJ phylogenetic tree, according to the K2P genetic distance model (Zhao et al., 2018).

From the results above, tea plants showed a close relationship with each other based on the bootstrap value. Tea plants that had varied morphological characteristics but shared the same SNP position or number of SNPs were grouped together. For example, TCR2, TCR50, TCR74, and TCM193 have red and purple young shoots, but they were categorised into different groups in morphological tree. Tea plants with doubly-serrate leaf margins 6 from 114 (TCM150, TCM162, TCM166, TCM176, and TCM231) were categorised *into* three different groups in the morphological tree. Other morphological characters could be found in any group of the phylogenetic tree in both *matK* and *psbA-trnH.* Varieties or the originally cultivated region of tea plants do not have an effect on SNPs and groups in the phylogenetic tree for either of the two DNA barcode regions. As a results, they maybe the same variety, but their morphology differed depending on geographical region and climate. Additionally, tea plants are a dominantly cross-pollinated plant and are normally found between tea varieties. Also, it could be due to phenotypic plasticity in response to pressures in the environment.

Our study showed that *matK* can distinguish wild tea plants by showing a higher number of SNPs and the number of SNP patterns than *psbA-trnH.* Therefore, the power of discrimination of *matK* better than *psbA-trnH* region. The results showed the successfully of using *psbA-trnH* and *matK* to explores the genetic diversity and variation in wild tea by using SNPs, genetic distance, and phylogenetic tree analysis. From our study, DNA extraction, PCR amplification, DNA sequencing, and BLAST are the techniques that are really crucial to use for species identification of wild tea

plants. All the techniques of DNA barcoding can be done in a rudimentary laboratory by technicians trained in basic molecular biology when compared with traditional techniques. Thus, DNA barcoding technique can be a powerful tool for identifying plant species and their origins.

REFERENCES

REFERENCES

Afzal, M., Safer, A., & Menon, M. (2015). Green tea polyphenols and their potential role in health and disease. *Inflammopharmacology*, *23*(4), 151-161. https://doi.org/10.1007/s10787-015-0236-1

Agarwal, U., Pathak, D. P., Bhutani, R., Kapoor, G., & Kant, R. (2017). Review on Camellia sinensis –Nature's Gift. *International Journal of Pharmacognosy and Phytochemical Research*, *9*(08), 1119-1126. http://doi.org/10.25258/phyto.v9i08.9619

Ahmed, S., Unachukwu, U., Stepp, J. R., Peters, C. M., Long, C. L., & Kennelly, E. (2010). Pu-erh tea tasting in Yunnan, China: Correlation of drinkers' perceptions to phytochemistry. *J. Ethnopharmacol*, *132*, 176–185. https://doi.org/10.1016/j.jep.2010.08.016

Alcázar, A., Ballesteros, O., Jurado, J., Pablos, F., Martín, M., Vilches, J., & Navalón, A. (2007). Differentiation of green, white, black, oolong, and pu-erh teas according to their free amino acids content. *Journal of Agricultural and Food Chemistry*, *55*(15), 5960-5965. https://doi.org/10.1021/jf070601a

Algarni, A. (2022). Molecular identification and phylogenetic analysis of *Aloe shadensis* from Saudi Arabia based on *matK*, *rbcL* and ITS DNA barcode sequence. *Saudi Journal of Biological Sciences*, *29*(2), 1125-1133. https://doi.org/10.1016/j.sjbs.2021.09.053

Amazing Thailand. (2022). *Northern and northeast regions*. https://www.amazing-thailand.com/north-northeast-thailand.html.

Beeck, M., Lievens, B., Busschaert, P., Declerck, S.,Vangronsveld, J., & Colpaert, J. V. (2014). Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *Plos One*, *9*(6), e97629. https://doi.org/10.1371/journal.pone.0097629

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiology*, *10*(1), 189. https://doi.org/10.1186/1471-2180-10-189

Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M., & Labra, M. (2010). Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine*, *124*(6), 595-603. https://doi.org/10.1007/s00414-010-0447-3

Cahyaningsih, Y., Sukartini, Sutanto, A., Santoso, P., Arsyad, M., & Larekeng, S. (2021). Amplification rate of *matK* and *rbcL* genes in three types of durian. *IOP Conference Series: Earth and Environmental Science*, *886*(1), 012011. https://doi.org10.1088/1755-1315/886/1/012011

Carneiro de Melo Moura, C., Brambach, F., Jair Hernandez Bado, K., Krutovsky, K. V., Kreft, H., Tjitrosoedirdjo, S. S., . . . Gailing, O. (2019). Integrating DNA barcoding and traditional taxonomy for the identification of Dipterocarps in remnant lowland forests of Sumatra. *Plants*, *8*(11), 461. https://doi.org/10.3390%2Fplants8110461

CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA, 106*(31), 12794-12797. https://doi.org/10.1073/pnas.0905845106

Chacko, S. M., Thambi, P. T., Kuttan, R., & Nishigaki, I. (2010). Beneficial effects of green tea: A literature review. *Chinese Medicine*, *5*(1), 1-9. https://doi.org/ 10.1186/1749-8546-5-13

Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, . . . Chaw, S. M. (2006). The chloroplast genome of Phalaenopsis aphrodite (*Orchidaceae*) : Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molecular Biology and Evolution, 23*(2), 279-291. https://doi.org/10.1093/ molbev/msj029

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madrinan, S., Petersen, G., . . . Wilkinson, M. (2007). A proposal for a standardised protocol tobarcode all land plants. *Taxon, 56*(2), 295-299. https://doi.org/10.1002/tax.562004

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., . . . Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *Plos One, 5*(1), e8613. https://doi.org/10.1371/journal.pone.0008613

Chen, D., Ding, Y., Chen, G., Sun, Y., Zeng, X., & Ye, H. (2020). Components identification and nutritional value exploration of tea (*Camellia sinensis* L.) flower extract: Evidence for functional food. *Food Research International, 132*(2), 109100. https://doi.org/ 10.1016/j.foodres.2020.109100

Chen, K., Liu, Y., Zhang, L., Liu, Z., Luo, K., & Chen, S. (2012). Species identification of Rhododendron (Ericaceae) using the chloroplast deoxyribonucleic acid PsbA-trnH genetic marker. *Pharmacognosy Magazine, 8*(29), 29-36. https://doi.org/10.4103%2F0973-1296.93311

Chen, Q., Wu, X., & Zhang, D. (2020). Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of Fritillariae cirrhosae bulbus and its adulterants. *Plos One, 15*(2), e0229181. https://doi.org/10.1371/journal.pone.0229181

Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., & Leon, C. (2014a). A renaissance in herbal medicine identification from morphology to DNA. *Biotechnology Advances*, *32*(7), 1237-1244. https://doi.org/10.1016/j.biotechadv. 2014.07.004

Chen, S., Zhu, Z., Ma, H., Yang, J., & Guo, Q. (2014b). DNA barcodes for discriminating the medicinal plant *Isatis indigotica Fort.* (*Cruciferae*) and its adulterants. *Biochemical Systematics and Ecology*, *57*, 287-292. https://doi.org/10.1016/j.bse.2014.08.007

Cuenoud, P., Savolainen, V., Chatrou, L., Powell, M., Grayer, R., & Chase, M. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, *89*(1), 132-144. https://doi.org/10.3732/ajb.89.1.132

Da Silva Pinto, M. (2013). Tea: A new perspective on health benefits. *Food Research International*, *53*(2), 558-567. https://doi.org/10.1016/j.foodres.2013.01.038

Dattner, C., Boussabba, S., & Emmanuelle, J. (2003). *The book of green tea*. Universe Books.

De Castro, O., Comparone, M., Di Maio, A., Del Guacchio, E., Menale, B., Troisi, J., . . . Guida, M. (2018). Correction: What is in your cup of tea? DNA Verity Test to characterize black and green commercial teas. *Plos One*, *13*(1), e0192334. https://doi.org/10.1371/journal.pone.0192334

De Costa, W., Mohotti, A., & Wijeratne, M. (2007). Ecophysiology of tea. *Brazilian Journal of Plant Physiology*, *19*(4), 299-332. https://doi.org/10.1590/S1677-04202007000400005

Duminil, J., & Di Michele, M. (2009). Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosystems-an International Journal Dealing with all Aspects of Plant Biology*, *143*(3), 528-542. https://doi.org/10.1080/11263500902722964

Eden, T. (1958). The development of tea culture. In T. Eden (Ed.), *Tea* (pp. 1-4). Longman.

Ellis, B., Daly, D. C., Hickey, L. J., Johnson, K. R., Mitchell, J. D., Wilf, P., & Wing, S. L. (2009). *Manual of leaf architecture*. Cornell University Press.

Fajarningsih, N. (2016). Internal transcribed spacer (ITS) as DNA barcoding to identify fungal species: A review. *Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology*, *11*(2), 37-44. https://doi.org/ 10.15578/squalen.v11i2.213

Fay, M. F., Swensen, S. M., & Chase, M. W. (1997). Taxonomic Affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin*, *52*(1), 111-120. https://doi.org/10.2307/4117844

Ferri, G., Corradini, B., Ferrari, F., Santunione, A., Palazzoli, F., & Alu, M. (2015). Forensic botany II, DNA barcode for land plants: Which markers after the international agreement?. *Forensic Science International: Genetics*, *15*, 131-136. https://doi.org/10.1016/j.fsigen. 2014.10.005

Fouad, A., Hamed, A., Amer, W., & Hafez, R. (2021). Barcoding of some plant species using rbcL gene in the Mediterranean oolitic sand dunes west of alexandria, Egypt. *Egyptian Journal of Botany*, *62*(1), 159-168. https://doi.org/10.21608/ejbo. 2021.74134.1683

Gao, T., Ma, X., & Zhu, X. (2013). Use of the psbA-trnH region to authenticate medicinal species of Fabaceae. *Biological and Pharmaceutical Bulletin*, *36*(12), 1975-1979. https://doi.org/10.1248/bpb.b13-00611

Gielly, L., & Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus *rbcL* sequences. *Molecular Biology and Evolution, 11*(5), 769-777. https://doi.org10.1093oxfordjournals.molbev.a040157

Graham, H. (1992). Green tea composition, consumption, and polyphenol chemistry. *Preventive Medicine*, *21*(3), 334-350. https://doi.org/10.1016/0091-7435(92)90041- f

Haghighi, A., Belduz, A., Vahed, M., Coskuncelebi, K., & Terzioglu, S. (2014). Phylogenetic relationships among Artemisia species based on nuclear ITS and chloroplast psbA-trnH DNA markers. *Biologia*, *69*(7), 834-839. https://doi.org/10.2478/s11756-014-0379-3

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. https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29

Hashim, A., Alatawi, A., Altaf, F., Qari, S., Elhady, M., Osman, G., & Abouseadaa, H. (2021). Phylogenetic relationships and DNA barcoding of nine endangered medicinal plant species endemic to Saint Katherine protectorate. *Saudi Journal of Biological Sciences*, *28*(3), 1919-1930. https://doi.org/10.1016/j.sjbs.2020.12.043

Hayat, K., Iqbal, H., Malik, U., Bilal, U., & Mushtaq, S. (2015). Tea and its consumption: Benefits and risks. *Critical Reviews in Food Science and Nutrition*, *55*(7), 939-954. https://doi.org/10.1080/10408398.2012.678949

Hebert, P. D., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings. Biological sciences*, *270*(1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

Hilal, Y., & Engelhardt, U. (2007). Characterisation of white tea comparison to green and black tea. *Journal Für Verbraucherschutz Und Lebensmittelsicherheit*, *2*(4), 414-421. https://doi.org/10.1007/s00003-007-0250-3

Hollingsworth, P., Graham, S., & Little, D. (2011). Choosing and using a plant DNA barcode. *Plos One*, *6*(5), e19254. https://doi.org/10.1371/journal.pone.0019254

Hoveka, L., van der Bank, M., Boatwright, J., Bezeng, B., & Yessoufou, K. (2016). The noncoding trnH-psbA spacer, as an effective DNA barcode for aquatic freshwater plants, reveals prohibited invasive species in aquarium trade in South Africa. *South African Journal of Botany*, *102*, 208-216. https://doi.org/10.1016/j.sajb.2015.06.014

Jiang, K., Zhang, R., Zhang, Z., Pan, B., & Tian, B. (2020). DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage. *Plant Diversity*, *42*(5), 376-385. https://doi.org/10.1016/j.pld.2020.07.007

Jusuf, N., Sundari., & Papuangan, N. (2019). Amplification and analysis of rbcl gene (Ribulose-1,5-Bisphosphate Carboxylase) of clove in Ternate island. *IOP Science*, *276*(1), 012061. https://doi.org/10.1088/1755-1315/276/1/012061

Kang, Y., Deng, Z., Zang, R., & Long, W. (2017). DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Scientific Reports*, *7*(1), 12564. https://www.nature.com/articles/s41598-017-13057-0

KC, Y., Parajuli, A., Khatri, B. B., & Shiwakoti, L. D (2020). Phytochemicals and quality of green and black teas from different clones of tea plant. *Journal of Food Quality*, *6*, 1-13. https://doi.org/10.1155/2020/8874271

Khan, N., & Mukhtar, H. (2007). Tea polyphenols for health promotion. *Life Sciences*, *81*(7), 519-533. https://doi.org/10.1016%2Fj.lfs.2007.06.011

Khan, N., & Mukhtar, H. (2013). Tea and health: Studies in humans. *Current Pharmaceutical Design*, *19*(34), 6141-6147. https://doi.org/10.2174%2F1381612811319340008

Khanongnuch, C., Unban, K., Kanpiengjai, A., & Saenjum, C. (2017). Recent research advances and ethno-botanical history of miang , a traditional fermented tea (*Camellia sinensis* var. *assamica*) of northern Thailand. *Journal of Ethnic Foods*, *4*(3), 135-144. https://doi.org/10.1016/j.jef.2017.08.006

Koch, W. (2019). Theaflavins, Thearubigins, and Theasinensins. In J. Xiao, S. D. Sarker, & Y. Asakawa (Eds.), *Handbook of Dietary Phytochemicals* (pp. 975-1003). Springer. https://doi.org/10.1007/978-981-15-4148-3_20

Koo, M., & Cho, C. (2004). Pharmacological effects of green tea on the gastrointestinal system. *European Journal of Pharmacology*, *500*(1-3), 177-185. https://doi.org/10.1016/j.ejphar.2004.07.023

Kosińska, A., & Andlauer, W. (2014). Antioxidant capacity of tea: Effect of processing and storage. In V. Preedy (Ed.), *Processing and Impact on Antioxidants in Beverages* (pp. 109-120). Academic Press. https://doi.org/10.1016/B978-0-12-404738-9.00012-X

Kress, W., & Erickson, D. (2007). A Two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *Plos One*, *2*(6), e508. https://doi.org/ 10.1371/journal.pone.0000508

Kress, W., & Erickson, D. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences, 105*(8), 2761-2762. https://doi.org/10.1073%2Fpnas.0800476105

Kress, W., Erickson, D., Jones, F., Swenson, N., Perez, R., Sanjur, O., & Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of The National Academy of Sciences, 106*(44), 18621-18626. https://doi.org/10.1073/pnas.0909820106

Kress, W., Wurdack, K., Zimmer, E., Weigt, L., & Janzen, D. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences*, *102*(23), 8369-8374. https://doi.org/10.1073/pnas.0503123102

Lee, K., Lee, J., Sebastin, R., Shin, M., Kim, S., Cho, G. & Hyun, D. (2019). Assessment of genetic diversity of tea germ plasm for its management and sustainable use in Korea genebank. *Forests*, *10*(9), 780. https://doi.org/10.3390/f10090780

Lee, S., Wang, C., Yen, C., & Chang, C. (2016). DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. *Journal of Food and Drug Analysis, 25*(2), 260-274. https://doi.org/10.1016/j.jfda.2016.06.008

Li, M., Wong, K., Chan, W., Li, J., But, P., Cao, H., . . . Shaw, P. (2012). Establishment of DNA barcodes for the identification of the botanical sources of the Chinese‘cooling’ beverage. *Food Control, 25*(2), 758-766. https://doi.org/10.1016/j.foodcont.2011.12.008

Li, X., & Zhu, X. (2016). Tea: Types, production, and trade. *Encyclopedia of Food and Health*, *5*, 279-282. https://doi.org/10.1016/B978-0-12-384947-2.00684-X

Liang, Y., Zhang, L., & Lu, J. (2005). A study on chemical estimation of pu-erh tea quality. *Journal of the Science of Food and Agriculture, 85*(3), 381-390. https://doi.org/10.1002/jsfa.1857

Lin, Y., Juan, I., Chen, Y., Liang, Y., & Lin, J. (1996). Composition of polyphenols in fresh tea leaves and associations of their oxygen-radical-absorbing capacity with anti proliferative actions in fibroblast cells. *Journal of Agricultural and Food Chemistry*, *44*(6), 1387-1394. https://doi.org/10.1021/jf950652k

Liu, J., Moller, M., Gao, L., Zhang, D., & Li, D. (2010). DNA barcoding for the discrimination of Eurasian yews (Taxus L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources, 11*(1), 89-100. https://doi.org/10.1111/j.1755-0998.2010.02907.x

Liu, M., Li, X., Liao, B., Luo, L., & Ren, Y. (2019). Species identification of poisonous medicinal plant using DNA barcoding. *Chinese Journal of Natural Medicines*, *17*(8), 585-590. https://doi.org/10.1016/S1875-5364(19)30060-3

Liu, M., Yao, H., Luo, k., Ma, P., Zhou, P., & Liu, P. (2012). Authentication of *Illicium verum* using a DNA barcode *psbA-trnH*. *Journal of Medicinal Plants Research*, *6*(16), 3156-3161. https://doi.org/10.5897/JMPR11.1562

Liu, Z., Zeng, X., Yang, D., Chu, G., Yuan, Z., & Chen, S. (2012). Applying DNA barcodes for identification of plant species in the family *Araliaceae*. *Gene*, *499*(1), 76-80. https://doi.org/10.1016/j.gene.2012.02.016

Lv, Y., Yang, C., Shi, L., Zhang, Z., Xu, A., Zhang, L. . . . Li, H. (2020). Identification of medicinal plants within the *Apocynaceae* family using ITS2 and *psbA-trnH* barcodes. *Chinese Journal of Natural Medicines*, *18*(8), 594-605. https://doi.org/10.1016/S1875-5364(20)30071-6

Ma, H., Zhu, Z., Zhang, X., Miao, Y., & Guo, Q. (2014). Species identification of the medicinal plant *Tulipa edulis* (*Liliaceae*) by DNA barcode marker. *Biochemical Systematics and Ecology*, *55*, 362-368. https://doi.org/10.1016/j.bse.2014.03.038

Meegahakumbura, M. K., Wambulwa, M. C., Li, M. M., Thapa, K. K., Sun, Y. S., Möller, M., . . . Gao, L. M. (2018). Domestication origin and breeding history of the tea plant (*Camellia sinensis*) in China and India based on nuclear microsatellites and cpDNA sequence data. *Frontiers in Plant Science*, *8*, 2270. https://doi.org/10.3389%2Ffpls.2017.02270

Meng, X., Li, N., Zhu, H., Wang, D., Yang, C., & Zhang, Y. (2018). Plant resources, chemical constituents, and bioactivities of tea plants from the genus Camellia s ection Thea. *Journal of Agricultural and Food Chemistry*, *67*(19), 5318-5349. https://doi.org/10.1021/acs.jafc.8b05037

Merrill, K., Coleman, C., Meyer, S., Leger, E., & Collins, K. (2016). Development of Single nucleotide polymorphism markers for *Bromus tectorum* (Poaceae) from a partially sequenced transcriptome. *Applications in Plant Sciences*, *4*(11), 1600068. https://doi.org/10.3732/apps.1600068

Mok, P., Chang, R., Wang, M., & So, K. F. (2008). A review on the laboratory investigations and epidemiological studies of black and pu-Erh Tea. In C. T. Ho, J. E. Simmon, F. Shahidi & Y. Shao (Eds.), *Dietary Supplements* (pp. 144-159). Oxford University Press. https://doi.org/10.1021/bk-2008-0987.ch009

Mulo, P., Sakurai, I., & Aro, E. (2012). Strategies for psbA gene expression in cyanobacteria, green algae and higher plants: From transcription to PSII repair. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics*, *1817*(1), 247-257. https://doi.org/10.1016/j.bbabio.2011.04.011

Nagini, S., & Senthil-Murugan, R. (2013). Cancer chemoprevention by black tea polyphenols. In V. Preedy (Ed.), *Tea in Health and Disease Prevention* (pp.737-750). https://doi.org/10.1016/B978-0-12-384937-3.00062-8

Newmaster, S. G., & Ragupathy, S. (2009). Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). *Molecular Ecology Resources*, *9*, 172-180. https://doi.org/10.1111/j1755-0998.2009.02642.x

Park, I., Kim, W. J., Yeo, S. M., Choi, G., Kang, Y. M., Piao, R., . . . Moon, B. C. (2017). The complete chloroplast genome sequences of Fritillaria ussuriensis Maxim. and Fritillaria cirrhosa D. Don, and comparative analysis with other Fritillaria species. *Molecules*, *22*(6), 982. https://doi.org/10.3390molecules22060982

Parveen, I., Singh, H., Malik, S., Raghuvanshi, S., & Babbar, S. (2017). Evaluating five different loci (*rbcL, rpoB, rpoC1, matK,* and ITS) for DNA barcoding of Indian orchids. *Genome*, *60*(8), 665-671. https://doi.org/10.1139/gen-2016-0215

Patil, S., Rane, N., Adsul, A., Gholave, A., Yadav, S., Jadhav, J., & Govindwar, S. (2016). Study of molecular genetic diversity and evolutionary history of medicinally important endangered genus Chlorophytum using DNA barcodes. *Biochemical Systematics and Ecology, 65*, 245-252. https://doi.org/10.1016/j.bse.2016.02.035

Rahman, M., & Ali, R. (2020). Application of rbcL in DNA barcoding of Phaleria macrocarpa (Scheff.) Boerl. and its processed commercial products. *Materials Today: Proceedings*, *31*, A58-A62. https://doi.org/10.1016/j.matpr.2020.12.131

Ramesh, G., Mathew, D., John, K., & Ravisankar, V. (2022). Chloroplast gene matK holds the barcodes for identification of Momordica (Cucurbitaceae) species from Indian subcontinent. *Horticultural Plant Journal*, *8*(1), 89-98. https://doi.org/10.1016/j.hpj.2021.04.001

Raveendar, S., Lee, J., Shim, D., Lee, G., Jeon, Y., Cho, G., . . . Chung, J. (2017). Comparative efficacy of four candidate DNA barcode regions for identification of Vicia species. *Plant Genetic Resources*, *15*(4), 286-295. http://doi.org/10.1017/S1479262115000623

Reha, L., Pallab, K., Prosenjit, B., Arnab, S., & Malay, B. (2021). Evolution of matK gene among the elite tea clones (Camellia sinensis) revealed by nucleotide substitution within the consensus region. *Journal of Applied Biology & Amp; Biotechnology*, *9*(1), 32-40. https://doi.org/10.7324/JABB.2021.9105

Rubel Mozumder, N., Lee, Y., Hwang, K., Lee, M., Kim, E., & Hong, Y. (2020). Characterization of tea leaf metabolites dependent on tea (Camellia sinensis) plant age through 1H NMR-based metabolomics. *Applied Biological Chemistry*, *63*(1), 10. https://doi.org/10.1186/s13765-020-0492-7

Schlichting, C. (1986). The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics*, *17*(1), 667-693. https://doi.org/10.1146/annurev.es17.110186.003315

Shao, W. F., Powell, C., & Clifford, M. N. (1995). The analysis by HPLC of green, black and pu'er teas produced in Yunnan. *Journal of the Science of Food and Agriculture*, *69*, 535-540. https://doi.org/10.1002/jsfa.2740690419

Sharma, R., Verma, S., & Kumar, D. (2021). Polyphenolics and therapeutic insights in different tissues extract and fractions of Camellia sinensis (L.) Kuntze (Kangra Tea).*Food Bioscience*, *42*, 101164. https:// doi.org/10.1016/j.fbio.2021.101164

Silk Road Teas. (2022). *Featured teas*. https://www.silkroadteas.com/

Soares, S., Grazina, L., Costa, J., Amaral, J., Oliveira, M., & Mafra, I. (2018). Botanical authentication of lavender (Lavandula spp.) honey by a novel DNA-barcoding approach coupled to high resolution melting analysis. *Food Control*, *86*, 367-373. https://doi.org/10.1016/j.foodcont.2017.11.046

Soni, R., Katoch, M., Kumar, A., Ladohiya, R., & Verma, P. (2015). Tea: Production, composition, consumption and its potential as an antioxidant and antimicrobial agent. *International Journal of Food and Fermentation Technology*, *5*(2), 95-106. https://doi.org/10.5958/2277-9396.2016.00002.7

Stoeckle, M., Gamble, C., Kirpekar, R., Young, G., Ahmed, S., & Little, D. (2011). Commercial teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports*, *1*(1), 42. https://doi.org/10.1038/ srep00042

Sun, W., Li, J., Xiong, C., Zhao, B., & Chen, S. (2016). The potential power of bar-HRM technology in herbal medicine identification. *Frontiers in Plant Science*, *7*, 367. https://doi.org/10.3389%2Ffpls.2016.00367

Sun, X., Zhu, Y., Guo, J., Peng, B., Bai, M., & Hang, Y. (2012). DNA barcoding the dioscorea in China, a vital group in the evolution of monocotyledon: use of matK gene for species discrimination. *Plos One*, *7*(2), e32057. https://doi.org/10.1371/journal.pone.0032057

Swetha, V., Parvathy, V., Sheeja, T., & Sasikumar, B. (2017). Authentication of *Myristica fragrans* Houtt. using DNA barcoding. *Food Control, 73*, 1010-1015. https://doi.org/10.1016/j.foodcont.2016.10.004

Tang, S., Liu, Y., Zheng, N., Li, Y., Ma, Q., Xiao, H., . . . Wu, L. (2020). Temporal variation in nutrient requirements of tea (Camellia sinensis) in China based on QUEFTS analysis. *Scientific Reports*, *10*(1), 1745. https://doi.org/10.1038/s41598-020-57809-x

Tang, W., Li, S., Liu, Y., Huang, M., & Ho, C. (2013). Anti-diabetic activity of chemically profiled green tea and black tea extracts in a type 2 diabetes mice model via different mechanisms. *Journal of Functional Foods*, *5*(4), 1784-1793. https://doi.org/10.1016/j.jff. 2013.08.007

Techen, N., Parveen, I., Pan, Z., & Khan, I. (2014). DNA barcoding of medicinal plant material for identification. *Current Opinion in Biotechnology*, *25*, 103-110. https://doi.org/10.1016/j.copbio.2013.09.010

Thansettakij. (2022). *Did you know? good tea is in Chiang Rai. Commerce pushes Thai tea to the world market with FTA.* https:/www.thansettakij.com/economy/514710

Theppakorn, T., Luthfiyyah, A., & Ploysri, K. (2014). Comparison of the composition and antioxidant capacities of green teas produced from Assam and the Chinese varieties cultivated Thailand. *Journal of Microbiology, Biotechnology and Food Sciences, 3*(5), 364-370. https://office2.jmbfs.org/index.php/JMBFS/article/view/6996

Thongkhao, K., Tungphatthong, C., Phadungcharoen, T., & Sukrong, S. (2020). The use of plant DNA barcoding coupled with HRM analysis to differentiate edible vegetables from poisonous plants for food safety. *Food Control*, *109*, 106896. https://doi.org/10.1016/j.foodcont.2019.106896

Tsuneki, H., Ishizuka, M., Terasawa, M., Wu, J., Sasaoka, T., & Kimura, I. (2004). Effect of green tea on blood glucose levels and serum proteomic patterns in diabetic (db/db) mice and on glucose metabolism in healthy humans. *BMC Pharmacology*, *4*(1), 18-21. https://doi.org/10.1186/1471-2210-4-18

Veldman, S., Ju, Y., Otieno, J., Abihudi, S., Posthouwer, C., Gravendeel, B., . . . De Boer, H. J. (2020). DNA barcoding augments conventional methods for identification of medicinal plant species traded at Tanzanian markets. *Journal of Ethnopharmacology*, *250*, 112495. https://doi.org/10.1016/j.jep.2019.112495

Vuong, Q., Bowyer, M., & Roach, P. (2011). L-Theanine: Properties, synthesis and isolation from tea. *Journal of the Science of Food and Agriculture*, *91*(11), 1931-1939. https://doi.org/10.1002/jsfa.4373

Wallace, L., Boilard, S., Eagle, S., Spall, J., Shokralla, S., & Hajibabaei, M. (2012). DNA barcodes for everyday life: Routine authentication of natural health products. *Food Research International*, *49*(1), 446-452. https://doi.org/10.1016/j.foodres.2012.07.048

Wang, X. C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J. H., . . . Li, J. Q. (2015). ITS1: a DNA barcode better than ITS2 in eukaryotes?. *Molecular Ecology Resources*, *15*(3), 573-586. https://doi.org/10.1111/1755-0998.12325

Wasser, S., Mailand, C., Booth, R., Mutayoba, B., Kisamo, E., Clark, B., . . . Stephens, M. (2007). Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences*, *104*(10), 4228-4233. https://doi.org/10.1073/pnas.0609714104

Whiting, D. C. (2017). *Plant Structures: Leaves*. http://www.ext.colostate.edu/mg/Gardennotes/134.html#function

Worthy, S. J., Bucalo, K., Perry, E., Reynolds, A., Cruse-Sanders, J., Pérez, Á., & Burgess, K. (2022). Ability of *rbcL* and *matK* DNA barcodes to discriminate between montane forest orchids. *Plant Systematics and Evolution, 308*, 19. https://doi.org/10.1007/s00606-022-01809-z

Wu, C., & Wei, G. (2009). Tea as a functional food for oral health. In M. Wilson (Ed.), *Food Constituents and Oral Health* (pp. 396-417). Woodhead Publishing. https://doi.org/10.1533/9781845696290.2.396

Xie, L., Wang, Y. W., Guan, S. Y., Xie, L. J., Long, X., & Sun, C. Y. (2014). Prospects and problems for identification of poisonous plants in China using DNA barcodes. *Biomedical and Environmental Sciences : BES, 27*(10), 794-806. https://doi.org/10.3967/bes2014.115

Yamashiro, T., Fukuda, T., Yokoyama, J., & Maki, M. (2004). Molecular phylogeny of Vincetoxicum (Apocynaceae-Asclepiadoideae) based on the nucleotide sequences of cpDNA and nrDNA. *Molecular Phylogenetics and Evolution, 31*(2), 689-700. https://doi.org/10.1016/j.ympev.2003.08.016

Yan, Z., Zhong, Y., Duan, Y., Chen, Q., & Li, F. (2020). Antioxidant mechanism of tea polyphenols and its impact on health benefits. *Animal Nutrition, 6*(2), 115-123. https://doi.org/10.1016/j.aninu.2020.01.001

Yang, C., Maliakal, P., & Meng, X. (2002). Inhibition of Carcinogenesis by Tea. *Annual Review of Pharmacology and Toxicology, 42*, 25-54. https://doi.org/10.1146/annurev.pharmtox.42.082101.154309

Ye, N. (2012). A Minireview of analytical methods for the geographical origin analysis of teas (*Camellia sinensis*). *Critical Reviews in Food Science and Nutrition, 52*(9), 775-780. https://doi.org/10.1080/10408398.2010.508568

Yonghua, L., Jinlan, R., Shilin, C., Jingyuan, S., Kun, L., Dong, L., & Hui, Y. (2010). Authentication of Taxillus chinensis using DNA barcoding technique. *Journal of Medicinal Plants Research, 4*(24), 2706-2709. http://doi.org/10.5897/JMPR09.677

Yoon, C. K. (1993). Forensic science. Botanical witness for the prosecution. *Science. 260*(5110), 894-895. https://doi.org/10.1126/science.8493521

Yu, J., Xue, J., & Zhou, S. (2011). New universal matK primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution, 49*(3), 176-181. http://doi.org/10.1111/j.1759-6831.2011.00134.x

Zhao, D., Parnell, J., & Hodkinson, T. (2017). Typification of names in the genus Camellia (Theaceae). *Phytotaxa*, *292*(2), 171-179. https://doi.org/10.11646/phytotaxa.292.2.4

Zhao, L., Yu, X., Shen, J., & Xu, X. (2018). Identification of three kinds of Plumeria flowers by DNA barcoding and HPLC specific chromatogram. *Journal of Pharmaceutical Analysis*, *8*(3), 176-180. https://doi.org/10.1016/j.jpha.2018.02.002

APPENDIX

APPENDIX A

FIGURES OF WILD TEA PLANT

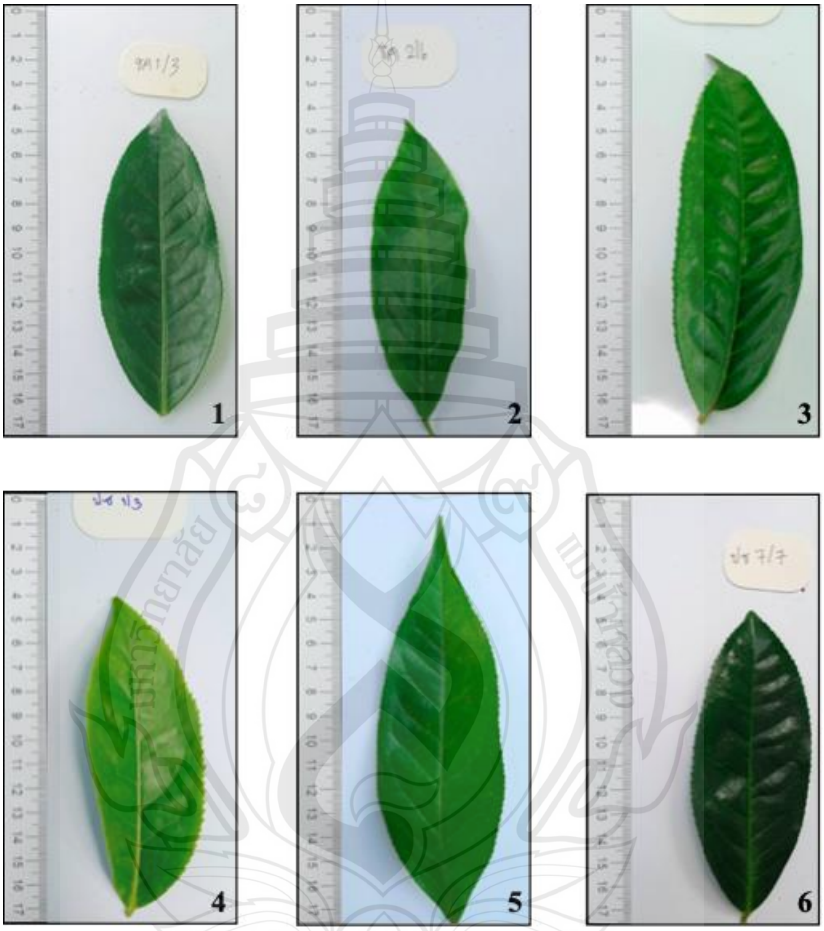


Figure A1 Picture of Tea Plant for TCR1 (1), TCR6 (2), TCR13 (3), TCR18 (4), TCR20 (5), and TCR30 (6)

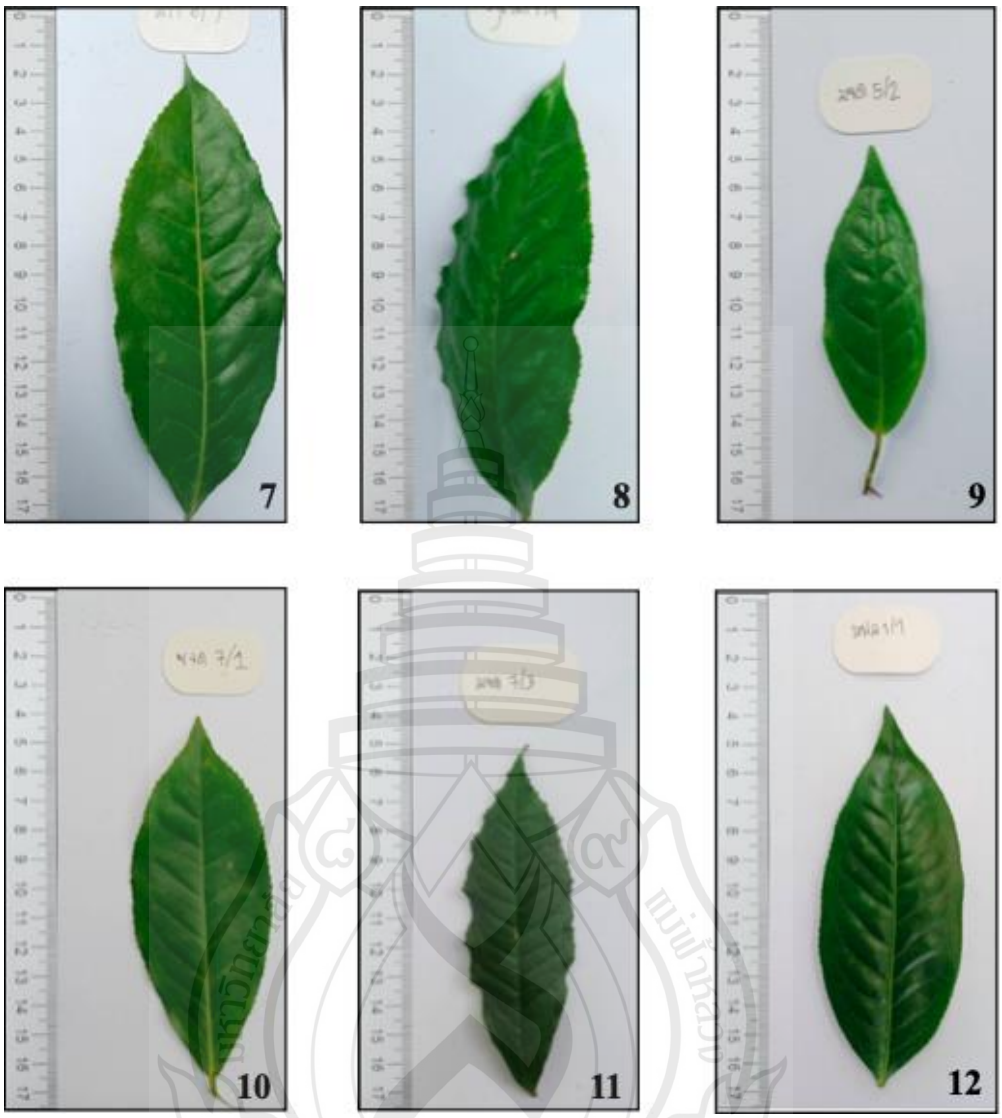


Figure A2 Picture of Tea Plant for TCR39 (7), TCR41 (8),TCR42 (9), TCR44 (10), TCR46 (11), and TCR47 (12)

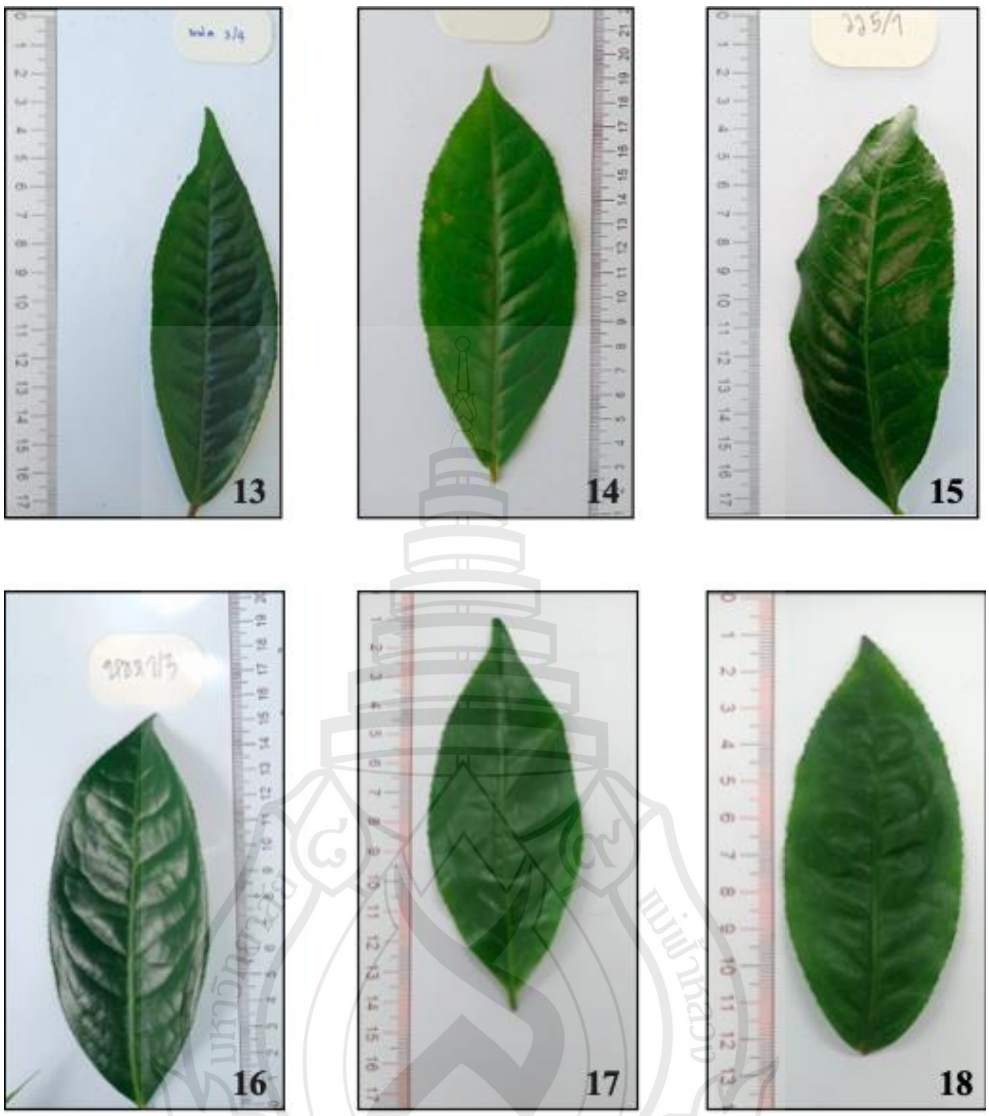


Figure A3 Picture of Tea Plant for TCR50 (13), TCR56 (14), TCR63 (15), TCR74 (16), TCM90 (17), and TCM98 (18)

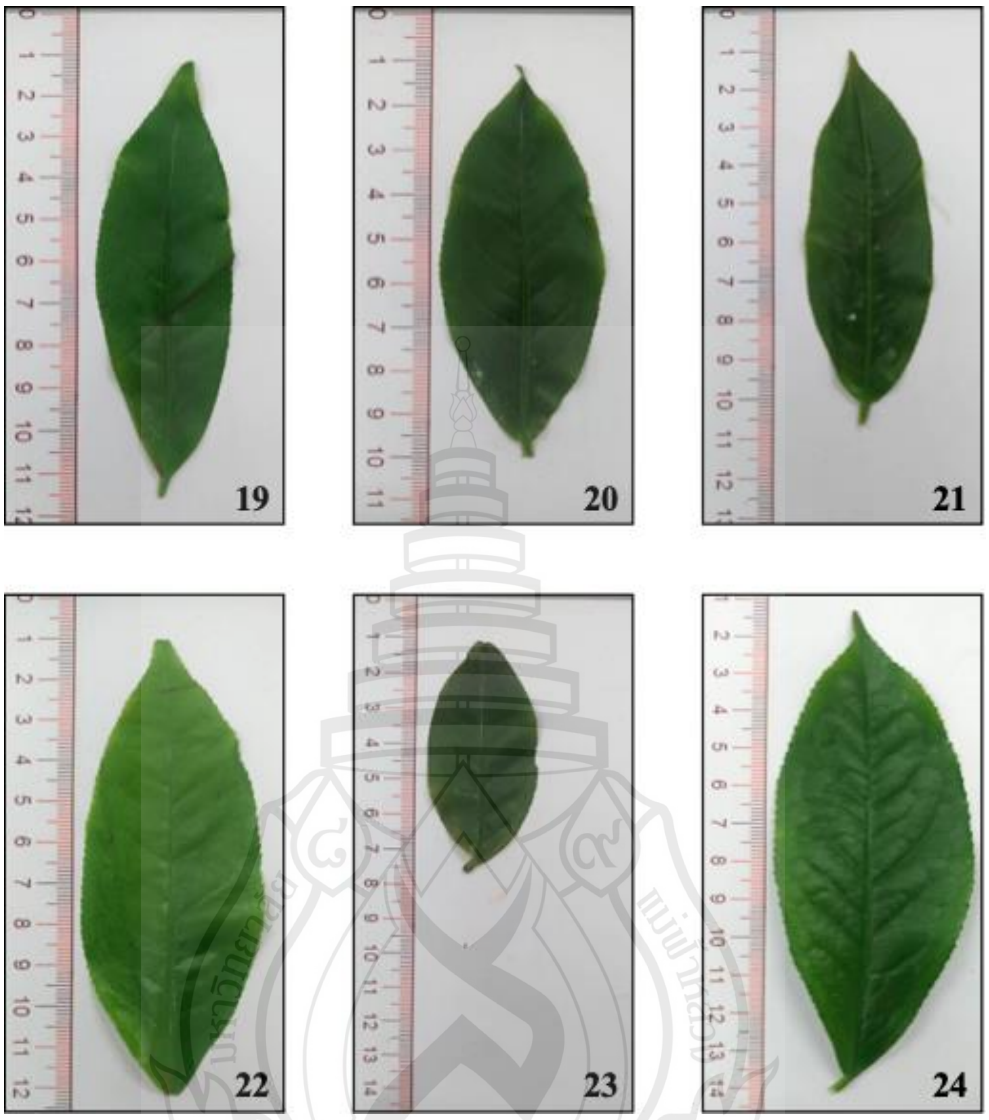


Figure A4 Picture of Tea Plant for TCM99 (19) , TCM100 (20) , TCM102 (21),TCM104 (22), TCM106 (23), and TCM107 (24)

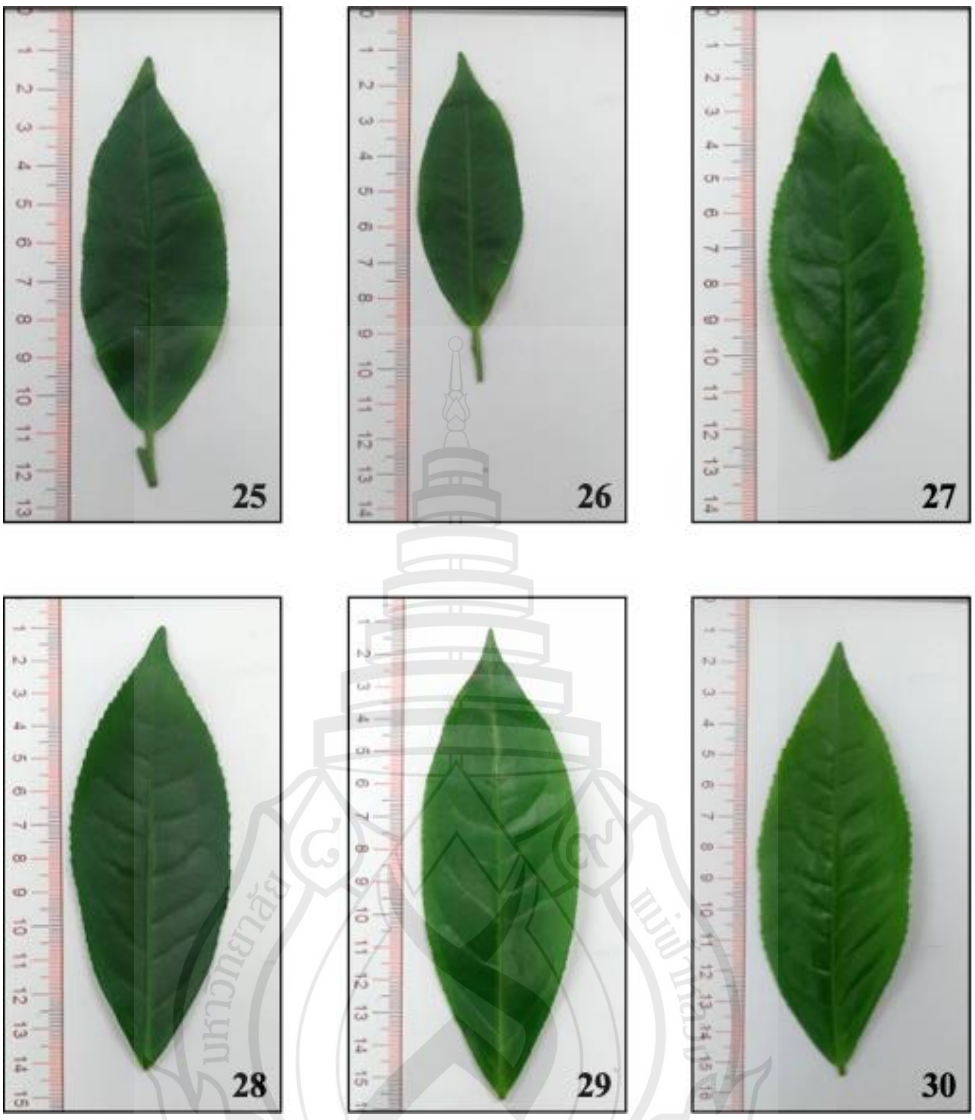


Figure A5 Picture of Tea Plant for TCM108 (25), TCM110 (26), TCM113 (27), TCM117 (28), TCM118 (29), and TCM123 (30)

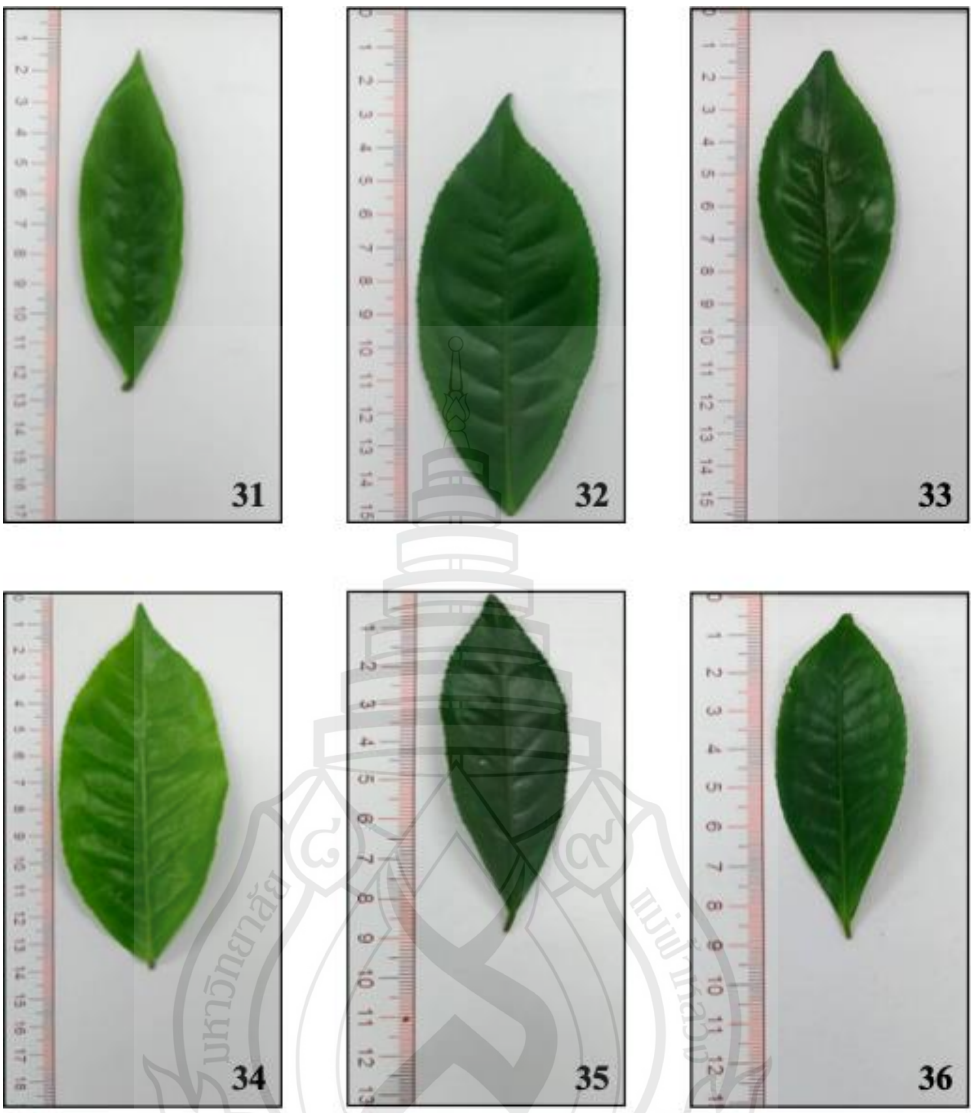


Figure A6 Picture of Tea Plants for TCM125 (31), TCM126 (32), TCM128.1 (33), TCM129 (34), TCM133 (35), and TCM139 (36)

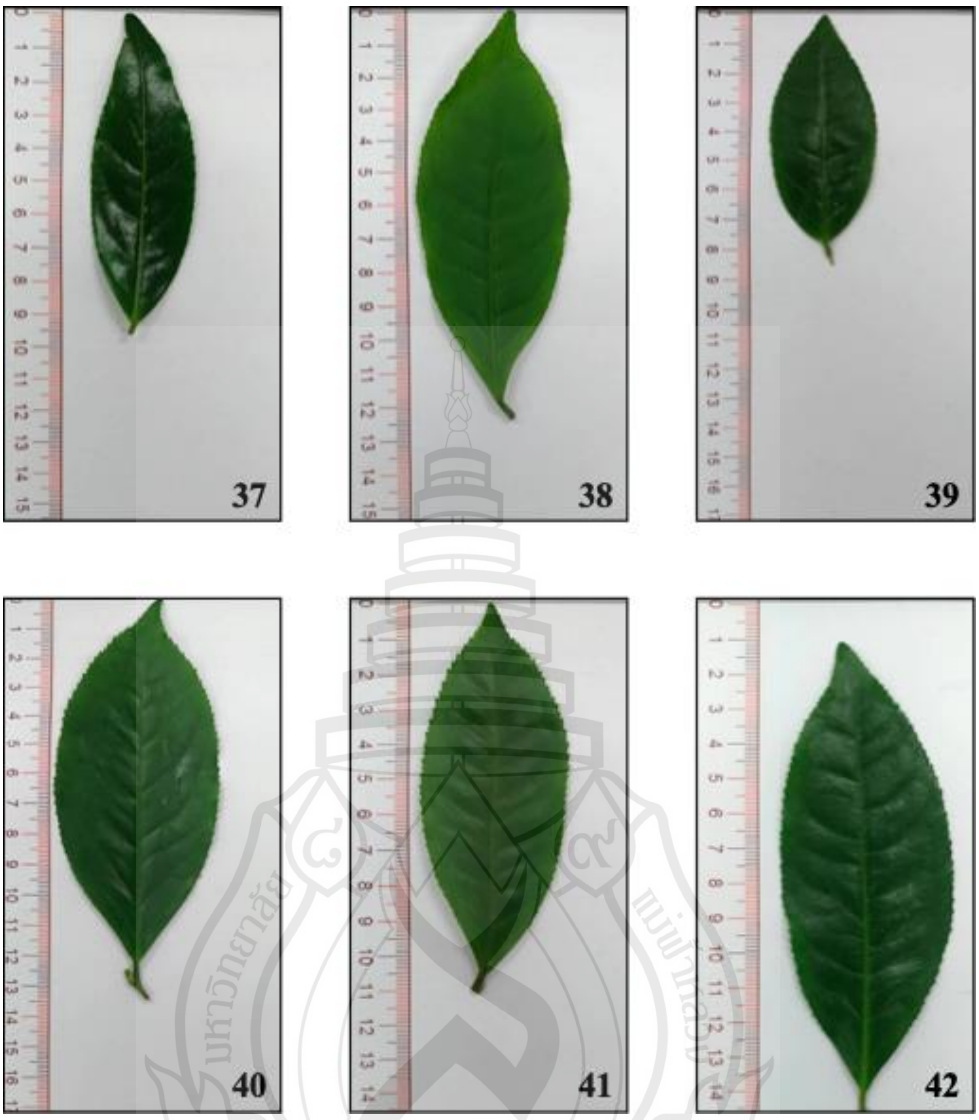


Figure A7 Picture of Tea Plants for TCM141 (37) , TCM142 (38) , TCM143 (39),TCM148 (40), TCM150 (41), and TCM154 (42)

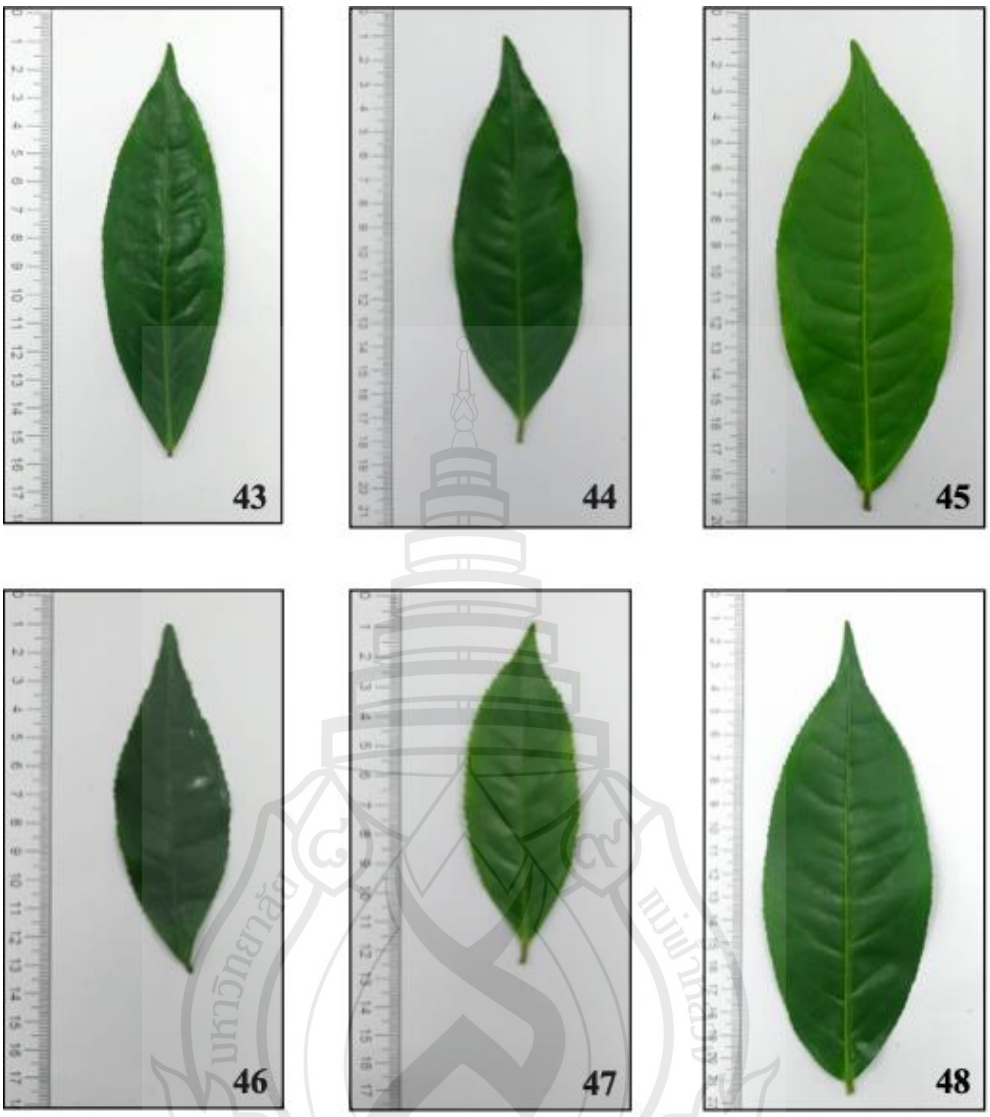


Figure A8 Picture of Tea Plants, TCM156 (43), TCM157 (44), TCM158 (45), TCM160 (46), TCM162 (47), and TCM163 (48)

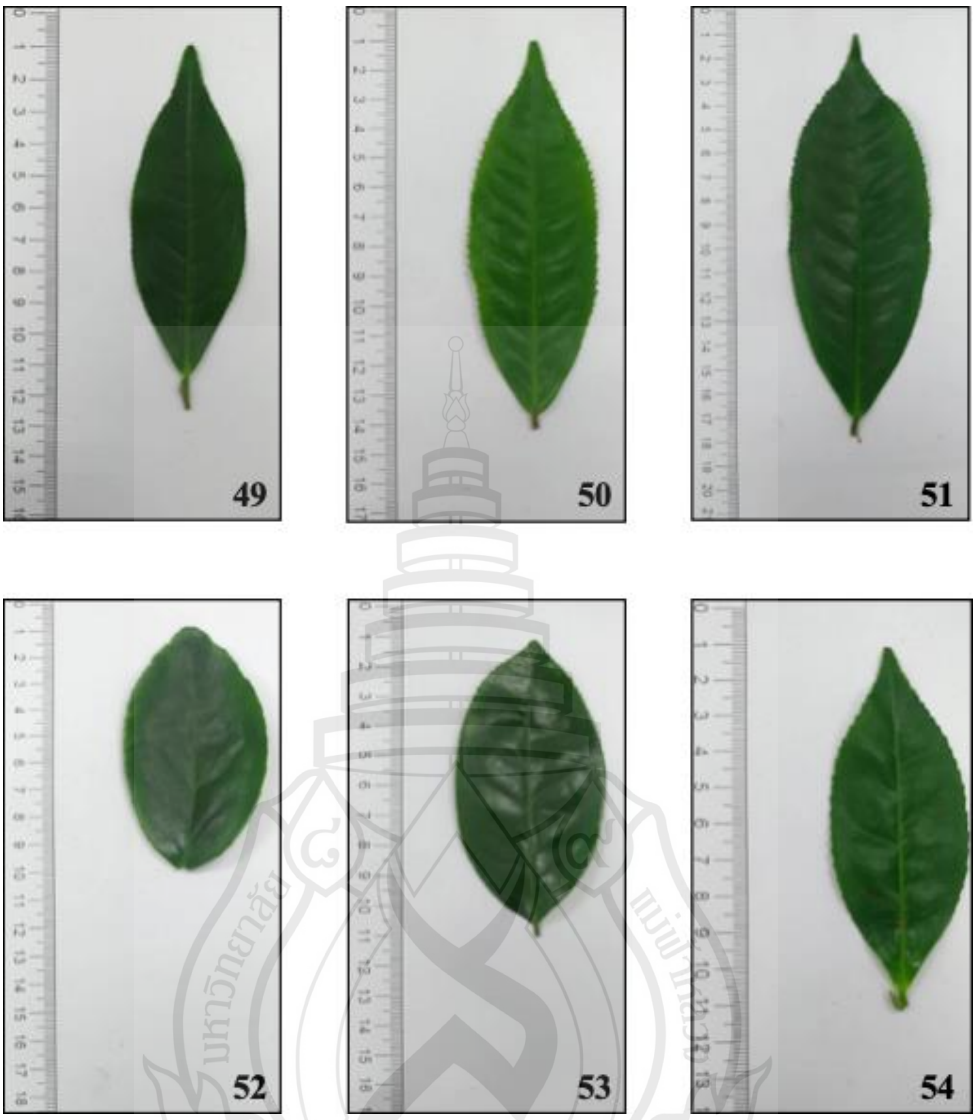


Figure A9 Picture of Tea Plants, TCM165 (49), TCM166 (50), TCM167 (51), TCM170 (52), TCM172 (53) and TCM175 (54)

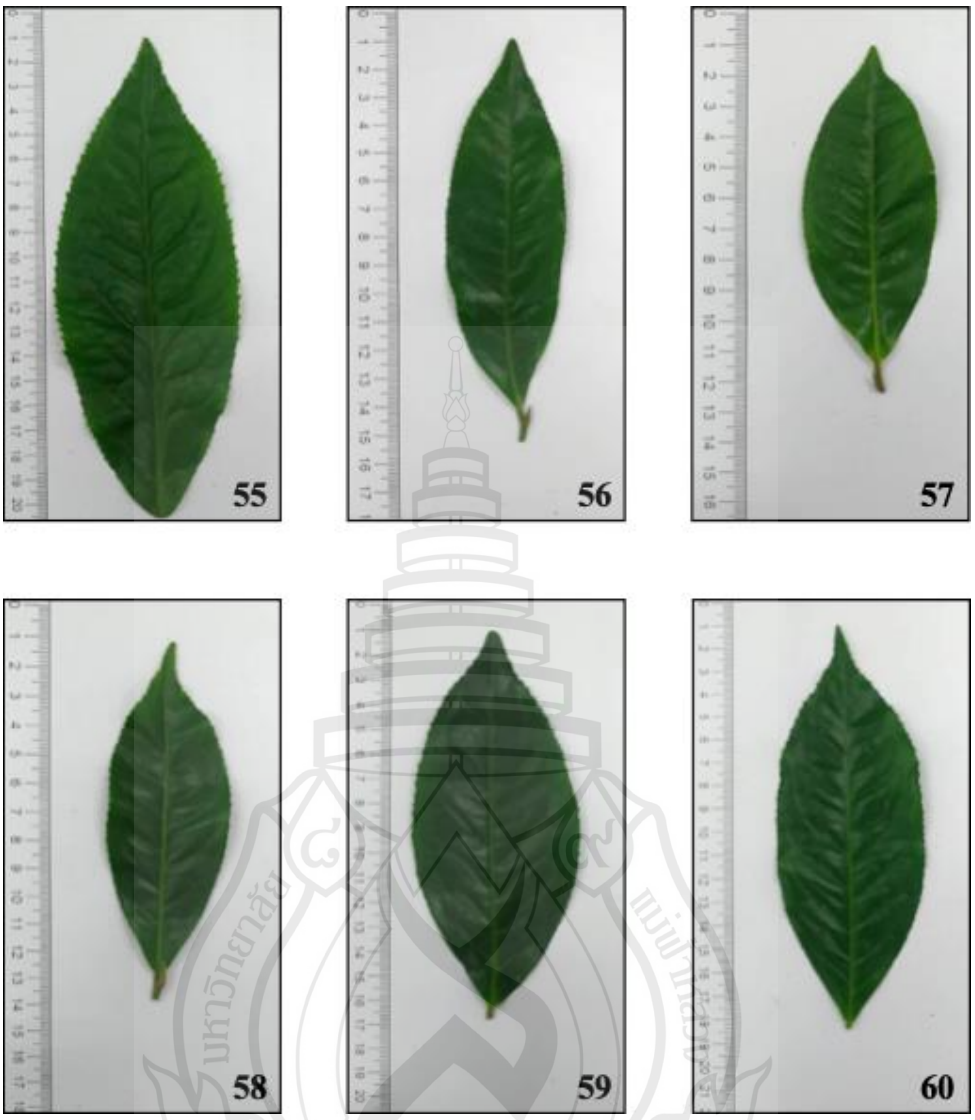


Figure A10 Picture of Tea Plants, TCM176 (55), TCM177 (56), TCM178 (57), TCM180 (58), TCM182 (59), and TCM183 (60)

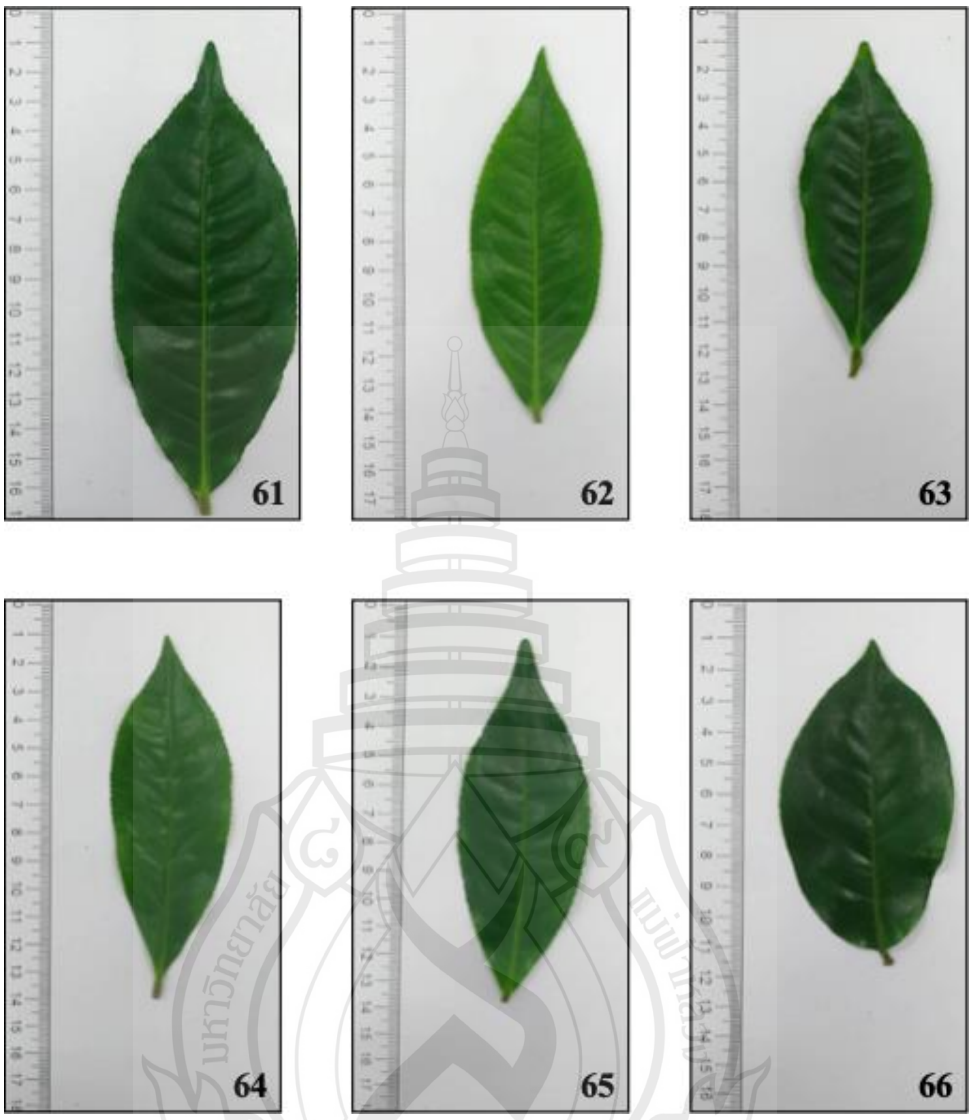


Figure A11 Picture of Tea Plants, TCM184 (61), TCM185 (62), TCM186 (63), TCM189 (64), TCM190 (65), and TCM193 (66)

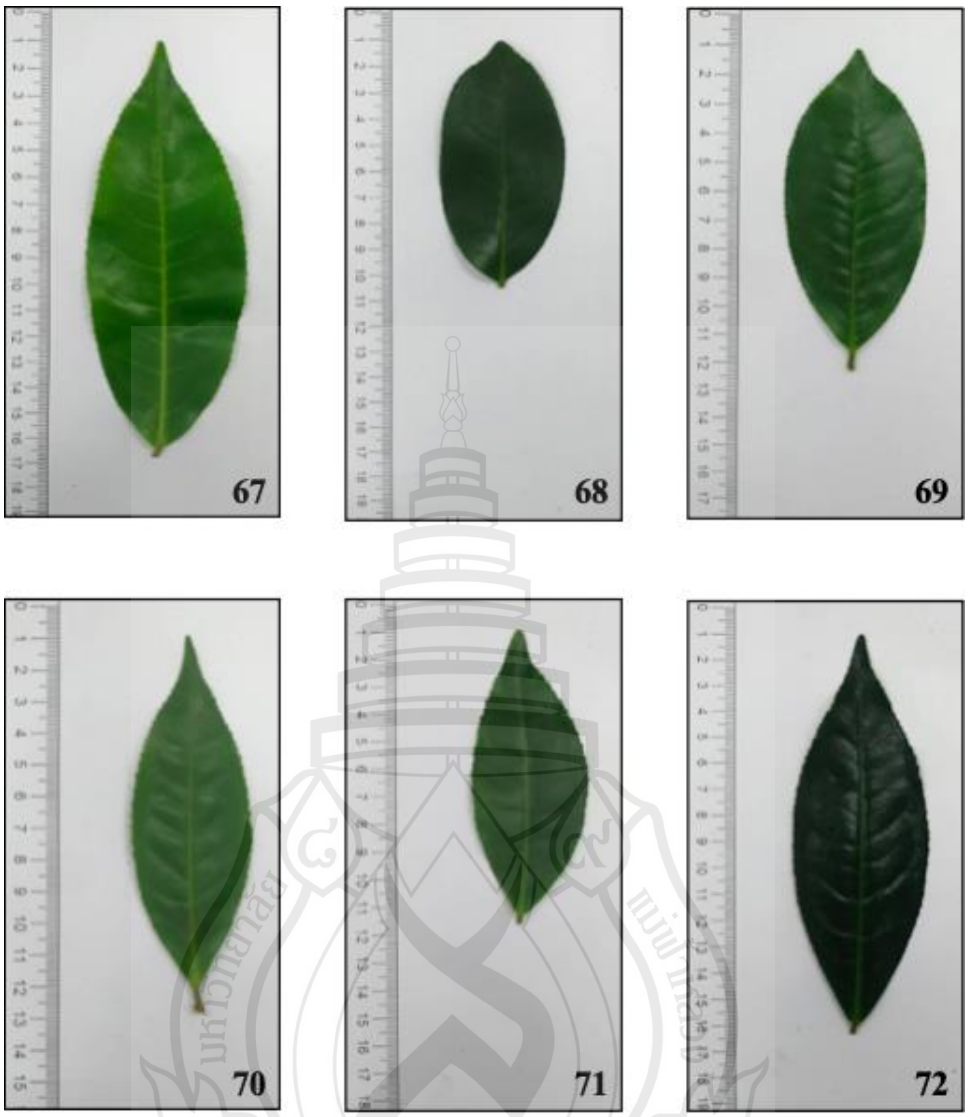


Figure A12 Picture of Tea Plants, TCM194 (67), TCM196 (68), TCM202 (69), TCM203 (70), TCM204 (71), and TCM205 (72)

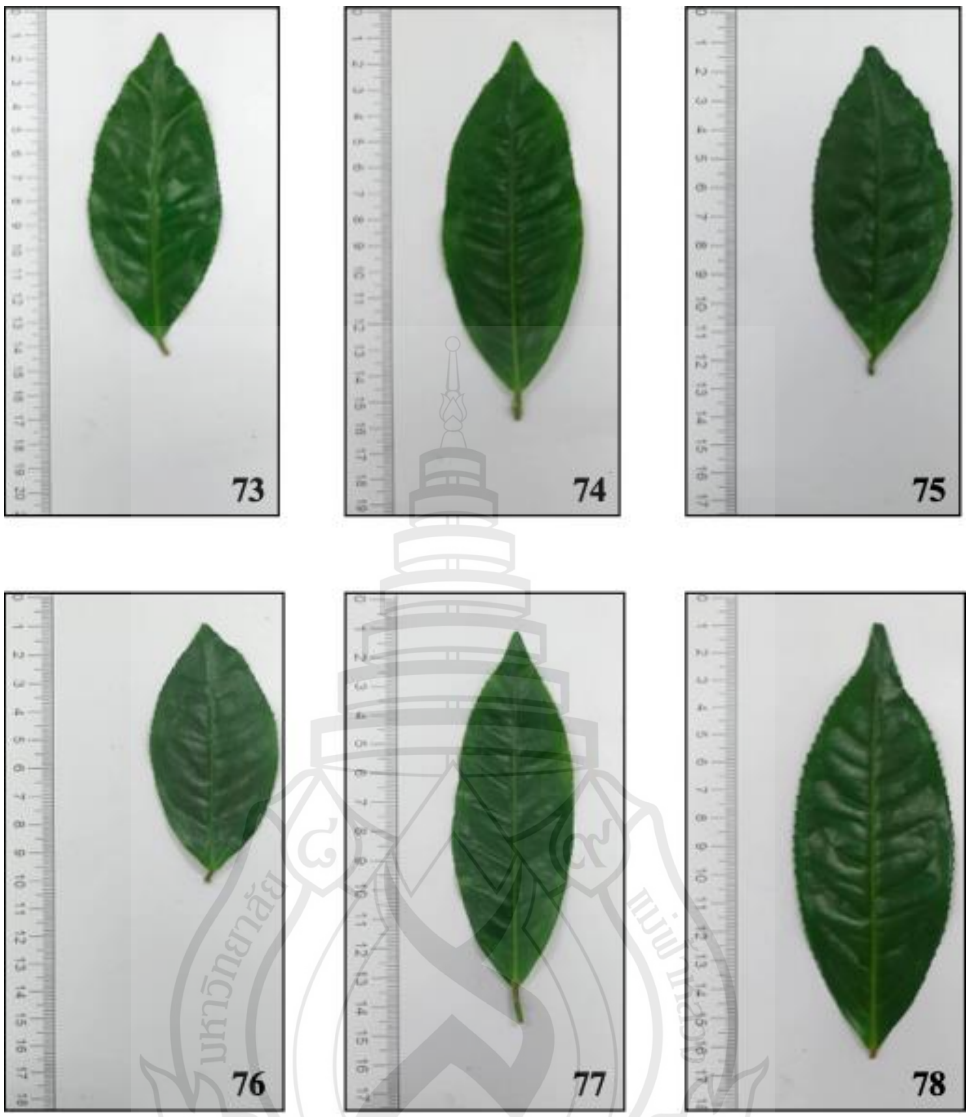


Figure A13 Picture of Tea Plants, TCM207 (73), TCM208 (74), TCM212 (75), TCM215 (76), TCM216 (77), and TCM217 (78)

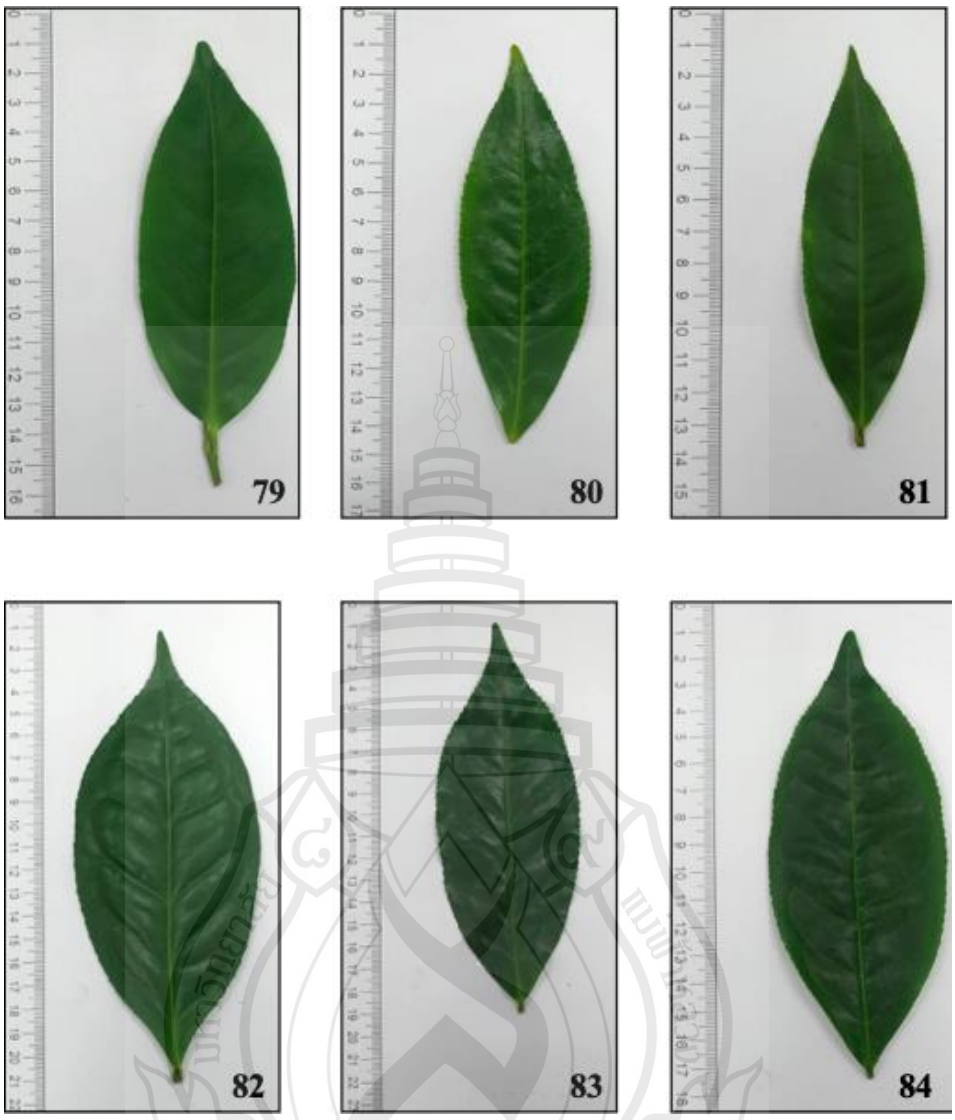


Figure A14 Picture of Tea Plants, TCM219 (79), TCM222 (80), TCM225 (81), TCM227 (82), TCM228 (83), and TCM229 (84)

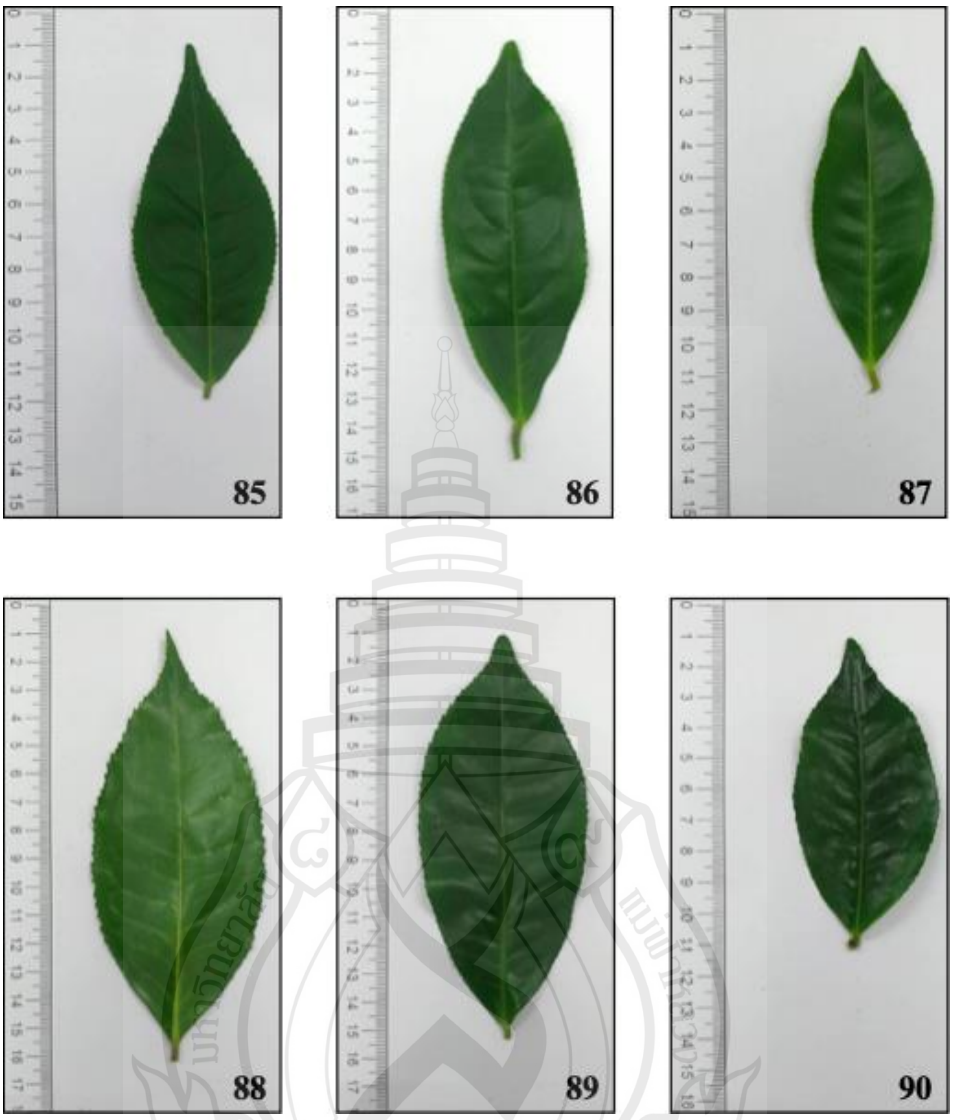


Figure A15 Picture of Tea Plants, TCM231 (85), TCM235 (86), TCM236 (87), TCM237 (88), TCM238 (89), and TCM239 (90)

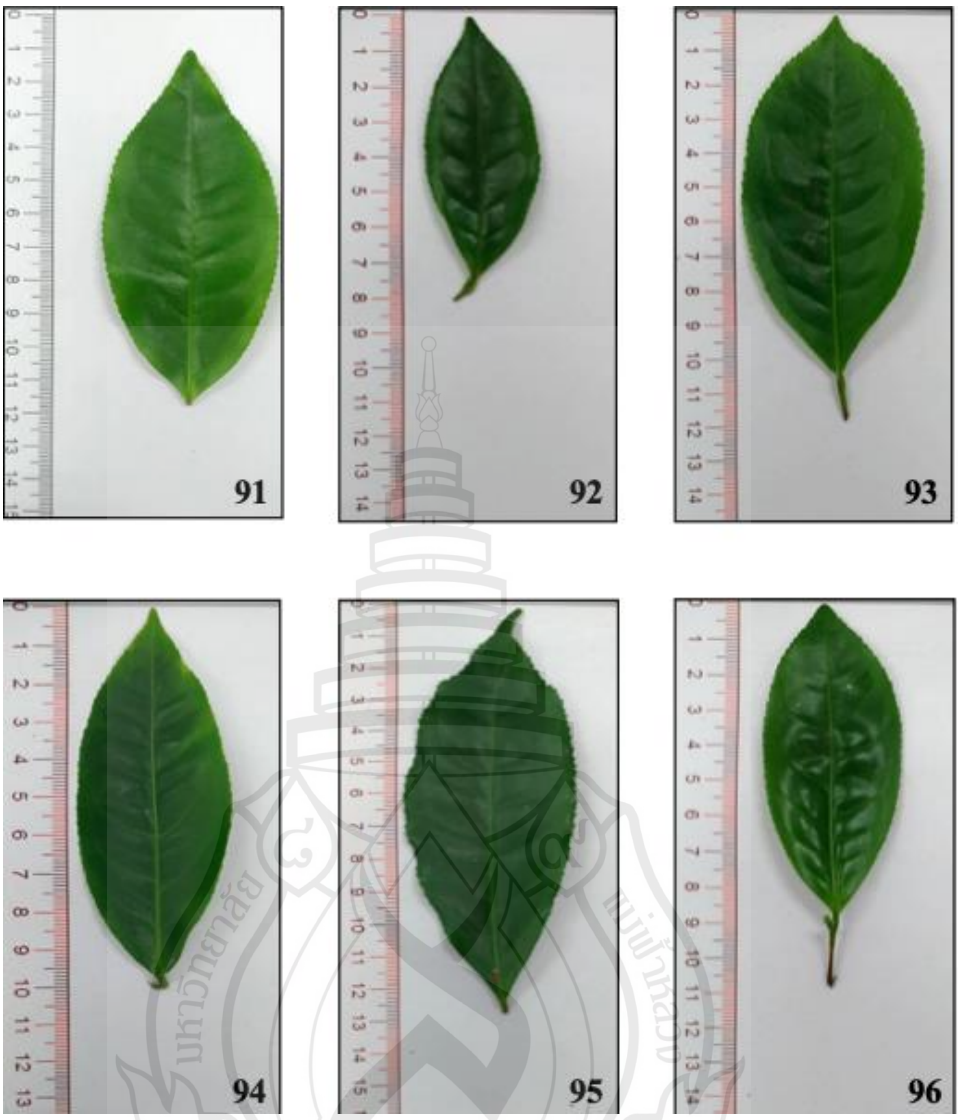


Figure A16 Picture of Tea Plants, TCM241 (91), TN242 (92), TN247 (93), TN249 (94), TN250 (95), and TN251 (96)

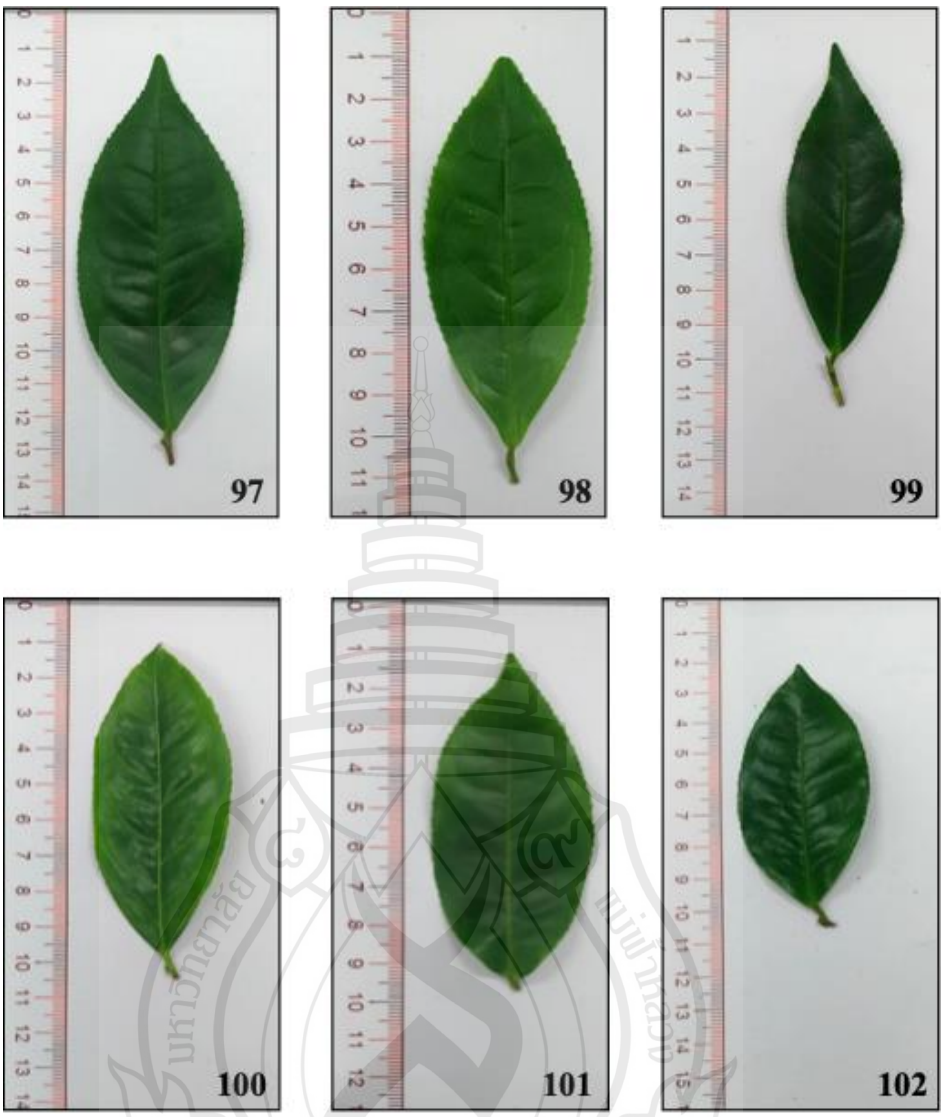


Figure A17 Picture of Tea Plants, TN252 (97), TN253 (98), TN254 (99), TN256 (100), TN262 (101), and TN268 (102)

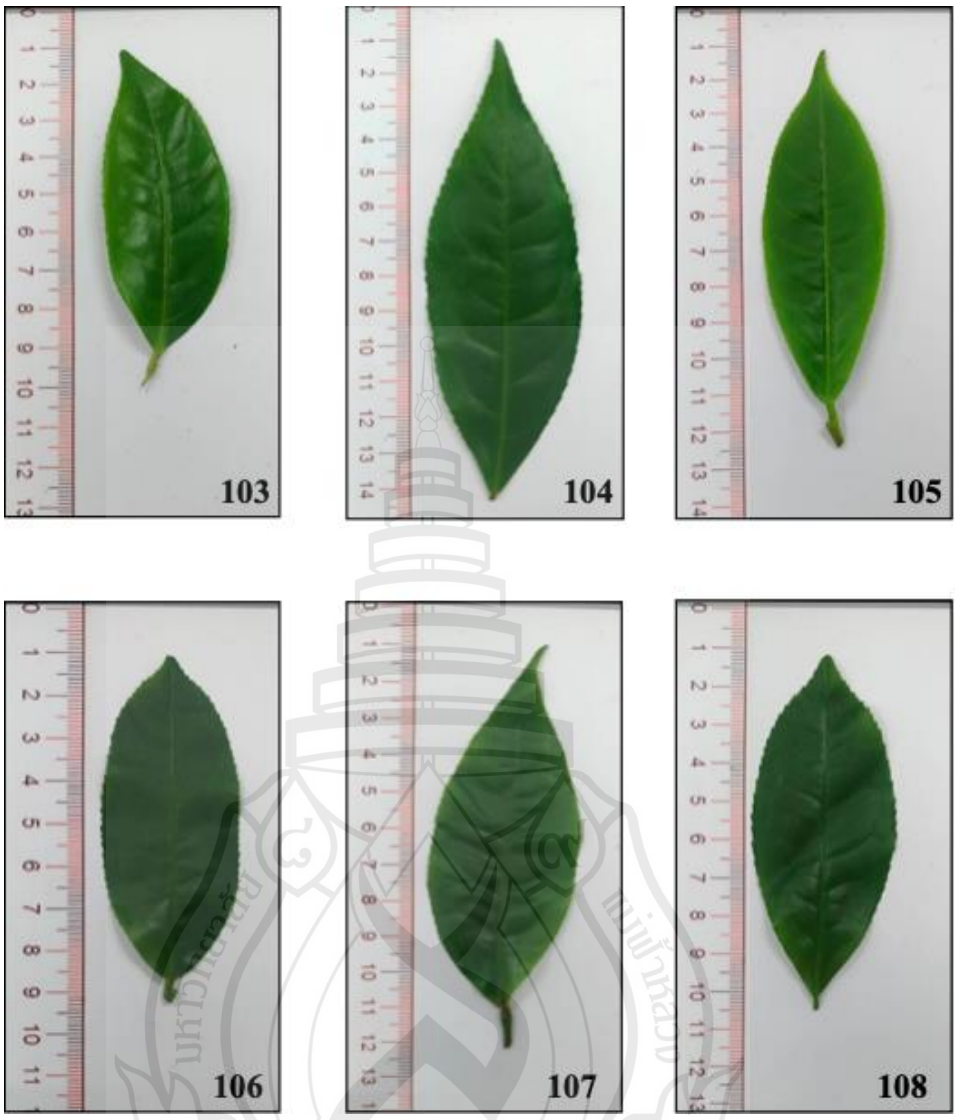


Figure A18 Picture of Tea Plants, TN273 (103), TN278 (104), TN280 (105), PPD281 (106), PPD282 (107), and PPD285 (108)

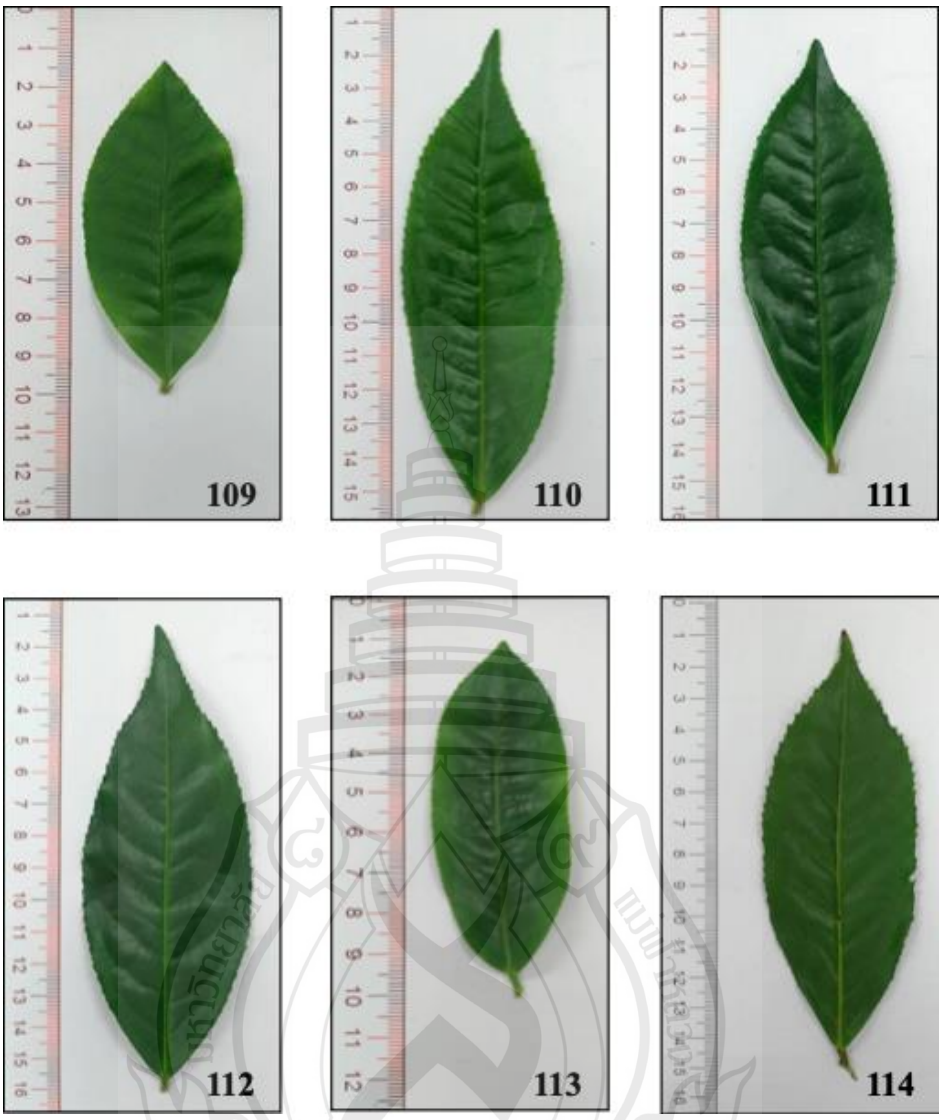


Figure A19 Picture of Tea Plants, PPD290 (109), TMS295 (110), TMS296 (111), TMS297 (112), TMS302 (113), and TMS306 (114)