DNA BARCODING COUPLED WITH HIGH RESOLUTION MELTING (BAR-HRM)

ANALYSIS FOR DIFFERENTIATION WHITE KWAO KHRUEA (*PUERARIA CANDOLLEI*),

BLACK KWAO KHRUEA (*MUCUNA MACROCARPA*), AND RED KWAO KHRUEA

(*BUTEA SUPERBA*)

Miss Nonthapon Vejpanich

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Pharmaceutical Sciences and Technology

Common Course

FACULTY OF PHARMACEUTICAL SCIENCES

Chulalongkorn University

Academic Year 2019

การใช้แถบรหัสดีเอ็นเอร่วมกับการวิเคราะห์การหลอมเหลวความละเอียดสูง (BAR-HRM) เพื่อ
แยกแยะระหว่างกวาวเครือขาว กวาวเครือดำ และกวาวเครือแดง

น.ส.นนทพร เวชพาณิชย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเภสัชศาสตร์และเทคโนโลยี ไม่สังกัดภาควิชา/เทียบเท่า
คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Thesis Title · · · · · · · · · · DNA BARCODING COUPLED WITH HIGH RESOLUTION
MELTING (BAR-HRM) ANALYSIS FOR DIFFERENTIATION
WHITE KWAO KHRUEA (*PUERARIA CANDOLLEI*),
BLACK KWAO KHRUEA (*MUCUNA MACROCARPA*),
AND RED KWAO KHRUEA (*BUTEA SUPERBA*)

By · · · · · · · · · · Miss Nonthapon Vejpanich

Field of Study · · · · · · · · · · Pharmaceutical Sciences and Technology

Thesis Advisor · · · · · · · · · · Associate Professor SUCHADA SUKRONG, Ph.D.

Thesis Co Advisor · · · · · · · · · · Assistant Professor TAKSINA CHUANASA, Ph.D.

Accepted by the FACULTY OF PHARMACEUTICAL SCIENCES, Chulalongkorn
University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the FACULTY OF
PHARMACEUTICAL SCIENCES

(Assistant Professor RUNGPETCH SAKULBUMRUNGSIL,
Ph.D.)

THESIS COMMITTEE

.... Chairman

(Associate Professor SORNKANOK VIMOLMANGKANG,
Ph.D.)

.... Thesis Advisor

(Associate Professor SUCHADA SUKRONG, Ph.D.)

.... Thesis Co-Advisor

(Assistant Professor TAKSINA CHUANASA, Ph.D.)

.... Examiner

(NATAPOL PORNPUTTAPONG, Ph.D.)

.... External Examiner

(Assistant Professor Watoo Phrompittayarat, Ph.D.)

นนทพร เวชพาณิชย์ : การใช้แถบรหัสดีเอ็นเอร่วมกับการวิเคราะห์การหลอมเหลวความละเอียด
สูง (BAR-HRM) เพื่อแยกแยะระหว่างกวาวเครือขาว กวาวเครือดำ และกวาวเครือแดง. (DNA
BARCODING COUPLED WITH HIGH RESOLUTION MELTING (BAR-HRM) ANALYSIS
FOR DIFFERENTIATION WHITE KWAO KHRUEA (*PUERARIA CANDOLLEI*), BLACK
KWAO KHRUEA (*MUCUNA MACROCARPA*), AND RED KWAO KHRUEA (*BUTEA
SUPERBA*)) อ.ที่ปรึกษาหลัก : รศ. ภญ. ร.ต.อ. หญิง ดร.สุชาดา สุขหร่อง, อ.ที่ปรึกษาร่วม : ผศ.
ภญ. ดร.ทักษิณา ชวนอาษา

กวาวเครือเป็นพืชสมุนไพรที่มีสรรพคุณเป็นยาอายุวัฒนะ กวาวเครือที่ใช้ในทางการแพทย์แผนไทย
ได้แก่ กวาวเครือขาว (*Pueraria candollei* Wall. ex Benth.) กวาวเครือดำ (*Mucuna macrocarpa* Wall.)
และกวาวเครือแดง (*Butea superba* Roxb. ex Willd.) โดยกวาวเครือแต่ละชนิดมีองค์ประกอบทางเคมีที่
แตกต่างกัน กวาวเครือขาวมีสารออกฤทธิ์คล้ายฮอร์โมนเอสโตรเจน ส่วนกวาวเครือดำและกวาวเครือแดงมี
สารออกฤทธิ์คล้ายฮอร์โมนเทสโทสเตอโรน โดยกวาวเครือดำมีฤทธิ์แรงกว่ากวาวเครือแดง การที่มีชื่อเรียก
คล้ายกันแต่มีองค์ประกอบทางเคมีต่างกัน ทำให้มีความจำเป็นที่ต้องระบุชนิดของสมุนไพรให้ถูกต้องเพื่อ
ประสิทธิภาพและความปลอดภัยของผู้บริโภค การศึกษานี้เป็นการระบุชนิดของสมุนไพรโดยการใช้ข้อมูลทาง
พันธุกรรม โดยการศึกษาแถบรหัสดีเอ็นเอจำนวนสี่บริเวณ ได้แก่ ยีนแมทเค ยีนอาร์บีซีแอล ดีเอ็นเอระหว่างยี
นทีอาร์เอ็นเอชและพีเอสบีเอ และดีเอ็นเอบริเวณไอทีเอส ผลการศึกษาพบว่าแถบรหัสดีเอ็นเอทั้งสี่บริเวณ
สามารถแยกแยะชนิดของกวาวเครือขาว กวาวเครือดำ และกวาวเครือแดงได้ นอกจากนี้ ได้มีการพัฒนาการ
ตรวจระบุเอกลักษณ์ของกวาวเครือทั้งสามชนิดด้วยการใช้แถบรหัสดีเอ็นเอร่วมกับการวิเคราะห์การ
หลอมเหลวความละเอียดสูงของผลผลิตพีซีอาร์ที่ได้จากปฏิกิริยาลูกโซ่พอลิเมอเรสในสภาพจริงของยีนแมทเค
และไอทีเอส พบค่าอุณหภูมิหลอมเหลวของผลผลิตพีซีอาร์ที่เพิ่มปริมาณจากยีนแมทเคของกวาวเครือแดง
กวาวเครือขาว และกวาวเครือดำที่อุณหภูมิ 73.6, 74.7 และ 74.8 องศาเซลเซียส ตามลำดับ ส่วนอุณหภูมิ
หลอมเหลวของผลผลิตพีซีอาร์ที่เพิ่มปริมาณจากดีเอ็นเอบริเวณไอทีเอสร่วมกับการวิเคราะห์การหลอมเหลว
ความละเอียดสูงนั้น พบอุณหภูมิหลอมเหลวที่ 88.1, 85.6 และ 84.5 องศาเซลเซียส ตามลำดับ กล่าวโดย
สรุป การศึกษานี้ประสบความสำเร็จในการตรวจระบุชนิดของกวาวเครือในตัวอย่างผลิตภัณฑ์เสริมอาหาร
และยาแผนโบราณที่มีสมุนไพรกวาวเครือเป็นส่วนประกอบหลักโดยการใช้แถบรหัสดีเอ็นเอร่วมกับการ
วิเคราะห์การหลอมเหลวความละเอียดสูง ซึ่งสามารถใช้วิธีที่พัฒนาขึ้นนี้ในการตรวจสอบชนิดของสมุนไพร
ก่อนขึ้นทะเบียนผลิตภัณฑ์สมุนไพรกวาวเครือได้

สาขาวิชา	เภสัชศาสตร์และเทคโนโลยี	ลายมือชื่อนิสิต ....
ปีการศึกษา	2562	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

\# \# 6176130233 : MAJOR PHARMACEUTICAL SCIENCES AND TECHNOLOGY

KEYWORD: Kwao Khruea, DNA barcode, Bar-HRM

Nonthapon Vejpanich : DNA BARCODING COUPLED WITH HIGH RESOLUTION MELTING (BAR-HRM) ANALYSIS FOR DIFFERENTIATION WHITE KWAO KHRUEA (*PUERARIA CANDOLLEI*), BLACK KWAO KHRUEA (*MUCUNA MACROCARPA*), AND RED KWAO KHRUEA (*BUTEA SUPERBA*). Advisor: Assoc. Prof. SUCHADA SUKRONG, Ph.D. Co-advisor: Asst. Prof. TAKSINA CHUANASA, Ph.D.

Kwao Khruea is the herbal product that known as a rejuvenation herb. In Thai traditional medicine, three species of Kwao Khruea including white Kwao Khruea (*Pueraria candollei* Wall. ex Benth.), black Kwao Khruea (*Mucuna macrocarpa* Wall.) and red Kwao Khruea (*Butea superba* Roxb. ex Willd.) have been used. Each Kwao Khruea accumulates different chemical constituents. White Kwao Khruea contains estrogenic biological activities while black and red Kwao Khruea accumulate testosterone-like activities. Black Kwao Khruea was reported to be potent than red Kwao Khruea. Because of similar common names with the difference in chemical compositions of Kwao Khruea herbs, it is necessary to differentiate those herbs for efficiency and safety purposes. Therefore, this study aims to differentiate the herbal species using genetic information by barcoding of the four core DNA regions including *mat*K gene, *rbc*L gene, *psb*A-*trn*H transgenic spacer and internal transcribed spacer (ITS). The result showed that all four core barcode regions can differentiate those three Kwao Khruea. In addition, DNA barcoding coupled with high resolution melting (Bar-HRM) analysis of the PCR amplicons by real-time PCR was performed to investigate a particular DNA sequence in the matK gene and ITS. The melting temperature (Tm) of PCR products amplified by matK gene of red, white and black Kwao Khruea were 73.6ºC, 74.7ºC and 74.8ºC, respectively while the Tm of PCR amplicons from the ITS were 88.1ºC, 85.6ºC and 84.5ºC, correspondingly. In conclusion, DNA barcode coupled with HRM analysis has been successful to identify Kwao Khruea species in dietary supplements and traditional medicines containing Kwao Khruea as the main mixture. Our developed method is potentially used for the authentication of Kwao Khruea products prior to registration.

Field of Study:	Pharmaceutical Sciences and Technology	Student's Signature ....
Academic Year:	2019	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The development of economics and scientific progress currently provide the advancement in medical technology allowing an increasing number of global elders and aging societies (Karako K, 2019). Elder people normally require more physical and medical supports to be able to maintain their normal activities. Hormones is the key factor that maintains the system of the body. There are many types of herbal products to maintain their health, both contemporary and traditional. In Thailand, traditional Thai medicine or the use of herbal products is often an alternative for health promotion of people; especially elders. Authenticated herbs are therefore very important for safety, efficacy and quality for the elderly people who have an imbalanced function of the body systems. Various herbal products claimed to enhance health performance are commercially available, for example, turmeric (*Curcuma longa* L.), black galingale (*Kaempferia parviflora* Wall. ex Baker.), brahmi (*Bacopa monnieri* (L.) Wettst.), shatavari (*Asparagus racemosus* Willd.), heart wood of sappan (*Caesalpinia sappan* L.), wan chak motluk (*Curcuma zanthorrhiza* Roxb.) and crude drugs named "Kwao Khruea".

Many herbs have been used for general health improvement whereas some of them have been claimed for gender-specific effects. Since hormones are one of the key factors that maintain the system of the body, herbs that affect male and female performances mostly accumulate hormones-like compounds. Herbs that have long been known to improve male health are Malasian ginseng (*Eurycoma longifollia* Jack) and saw palmetto (*Serenoa repens* (W.Bartram) Small) while soybean (*Glycine max* (L.) Merr.) has beneficial effects on females.

In Thai traditional medicine, the tuberous roots called "Kwao Khruea" are used for rejuvenation. However, there are three types of Kwao Khruea: white (*Pueraria candollei* Wall. Ex Benth.), red (*Butea* superba Roxb.ex Willd.) and black (*Mucuna macrocarpa* Wall.). They belong to the family Fabaceae. Although all of these species are indicated for health revitalization, each differs in its medicinal properties. Two

varieties of *P. candollei*, var. *mirifica* and var. *candollei*, affect females, whereas

B. superba and *M. macrocarpa* exhibit effects on males. Therefore, the correct

differentiation of these plant at the first step is fundamental for efficacy and safety in a

later application.

(A) (B) (C)

Figure 1 Kwao Khruea crude drugs. White Kwao Khruea (A) Black Kwao Khruea (B) and
Red Kwao Khruea (C)

The appearance of three Kwao Khruea freshy plants is different; however, the

root as commonly used parts of these plants share similarities in the features (**figure 1**).

Additionally, the roots in terms of crude drugs are normally available in shredded or

powdered form leading to excessive difficulty in identifying species of their origins. A

previous study showed that the roots of white Kwao Khruea (*Pueraria candollei* Wall. ex

Benth.) (**figure 1A**) accumulates several estrogen-like compounds including miroestrol,

deoxymiroestrol, coumesttol, daiazein, genistein, kwakhurin and mirificoumestan which

are all benefit for females (Yusakul G, 2011). As reported in another study, the roots of

black Kwao Khruea (*Mucuna macrocarpa* Wall.) (**figure 1B**) and red Kwao Khruea

(*Butea superba* Roxb. ex Willd.) (**figure 1C**) exhibited effects on males due to an

accumulation of phytoandrogens which are flavonoid-rich compounds including butenin,

formononetin and prunetin. These phytoandrogens exhibited biological functions similar

to testosterone in men (Roengsumran et al., 2000). Once these three Kwao Khruea plants are misidentification and misused, adverse and toxic effects are conceivably occurred due to their gender-specific chemicals.

Traditional ways to differentiate the plants are using morphological and microscopic analyses that require expertise for the operation. Besides, a chemical analysis is also exploited to help identify plants by providing only chemical information retrieved from samples. Traditional DNA analysis methods, random amplification of polymorphic DNA (RAPD) and PCR-Restriction Fragment Length Polymorphism (PCR-RFLP), are still in practice for plant species identification but become less recognized due to several steps of operations and analyses.

DNA barcoding is a current and promising technique capable to identify individual species. One of the plant DNA sequences that is most frequently used for DNA barcoding is *mat*K gene (Jing & Zon, 2011). There have been many reports on plant authentication and differentiation issues using this area but, in some cases, using only a region is insufficient to differentiate species. Many researchers use four regions that are known as DNA core barcode regions including *mat*K, *rbc*L, *psb*A-*trn*H intergenic spacer and internal transcribed spacer (ITS). To simplify the differentiation method base on DNA barcode information, the Bar-HRM analysis is a preferable technique since it provides a rapid and low-cost detection.

This study was aimed to discriminate white Kwao Khruea, black Kwao Khruea and red Kwao Khruea by using DNA barcoding of 4 areas coupled with high resolution melting (Bar-HRM) analysis. Herbal products containing Kwao Khruea were also examined to prove the concept for an application of this technique.

CHAPTER II

LITERATURE REVIEWS

In this chapter, background information including morphology and chemical constituent of three Kwao Khruea species are described. Besides, both chemical and DNA methods for plant identification are also provided.

2.1 Kwao Khruea herbs

Three species of Kwao Khruea (**Figure 2**, **Figure 3**); white Kwao Khruea (*Pueraria candollei* Wall. ex Benth.) black Kwao Khruea (*Mucuna macrocarpa* Wall.) and red Kwao Khruea (*Butea superba* Roxb. ex Willd.), are used in Thai medical therapy. All of them are liana in family Fabaceae that accumulate their active compounds in their tuberous roots.

2.1.1 Morphology of Kwao Khruea herbs

White Kwao Khruea have two variety *P. candollei* Graham ex Benth. var. *mirifica* (Airy Shew Savat.) (**Figure 2A**) and *P. candollei* Graham ex Benth. var. *candollei*, (**Figure 2B**) both of them have a similar morphology except their pods were different. They are woody climber that the leaves are trifoliate pinnately compound leave and stipulate; terminal leaflet. They have white tuberous roots that are varied in sizes and shapes depending on environmental conditions. The flowers (**Figure 3A**) are bluish-purple in legume shaped. The flower season will be due on late January to early April. The pods are slender typically short or elongate, with hairs (var. *mirifica*) or no hair (var. *candollei*), including 1-10 single seeds when fully developed and turn into brown color when dried. (Yodpetch, 2007) *P. candollei* Graham ex Benth. var. *mirifica* is widely spread in Saraburi province their

fruits have longer hair than *P. candollei* Graham ex Benth. var. *candollei* that
widely spread in Lampang and Kanchanaburi (Chandorn, 2016).

Black Kwao Khruea (*M. macrocarpa*) (**Figure 2C**) is a woody climber
(up to 70 m), which the stems and petioles with sparse to dense cover with
light brown or red-brown fine pubescence. The leaves are trifoliate pinnately
compound with terminal leaflets that are in the elliptic to ovate (or obovate)
shape. Lateral leaflets usually markedly asymmetrical, young leaflets are
hairy like the stem. The stipes of the leaf is absent. The root is black and
tuberous with black latex. Flowers (**Figure 3B**) are inflorescences from old
wood and usually bicolored in form of papilionaceous with conspicuously
pubescent petal margins at the apex, fruit woody, greenish, linear-oblong 6-
15 seeded with margins constricted between seed and linear with thickened
but unwinged margin. (Wallich et al., 1830; Wilmot-Dear, 1992).

Red Kwao Khruea (*B. superba*) (**Figure 2D**) is large size woody
climber or shrub, trifoliate pinnately compound leave, acuminate leaflet and
long leafstalk. The tuberous root containing red sap that appears when it
causes a wound. The flowers (**Figure 3C**) cinnabarine with yellowish-orange
color. The fruits are legumes oblong with silvery silky short hair and have
one seed at the top. The bark has red latex (Chen et al., 2014).
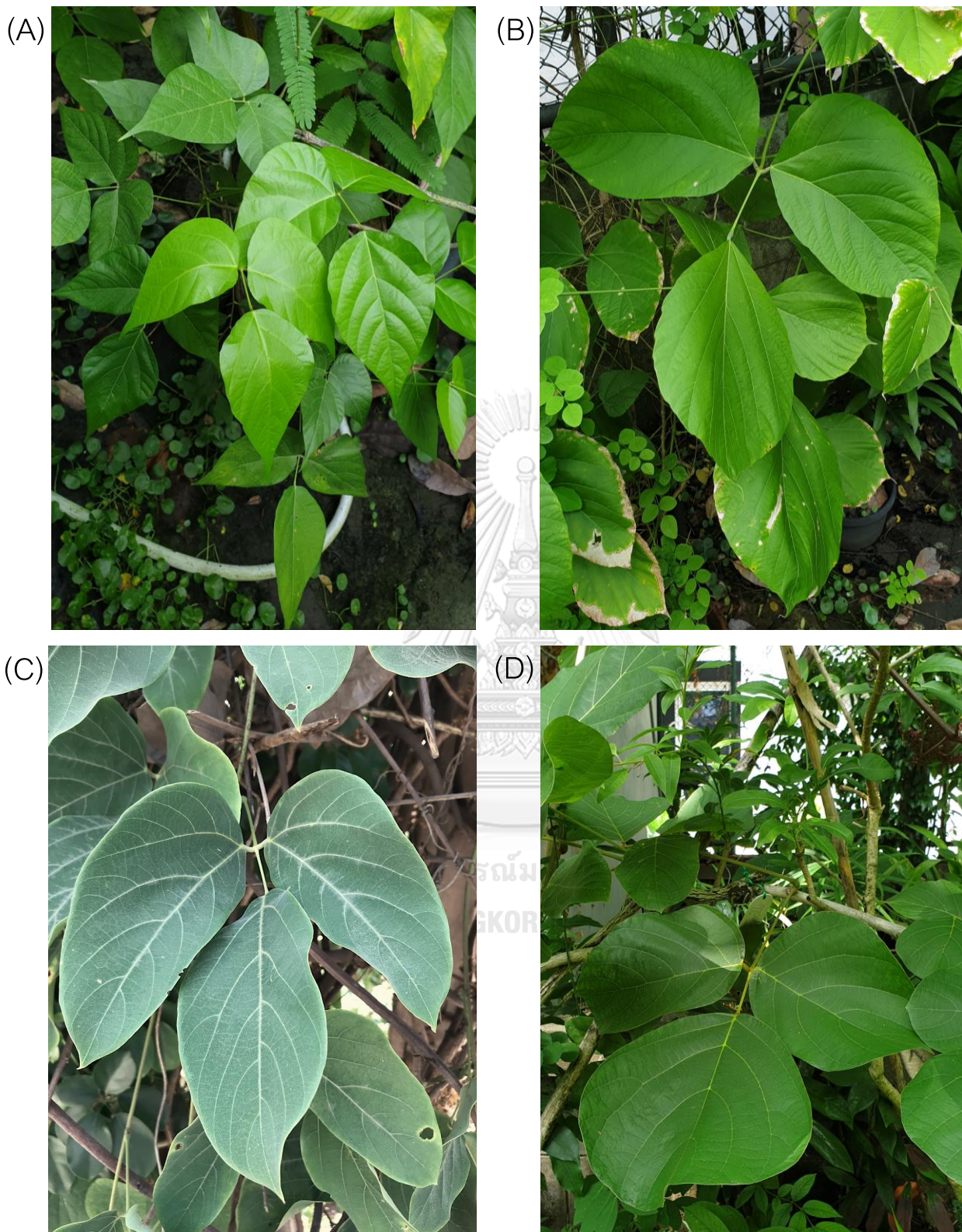
Figure 2 Pueraria candollei Wall. ex Benth. var. *mirifica* (A) *Pueraria candollei* Wall. ex Benth. var. *candollei* (B) *Mucuna macrocarpa* Wall. (C) and *Butea superba* Roxb. ex Willd. (D)

(A)

(B)

(C)
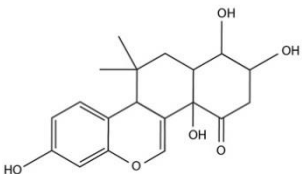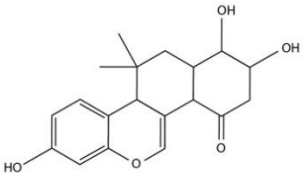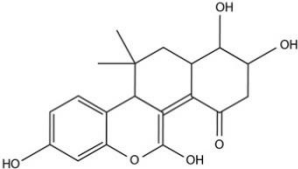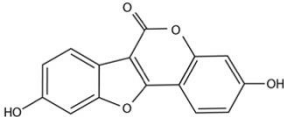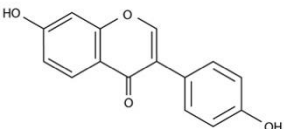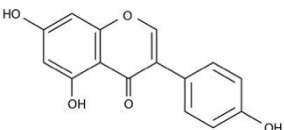
Figure 3 Kwao Khruea Flower White Kwao Khruea (A) Black Kwao Khruea (B) and Red Kwao Khruea (C)
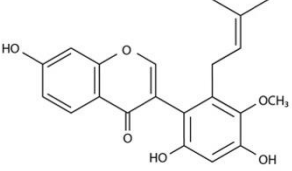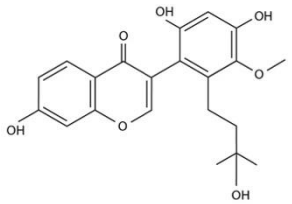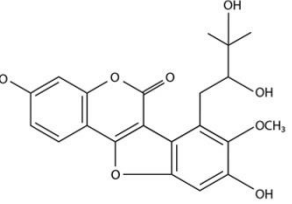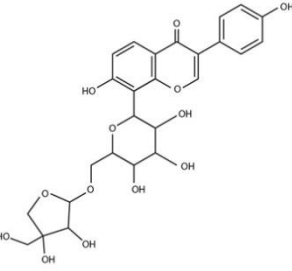
2.1.2 Chemical constituents and bioactivities of Kwao Khruea herbs
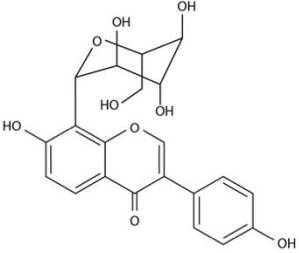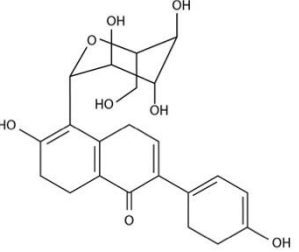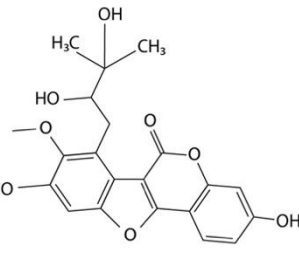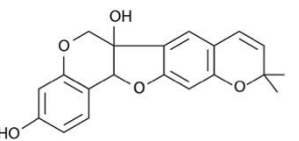
White Kwao Khruea was found to accumulate several estrogen-like compounds which are all beneficial for females (Yusakul G, 2011). Both black and red Kwao Khruea showed effects on males due to phytoandrogens (Roengsumran et al., 2000). These substances are classified into severed phytochemical groups based on their chemical structures. Various bioactivities of these compounds were observed in a number of researches. The chemical constituents and bioactivities of the three Kwao Khruea herbs are showed on **Table 1**.
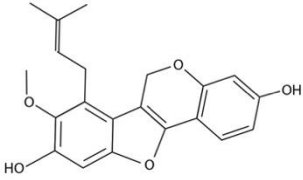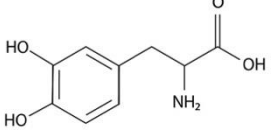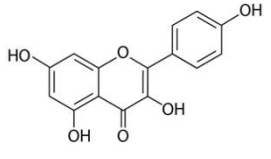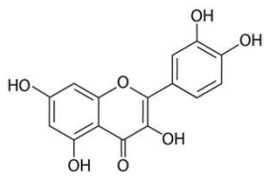
Table 1 Chemical constituent and bioactivities of Kwao Khruea herbs
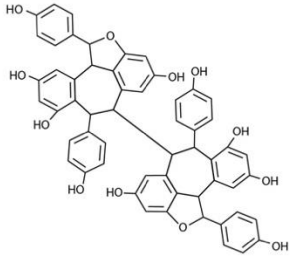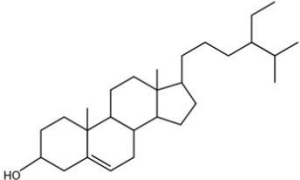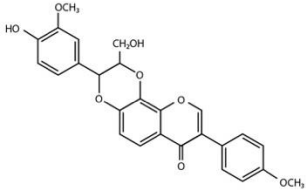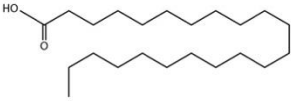
Plant sample	Chemical structure and their name	Classification	Bioactivities	References
white Kwao Khruea (*Pueraria candollei* Wall. ex Benth.)	Miroestrol	Chromene	Fulfillment of the processes in pharmacokinetics	(Chansakao w et al., 2000a)
	Deoxymiroestrol	Chromene	Effects on MCF-7 human breast cancer cells	(Chansakao w et al., 2000b)
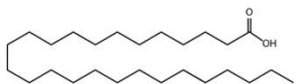
Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	Isomiroestrol	Chromene	Estrogen steroid hormone biosynthesis pathway	(J. H. Lee et al., 2017)
	Coumesttol	Coumestan	Fulfillment of the processes in pharmacokinetics	(Chansakao w et al., 2000a)
	Daidzein	Isoflavone glycoside	Inhibit key enzymes in the steroid metabolism pathway	(Poschner et al., 2017)
	Genistein	Isoflavone glycoside	Angiogenesis inhibitor	(Coward, Barnes, Setchell, & Barnes, 1993)
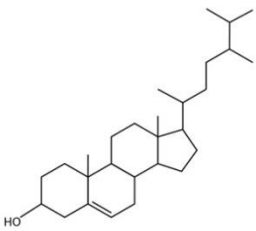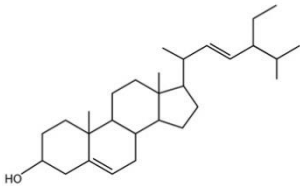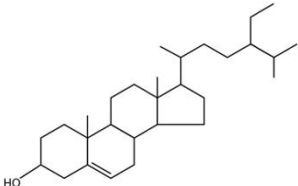
Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	 Kwakhurin	Isoflavone	Exhibited cytotoxicity effect for MCF-7 cells	(Selepe & Van Heerden, 2013)
	 Kwakhurin hydrate	Isoflavone	Inhibition of LDL-cholesterol oxidation Anti-platelet aggregation	(Kapiotis et al., 1997)
	 Mirificoumestan	Coumestan	Fulfillment of the processes in pharmacokinetics	(Ingham, Tahara, & Dziedzic, 1988)
	 Mirificin	Isoflavone glycoside	Inhibition LDL-cholesterol oxidation Anti-platelet aggregation	(Kapiotis et al., 1997)

Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	Puerarin	Isoflavone glycoside	Inhibition of LDL-cholesterol oxidation	(Kapiotis et al., 1997)
	Puerarin-6"-monoacetate	Isoflavone glycoside	Inhibit LDL-cholesterol oxidation	(Kapiotis et al., 1997)
	Mirificoumestan glycol	Coumestan	Fulfillment of the processes in pharmacokinetics	(Kapiotis et al., 1997)
	Tuberosin	Pterocarpene	Antioxidation	(Ingham et al., 1988)

Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	Puemiricarpene	Isoflavone	Low estrogenic activity	(Shimokawa et al., 2013)
black Kwao Khruea (*Mucuna macrocarpa* Wall.)	L-3, 4-dihydroxyphenylalanine	Amino acid	Increased urinary morphine, codeine and tetrahydropapaveroline in Parkinsonian patient	(Matsubara, Collins, & Neafsey, 1992)
	Kaempferol	Flavonoid	Lowing blood sugar levels	(Roengsumran et al., 2001)
	Quercetin	Polyphenols	Antioxidation	(Roengsumran et al., 2001)

Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	Hopeaphenol	Stilbenoid	Antioxidation	(Roengsumran et al., 2001)
red Kwao Khruea (*Butea superba* Roxb. ex Willd.)	Beta-sitosterol	Flavonoid	cAMP phosphodiesterase inhibitor Induction of penile erection	(Tocharus, Jamsuwan, Tocharus, Changtam, & Suksamrarn, 2012)
	Butesuperin	Flavonoid	Anti-estrogenic activity (aphrodisiac effect)	(Tocharus et al., 2012)
	Behenic acid	Saturated fatty acid	Antioxidation	(Badea, Lăcătuşu, Badea, Ott, & Meghea, 2015)

Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	 Tricosanoic acid	Saturated fatty acid	Inhibition of urease	(Rauf, Uddin, Arfan, & Muhammad, 2013)
			DPPH radical scavenging activities	
	 Lignoceric acid	Saturated fatty acid	Antifungal activity	(Rauf et al., 2013)
			Antibacterial activity	
			Inhibition of urease	
	 Pentacosanoic acid	Saturated fatty acid	Antifungal activity	(Kurdyukov et al., 2006)
	 Hexacosanoic acid	Saturated fatty acid	Antimicrobial activity	(Singh & Singh, 2003)

Plant sample	Chemical structure and their name	Classification	Bioactivities	References
	Campesterol	Steroid	Precursor of anabolic steroid boldenone	(Anwar et al., 2013)
	Stigmasterol	Steroid	Increasing phytosterol contents Lowering LDL levels	(D.-U. Lee, Park, & Lee, 2020)
	Beta- sitosterol	Steroid	Decrease benign prostatic hyperplasia Immune system modulation Prevention of cancer	(D.-U. Lee et al., 2020)

2.2 Methods for differentiation of the medicinal plants

There are several methods of plant differentiation including analysis of morphological characteristics, an examination of chemical constituents and investigation of genetic information. Currently, morphological analysis the first required step to identify the plant species when none of the advanced technology is available for identification of the plant of interest. Chemical and nucleotide analyzes of plant species are useful in helping identify plants with greater confidence. In this

topic, the information of both chemical and genetic methods for plant differentiation are informed in detail.

2.2.1 Chemical methods

Chemical analyses are approaches to assist in plant identification. A basic chemical analysis is thin layer chromatography (TLC). This is a method of separating the chemical composition of a sample by using different ratios of solvents passing through the TLC plate. Another chemical analysis is high performance liquid chromatography (HPLC). This method permits the separation of chemicals in the sample by the real-time change of the mobile phase. HPLC is generally combined with mass spectrometry (MS) for chemical identification resulting in the generation of the mass spectrum and structure elucidation of separated compounds. Both TLC and HPLC techniques are normally used to compare samples of interest and reference authentic samples (Zhao et al., 2006).

Although chemical methods are extensively used as standard analyses, they have some limitations. Since closely related plants accumulate similar phytochemicals, chemical methods are not effective enough for the differentiation of plants. Besides, many factors; for example, the growing environment and varieties of the same species, play a role in the chemical accumulation of a plant. These factors have the potential to alter the levels of accumulated chemical constituents and allow inaccurate interpretations of plant species if chemical methods are the only techniques for plant differentiation.

2.2.2 Molecular biology methods

Molecular biology is about the study of the genes that is unique in single species. The gene has its function and is inherited from generation to generation. For authentication the herbal product, DNA is one of the popular

choices because of the unique among species. This method can difference plant species with specificity, accuracy and fluently. It not only detectes the fresh plants but also processed plants that are in the form of medicine, capsule and others. Small amount of samples with the broken pieces of DNA are sufficient for species differentiation by multiplying a particular region of DNA.

The DNA-sequence of interest is amplified by polymerase chain reaction (PCR) that is the basic method of DNA analysis to generate amplicons. The PCR products are further analyzed by a single or several methods depending on the purpose of the study. Dominant and co-dominant marker systems allow for analyzing multiple and single loci at one time. The fundamental dominant marker systems, such as AFLP, RAPD, iPBS, SRAP and ISSR, permit DNA analyses with no requirement of detailed sequence information. Although this method is the ease of DNA examination, it is not suitable for investigating heterozygotes of which DNA is from the nuclear genome (Loera-Sánchez, Studer, & Kölliker, 2019). On the other hand, the co-dominant marker distinguishes heterozygotes from homozygotes, allowing the determination of genotypes and allele frequencies. After finishing PCR amplification, PCR products are also examined by DNA sequencing and the sequencing results are subsequently analyzed by DNA alignment. The DNA sequence would show the identity of the plant species and the result can be used to generate their DNA barcode.

2.2.1.1 Genome sequencing and compare

There are many sequencing technologies, the first generation of DNA sequencing since 1970s is sanger sequencing. This method claims to be the gold standard of DNA sequencing because of the accuracy. The nucleotides were detected singerly by fluorescent that run into the buffer pass-through special agarose gel or glass capillary.

The result of sanger sequencing was analyzed by DNA alignment and used as DNA barcoding of the species. Sanger sequencing can show the single nucleotide signature (SNP) of the species in a short sequence (not more than 800 bp). The pros of this technique are low cost with high accuracy nucleotide detected. However, this technique has some limitations. The number of nucleotides are read by the distance of agarose gel or glass capillary that has generally been used (Wangler et al., 2017). Also, the limitation of the number of samples per round, whole genome sequencing and the way to prepare the sample automatically. For longer sequence, DNA need to be clone in bacteria (Md & Abhijit, 2012)

Another sequence technology since 2000s is next-generation sequencing (NGS) which is a high-though put technique. The difference of second-generation sequencing and third-generation sequencing is how long of reads lengths per sample and detected DNA without the amplification step. (Ari & Arıkan, 2016) Second-generation sequencing is proper for short read lengths (100-500 bp) and third-generation sequencing is proper for long read lengths (more than 14 kb). The limitation of second-generation sequencing is the amplification biases resulting and high costs per sample. The limitation of third-generation sequencing is still in high error rate, biased towards long fragments and higher cost per sample than Sanger sequencing. (Leblanc & Marra, 2015)

2.2.1.2 DNA barcoding technique

DNA barcoding is the pattern of a nucleotide sequence that represents the uniqueness of the species. There are four universal barcode areas that are used for plant differentiation called DNA core

barcode areas (**Figure 4**). The Barcode of Life Plant Working Group recommends the genomic regions of *mat*K (**Figure 4A**) and *rbc*L (**Figure 4B**) in the chloroplast genome for barcoding (Hollingsworth et al., 2009). Another recommends the genomic regions are *psb*A-*trn*H transgenic spacer (**Figure 4C**) that is also in the chloroplast region and Internal transcribed spacer (ITS) (**Figure 4D**) that is the nuclear region. (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005).

DNA barcode of a plant can show the nucleotide signature of the species that can be used for other rapid, convenient and reasonable costs for differentiating herbal products.

Although four areas of DNA are recommended to generate the DNA barcode, it was reported that the four areas were sufficient for differentiating plants within the same lineage but not for multi-species lineages (Stallman, Funk, Price, & Knope, 2019).

2.2.1.3 DNA Barcoding couple with high-resolution melting (Bar-HRM) analysis

An alternative analyzing result after finishing the DNA barcoding of the species is to combine with the high-resolution melting analysis. This technique is the post-PCR method that discriminates double-stranded DNA based on sequence length, GC content and strand complementarity by detecting the single base difference in mutation and indels of the nucleotide sequence (Simko, 2016). Thus, it shortens the time use and reduces the cost per sample. In brief, the first step of this technique is to design primers that cover the short sequence of nucleotide signature areas of the sample. The real-time PCR is then exploited to detect the diminishes of fluorescence dye that releases from double-stranded DNA when the change of temperature is captured in the thermal cycler. The

fluorescent dye intercalates double-strand DNA while PCR products are amplified (**Figure 5**). After finishing real-time PCR, the melting temperature of PCR amplicons is subsequently examined from 60 to 90°C in 0.2°C increments which are more constricted than those of standard melting curve protocols. The results from HRM analysis are shown in graph of melting curves (**Figure 6**) that have different shape and peak for each sample (Osathanunkul M., 2015). HRM software permits the detection of pre- and post-melt fluorescent intensity and automatically normalizes raw data to generate a normalized plot, a graph with relative values of 1.0 (pre-melting) and 0 (post-melting) (**Figure 6A**). HRM data are usually plotted in the difference plot (**Figure 6B**) to visualize the difference cluster of the melting profile among the samples. The same cluster represents the same group of samples.
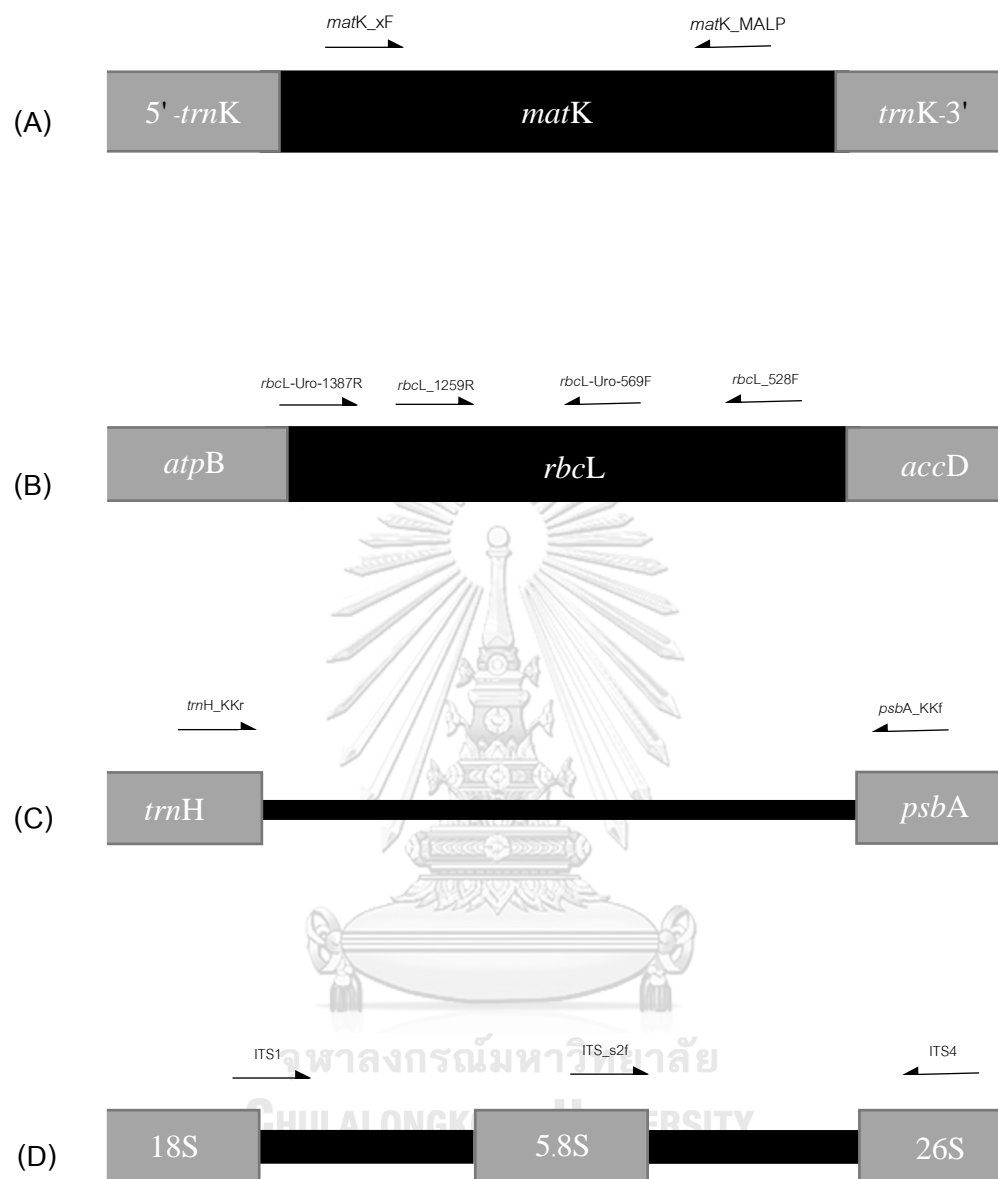
Figure 4 The schematic diagram of the application loci and primer pairs for amplifications and sequencing *mat*K (A), *rbc*L (B), *psb*A-*trn*H transgenic spacer (C) and internal transcribed spacer (ITS) (D)
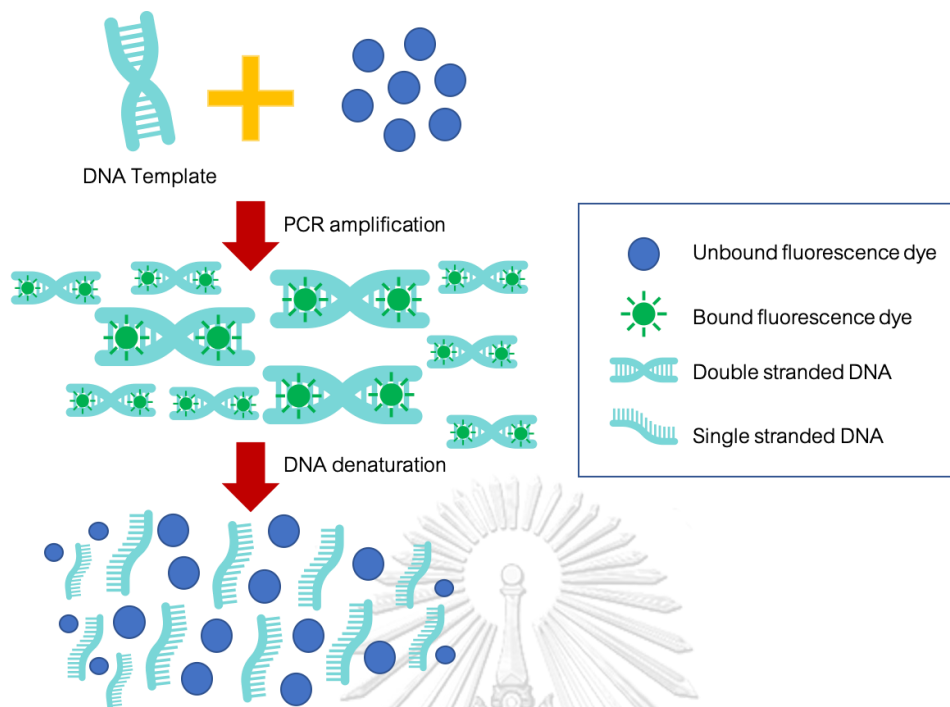
Figure 5 The diagram of fluorescence dye activities while pre-PCR, PCR amplification and post-PCR (DNA denaturation) which inferred to the HRM analysis
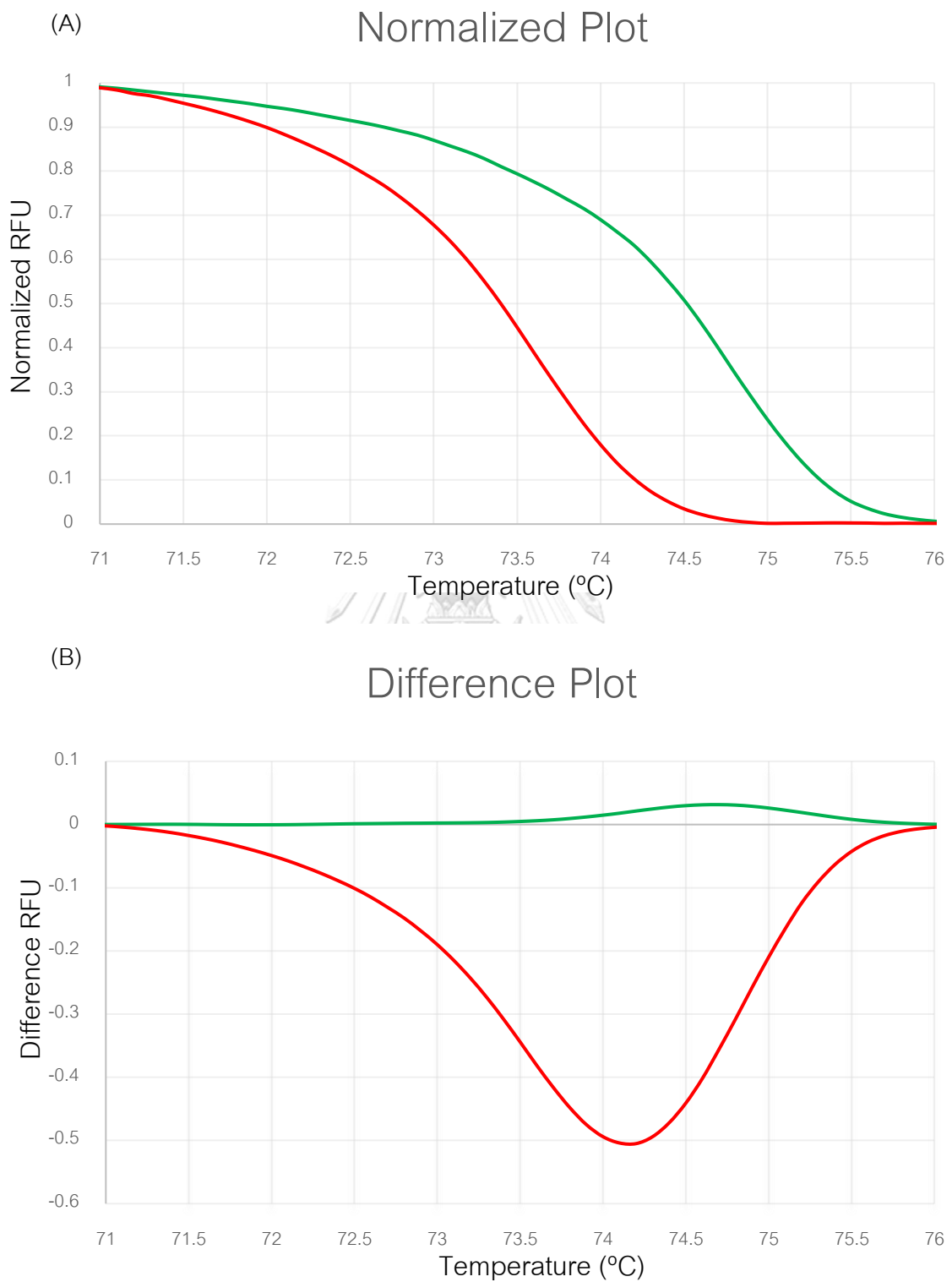
Figure 6 The schematic of melting curve from real-time PCR couple with HRM analysis for plant differentiation Normalized Plot (A) Difference Plot (B)

CHAPTER III

MATERIALS AND METHODS

In this study, the three Kwao Khruea species collected from several locations were initially analyzed to retrieve their DNA core barcode sequences. Barcoding primers were used to amplify four standard regions including *rbc*L, *mat*K, ITS and *psb*A-*trn*H transgenic spacer. Since barcoding generation from *mat*K and ITS sequences was successful, primers designed for Bar-HRM analysis of these two regions were created. In addition, Thai herbal products claimed to have these three Kwao Khruea species as main ingredients were investigated for propriety of plant species using Bar-HRM analysis.

3.1 Collection of plant materials and herbal products

Three species of Kwao Khruea, *P. candollei* (white Kwao Khruea), *M. macrocarpa* (black Kwao Khruea) and *B. superba* (red Kwao Khruea), considered as authentic samples were collected from various locations in Thailand (**Table 2**). All samples were identified by Associate Professor Thatree Phadungcharoen at the Chulalongkorn University. Herbal products that labeling with Kwao Khruea as an ingredient were randomly bought from herbal markets at various locations in Thailand (**Table 3**). All herbal product samples in this study were in capsule formulation.

Table 2 Information of plant materials for authentic samples

Scientific name	Common name	Place of collection (Province)	Voucher no.	Accession number			
				matK	*psbA-trnH*	*rbcL*	ITS
Pueraria candollei Graham ex Benth. var. *mirifica* (AiryShewSavat.)	white Kwao Khruea	Bangkok	WY081205	LC570920	LC570937	LC570932	LC573986
		Nonthaburi	TH090505	LC570921	LC570938	LC570933	LC573987
		Lampang	MUS-H3852	-	LC570939	-	-
		Saraburi	MUS-H3853	-	LC570940	-	-
		Chiang Mai	MUS-H3854	-	-	-	-
		Loei	MUS-H3855	-	-	-	-
		Tak	MUS-H3856	-	-	-	-
Pueraria candollei Graham ex Benth. var. *candollei*		Nakhon Pathom	WY270106	-	-	-	-
		Kanchanaburi	MUS-H3860	-	-	-	-
		Prachuap Khiri Khan	MUS-H3858	-	-	-	-
Mucuna macrocapa Wall.	black Kwao Khruea	Lampang	SS-591	LC570922	LC570941	LC570931	LC573988
		Chiang Mai	MUS-H3849	-	LC570942	-	LC573989
		Bangkok	MUS-H3848	-	-	-	-

Scientific name	Common name	Place of collection (Province)	Voucher no.	Accession number			
				matK	psbA-trnH	rbcL	ITS
		Bangkok	TH050706	LC570919	LC570935	LC570934	LC577560
		Prachin Buri	SS-700	-	LC570936	-	LC577561
Butea superba Roxb. ex Willd.	red Kwao Khruea	Chachoengsao	TH130206	-	-	-	-
		Lampang	MUS-H3861	-	-	-	-
		Prae	MUS-H3862	-	-	-	-
		Kanchanaburi	MUS-H3863	-	-	-	-
		Kalasin	BR010807	-	-	-	-

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Table 3 Information of commercial herbal products from Kwao Khruea used in this study

Code name	Commercial herbal products	Species on the label	Place of collection (Province)
WC01		white Kwao Khruea	Bangkok
WC02		white Kwao Khruea	Bangkok
WC03		white Kwao Khruea	Bangkok
WC04		white Kwao Khruea	Bangkok
WC05		white Kwao Khruea	Bangkok

Code name	Commercial herbal products	Species on the label	Place of collection (Province)
WC06		white Kwao Khruea	Bangkok
BC01		black Kwao Khruea	Lampang
RC01		red Kwao Khruea	Bangkok
RC02		red Kwao Khruea	Bangkok
RC03		red Kwao Khruea	Bangkok
RC04		red Kwao Khruea	Bangkok

3.2 Genomic DNA extraction

Total genomic DNA was extracted from 30-40 mg of fresh leaves and individual herbal product sample. The samples were ground into fine powder by Tissue lyser LT (Qiagen, Germany). Genomic DNA was extracted by using DNeasy plant pro DNA extraction kit (Qiagen, Germany). The quality of genomic DNA was investigated by separating DNA in 0.8% agarose gel electrophoresis, staining with red safe dye and visualizing under UV light of GelDocTM XR (Bio-Rad, U.S.A.). The quantity of genomic DNA was assessed in O.D. using NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer with Wi-Fi (ThermoFisher scientific, Germany). The DNA samples were stored at -20ºC until further use.

3.3 DNA barcode generation of Kwao Khurea

3.3.1 Primer design

The universal primer for four loci DNA regions including ITS, *mat*K, *rbc*L and *psb*A-*trn*H intergenic spacer were used in PCR reaction (**Table 4**) and the schematic diagram of the amplificated loci and primer pairs for PCR and sequencing. (**Figure 4**)

3.3.2 PCR amplification

The PCR reactions were performed in 25 µl amplification volume consisting of 1X PCR buffer with 1.5mM $MgCl_2$ (Invitrogen), 0.2mM $dNTP_{mix}$ (Invitrogen), 10µM primers, 0.1 unit of Platinum *Taq* DNA Polymerase (Invitrogen) and 20 ng of genomic DNA and carried out in GS-96 PRO Gradient Touch Thermal Cycler (Hercuvan Lab system, England) using the cycling condition at 94ºC for 4min, followed by 30 cycles of 94ºC for 30s, 55 ºC 30s for ITS and *psb*A-*trn*H intergenic spacer loci or 57 ºC 30s for *mat*K locus or 58 ºC 30s for *rbc*L locus then 72ºC for 30s, and final extension at 72 ºC for 5min (**Table 5**). The amplified PCR products were examined by 2%

agarose gel electrophoresis in 1% TAE buffer and visualized under UV light

by using Ultrepower™ Nucleic Acid Staining in GelDoc™ XR (Bio-Rad,

U.S.A.) and then photographed.

Table 4 Primers used for DNA barcoding

Location	Primer names	Primers sequences (5'⟶ 3')	References
*mat*K	*mat*K_xF	TAA TTT ACG ATC AAT TCA TTC	Ford et al. 2009
	*mat*K_MALPR1	ACA AGA AAG TCG AAG TAT	Dunning & Savolainen 2010
*psb*A-*trn*H intergenic spacer	*psb*A_KKf	GCG ATC GAA GCT CCA TCT AT	Redesigned for this study
	*trn*H_KKr	GCT ACA TCC GCC CTT ATA CT	
*rbc*L	*rbc*L-Uro-569F	ATG AAT GTC TTC GCG GTG GAC	Redesigned for this study
	*rbc*L-Uro-1387R	TTA CAA AGT ATC CAT TGC TTG AAA T	
	*rbc*L_528f	CTG CAT GCA TTG CAC GAT GT	Redesigned for this study
	*rbc*L_1259R	CGG TAG CTG CCG AAT CTT CT	
ITS	ITS1	TCC GTA GGT GAA CCT GCG G	White et al. 1990
	ITS4	TCC TCC GCT TAT TGA TAT GC	
	ITS-s2f	ATG CGA TAC TTG GTG TGA TGA AT	Chen et al. 2010

Table 5 The PCR reaction and condition for Kwao Khruea barcoding

Reaction / PCR components	1X	PCR condition				
		Pre-denaturation	30 cycles			Post-extension
			Denaturation	Annealing	Extension	
PCR buffer	1X	94ºC 4min	94ºC 30s	55 ºC 30s for ITS and *psb*A-*trn*H intergenic spacer	72ºC 30s	72 ºC 5min
MgCl$_2$	1.5mM					
dNTP$_{mix}$	0.2mM					
Forward primers	10µM			57 ºC 30s for *mat*K		
Reverse primers	10µM					
DNA polymerase	0.1 unit			58 ºC 30s for *rbc*L		
Genomic DNA	20ng					

3.3.3 Nucleotide sequence by Sanger sequencing

For DNA sequencing, the PCR products were bi-directionally sequenced using the same pair of primers of each region by ABI 3730XL DNA analyzer at BIONEER Co., Ltd. The raw DNA sequences from four regions were aligned, edited, corrected, verified and compiled using MEGA version 7 to confirm the quality of sequences. The consensus sequences of obtained sequences were analyzed by MUSCLE algorithm (Edgar, 2004). All sequences were submitted to GenBank database with their accession numbers. (**Table 2**)

3.4 Barcode DNA-High Resolution Melting (Bar-HRM) analysis

3.4.1 Primer design for Bar-HRM analysis

The primers for Bar-HRM analysis were designed based on nucleotide signatures from sequencing results. The primer pair was located on a representative position that was able to differentiate the three species at once. The amplification reaction was run for three different regions of *mat*K

and one region of ITS that generated amplicons with 113, 47, 114, 134 bps in sizes, respectively. The specific primer sets were shown in table 6. (**Table 6**).

Table 6 Primers for real-time PCR coupling with HRM analysis*.*

Primer set	Primer names	Primers sequences (5'$\longrightarrow$ 3')	Amplicon size
KK*mat*K1	KK*mat*K1_F	GCG ATC AAT TCA TTC CAT TTT TCC C	113 bp
	KK*mat*K1_R	GGA TTT GAA CCA AGA TTT CCA GAT GG	
KK*mat*K2	KK*mat*K2_F	GTC AGA TAT ACG AAT ACC C	47 bp
	KK*mat*K2_R	CCA TCT GGA AAT CTT GGT TC	
KK*mat*K3	KK*mat*K3_F	GGA ATA GTC TTT TTA CTC C	114 bp
	KK*mat*K3_R	GGG AAT ATG AAT CTA TCT TTC	
ITS_KK1	ITS_KK1_F	CTG CCT GGG TGT CAC ACA TC	134 bp
	ITS_KK1_R	GGC CAC GAA CTC GAT TTT CAA C	

3.4.2 HRM analysis using Real-time PCR

The real-time PCR was used to perform HRM analysis by investigating the melting temperature of PCR amplicons. The DNA melting character of each locus from individual Kwao Khruea species was obtained by CFX96 Touch™ Real-time PCR Detection System (Bio-Rad). The real-time PCR reactions were carried out in a total volume of 20 µl containing 1X of Ssofast Evagreen Supermix® (Bio-Rad), 0.4 µM of forward and reverse primers and 10 ng of genomic DNA for all samples. PCR reactions were performed in individual PCR tube and the cycling condition consisted of the initial denaturing step at 98ºC for 1 min followed by 40 cycles of 98ºC for 15s, 56 ºC for KK*mat*K1 and ITS_KK1 primer set and 62 ºC for KK*mat*K2 and KK*mat*K3 primer set for 15s and 72ºC for 15s. (**Table 7**). The fluorescent data for PCR amplification were collected during

extension step as a green channel. In fluorescent data for PCR amplification, PCR products were denatured at 95ºC for 1 min, and then re-annealed at 60 ºC for 1 min to reform random DNA duplexes. For the melting curve analysis, the PCR product was melted in the ramped steps from 60ºC to 90ºC in 0.2 ºC increments and the fluorescence intensity was then collected on every increasing point.

The melting curves were analyzed using CFX Manager[TM] Software Upgrade of version 3.1 and Precision Melt Analysis[TM] software upgrade of version 3.1. All reactions were performed in triplicate and shown in melting curve as a normalized plot and then transformed into a difference plot. The results of individual transformed curves were shown in clusters.

Table 7 The PCR reaction and condition for HRM analysis

Reaction PCR components	1X	PCR condition				Melting curve analysis		
		Pre-denaturation	40 cycles			Re-denaturation	Re-annealed	Post-extension
			Denaturation	Annealing	Extension			
SsoFast EvaGreen supermix	1X	98ºC 1min	98ºC 15s	62 ºC 15s for KK*mat*K2 and KK*mat*K3 primer set	72ºC 15s	95 ºC 1min	from 60 ºC to 90 ºC in 0.2 ºC increments	95 ºC
Forward primers	0.4µM			56 ºC 15s for KK*mat*K1 and ITS_KK1 primer set				
Reverse primers	0.4µM							
Genomic DNA	10ng							

CHAPTER IV
RESULTS

In this chapter, agarose gel electrophoresis of genomic DNA of the three Kwao Khruea species (**Figure 7**) and their products (**Figure 8**) including of the four areas of DNA core barcodes was reported. Bar-HRM analysis of the *mat*K gene and ITS regions of three Kwao Khruea herbs has been successful developed. Also, Bar-HRM technique can be applied to test Thai herbal products that claimed to be specific Kwao Khruea species as main ingredients.

4.1 Samples collection and DNA extraction

Fresh samples of these three species of Kwao Khruea were collected from various area distributed throughout Thailand (**Table 2**) and the samples of herbal products were collected from both offline and online markets (**Table 3**).

Genomic DNA were extracted from fresh samples then checked the DNA quality by agarose gel electrophoresis (**Figure 7**). The genomic DNA of Kwao Khruea product also run on agarose gel electrophoresis (**Figure 8**).
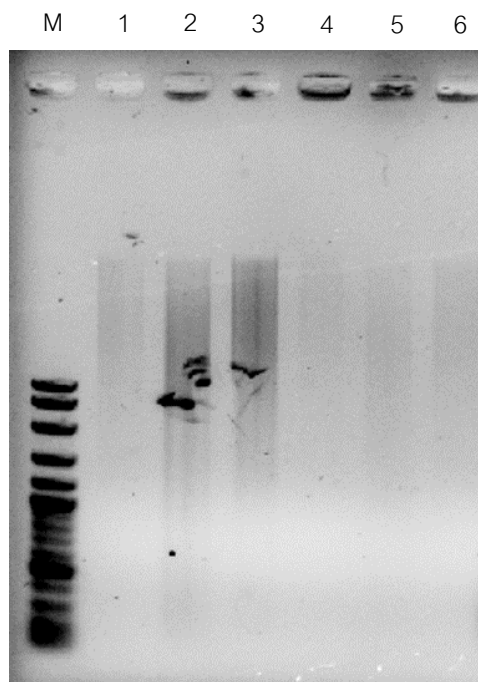
Figure 7 0.8% agarose gel electrophoresis of genomic DNA of three Kwao Khruea herbs and their leaves; M = DNA marker: lamda DNA/HindIII marker; lane 1 = *M. macrocarpa* 1'elution; lane 2 = *B. superba* 1'elution; lane 3 = *P. candollei* 1'elution; lane 4 = *M. macrocarpa* 2'elution; lane 5 = *B. superba* 2'elution; lane 6 = *P. candollei* 2'elution
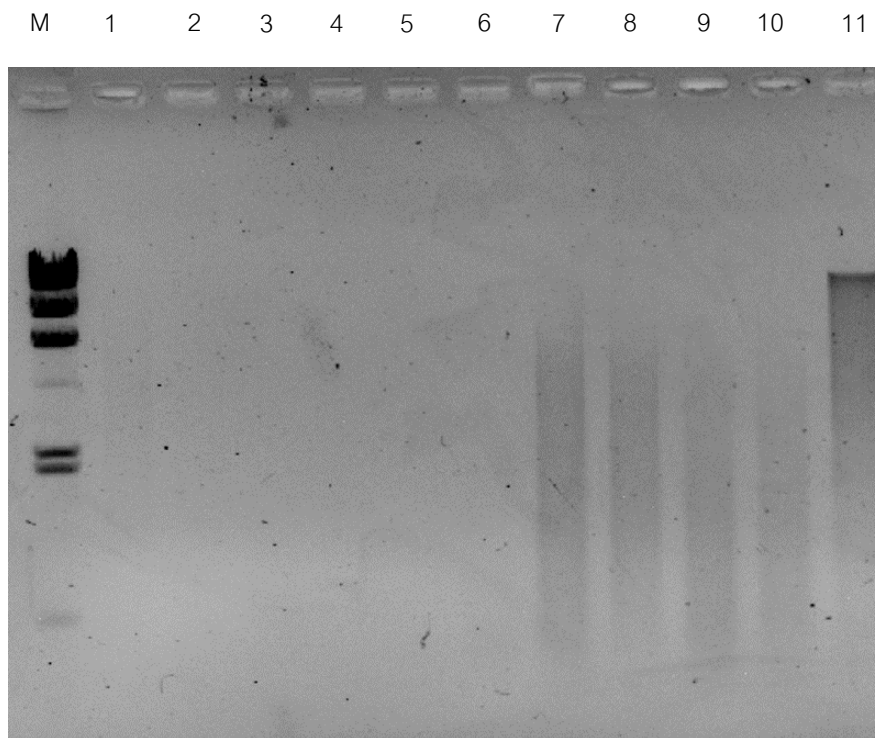
Figure 8 0.8% agarose gel electrophoresis of genomic DNA of Kwao Khruea product; M = DNA marker: lamda DNA/HindIII marker; lane 1 = WC1; lane 2 = WC2; lane 3 = WC3; lane 4 = WC4; lane 5 = WC5; lane 6 = WC6; lane 7 = BC1; lane 8 = RC1; lane 9 = RC2; lane 10 = RC3; lane 11 = RC4

4.2 Core DNA barcoding of the three Kwao Khruea herbs

The genomic DNA of the three Kwao Khruea herbs collected from different locations of Thailand was successfully extracted for DNA barcoding. Core DNA loci were amplified from all genomic DNA samples of fresh herbs and further sequenced. The partial sequences of the four core DNA loci of the three Kwao Khruea herbs were consisted of ITS, *mat*K, *rbc*L and *psb*A-*trn*H intergenic spacer. The primer location of each area was shown in the schematic diagram (**Figure 4**) The difference lengths of DNA sequences were various with their own nucleotide signatures (**Table 6**). All sequence alignments and their nucleotide signatures were shown in **Appendix**

Table 8 Sequence analysis of four core DNA barcode regions of Kwao Khruea plants

DNA region	*mat*K	ITS	*psb*A-*trn*H	*rbc*L
Length (bp)	1518-1527	366-393	272-357	657-810
Variation position (bp)	123	119	53	20
% Variation	8.0815	31.3570	16.8521	29.6296
Average %GC content	28.5	61	22.6667	41.3333

The complete sequences of maturase K (*mat*K) gene from all Kwao Khruea plants were amplified and subsequently bidirectional sequenced. The amplicon was inspected by gel electrophoresis and the PCR product sizes were ranged 1500 bp. The complete sequence of *mat*K from *P. candollei* var. *mirifica, P. candollei* var. *candollei, M. macrocarpa* and *B. superba* were 1521, 1521, 1518 and 1527 bp, respectively. The partial sequences of *mat*K element from *P. candollei* var. *mirifica, P. candollei* var. *candollei* and *M. macrocarpa* were 534 bp and *B. superba* was 540 bp. The *mat*K sequences of these three Kwao Khruea were aligned and the alignment showed the high number variable sites (123 of 540 sites or 8.0815%) and high %GC content (28.5%). The *mat*K sequences were identical within species. The *mat*K sequence alignments and their nucleotide signature were submitted to GenBank database and all sequenced were shown in **Appendix**.

The partial sequences of ITS region from all Kwao Khruea plants were amplified and subsequently bidirectional sequenced. The amplicon was inspected by gel electrophoresis and the PCR product sizes were approximately 400 bp. The partial sequences of ITS element from *P. candollei* var. *mirifica, P. candollei* var. *candollei, M. macrocarpa* and *B. superba* were 378, 369, 393 and 366 bp, respectively. The ITS sequences of these three Kwao Khruea were aligned and the alignment showed the high number variable sites (119 of 393 sites or 31.3570%). Moreover, %GC content of ITS was 61% which was higher

than that of the *mat*K aligned. The ITS sequences of any samples were identical within species. The ITS sequence alignments and their nucleotide signature were also submitted to GenBank database and all sequenced were shown in **Appendix**.

The partial sequences of photosystem II protein D1 – tRNA-Histidine intergenic spacer (*psb*A-*trn*H intergenic spacer) were amplified and sequenced. The PCR products were inspected by gel electrophoresis and the PCR product sizes were approximately 350 bp. The partial sequences of *psb*A-*trn*H intergenic spacer from *P. candollei* var. *mirifica, P. candollei* var. *candollei*, *M. macrocarpa* and *B. superba* were 272, 272, 320 and 357 bp, respectively. All *psb*A-*trn*H intergenic spacer sequences were aligned and the alignment showed 53 variation sites (16.8521%) with 22.6667% of GC content. Moreover, %GC content of *psb*A-*trn*H intergenic spacer was lower than that of the *mat*K aligned. The *psb*A-*trn*H intergenic spacer sequences were identical within species. The *psb*A-*trn*H intergenic spacer sequence alignments and their nucleotide signature were also submitted to GenBank database and all sequenced were shown in **Appendix**.

The partial sequences of ribulose-1,5-bisphosphate carboxylase/ oxygenase large chain gene (*rbc*L) were amplified and subsequently bidirectional sequenced. The PCR products were inspected by gel electrophoresis and the PCR product sizes were approximately 800 bp. The partial sequences of from *P. candollei* var. *mirifica, P. candollei* var. *candollei*, *M. macrocarpa* and *B. superba* were 680, 831, 788 and 658 bp, respectively. All *rbc*L sequences of these three Kwao Khruea species were aligned using bioedit and the alignment showed 20 positions (29.6296%) of variable sites obtained from sequence alignment with 41.3333% of GC content. The *rbc*L sequence obtained within same species were identical. The *rbc*L region was the highest %GC content of these three species of Kwao Khruea.

4.3 Bar-HRM analysis of the Kwao Khruea herbal products

There are two loci, *mat*K and ITS, which were suitable for analysis using Bar-HRM method and identified for their nucleotide signatures by DNA alignment. The length of *mat*K sequences was 1527 bp. A short DNA region at nucleotide position 375 to 487 bp was to yield and it was selected for Bar-HRM analysis and amplified by KK*mat*K1 primer set (PCR product length 113 bp) (**Figure 9**). Both varieties of *P. candollei* have 33.6283% of GC content. Also, *M. macrocarpa* and *B. superba* have 32.7433% and 30.673% of GC content, respectively. Six positions of nucleotide polymorphism have been found in the three Kwao Khruea species (**Table 9**). The result of melting temperatures (T_m) of the three Kwao Khruea species were divided into two clusters. The melting temperatures of the first cluster, *P. candollei* var. *mirifica, P. candollei* var. *candollei* and *M. macrocarpa* were 74.4-74.9°C and the melting temperatures of the second cluster, *B. superba* was 73.6-74.0°C. By the way, DNA position 435 to 482 bp was amplified by KK*mat*K2 primer set (PCR product length 47 bp) (**Figure 10**). Both variety of *P. candollei* were contained 40.4255% of GC content while 42.5532% of that *M. macrocarpa* and *B. superba* have 40.4255% of GC content. There is only one nucleotide difference among those three Kwao Khruea species (**Table 10**). The melting temperatures of the three Kwao Khruea species were divided into two clusters. The melting temperatures of the first cluster, *M. macrocarpa* was 71.0-71.4°C while the melting temperatures of the second cluster, *P. candollei* var. *mirifica, P. candollei* var. *candollei* and *B. superba* were 70.7-70.9°C. DNA position 556 to 676 bp was amplified by KK*mat*K3 primer set (PCR product length 114 bp) (**Figure 11**). Both variety of *P. candollei* were contained 24.5614% of GC content while 25.4386% of that *M. macrocarpa* and *B. superba* have 25.8333% of GC content. There are twelve nucleotides difference among those three Kwao Khruea species (**Table 11**). The melting temperatures of the three Kwao Khruea species were divided into three clusters. The melting temperatures of the first cluster, *M. macrocarpa* was 70.7-71.4°C

while the melting temperatures of the second cluster, *P. candollei* var. *mirifica, P. candollei* var. *candollei* was 70.4-71.2 and *B. superba* were 71.5-72.7°C. The ITS signature was located in nucleotide position 479 to 611 bp that covered partial 5.8S region and partial ITS2 region. The nucleotide signature was selected for Bar-HRM analysis and amplified by ITS_KK1 primer set (**Figure 12**). The amplicons were produced using real-time PCR coupled with HRM analysis. There are forty-six positions of nucleotide polymorphism among these three Kwao Khruea species (**Table 12**). The melting temperatures result of three Kwao Khruea species were divided into three clusters. The melting temperatures of the first cluster that contain both varieties of *P. candollei* were 87.2-87.4°C. The melting temperatures of the second cluster, only *M. macrocarpa* was 87.6-88.7°C. The melting temperatures of the third cluster *B. superba* was 85.3-85.8°C.
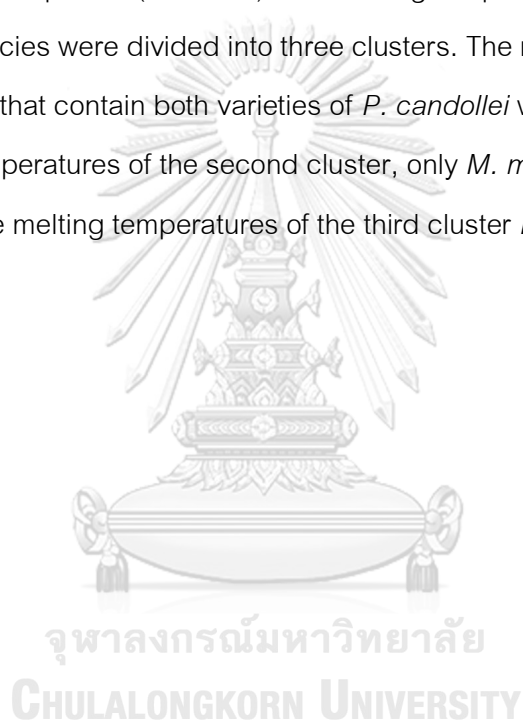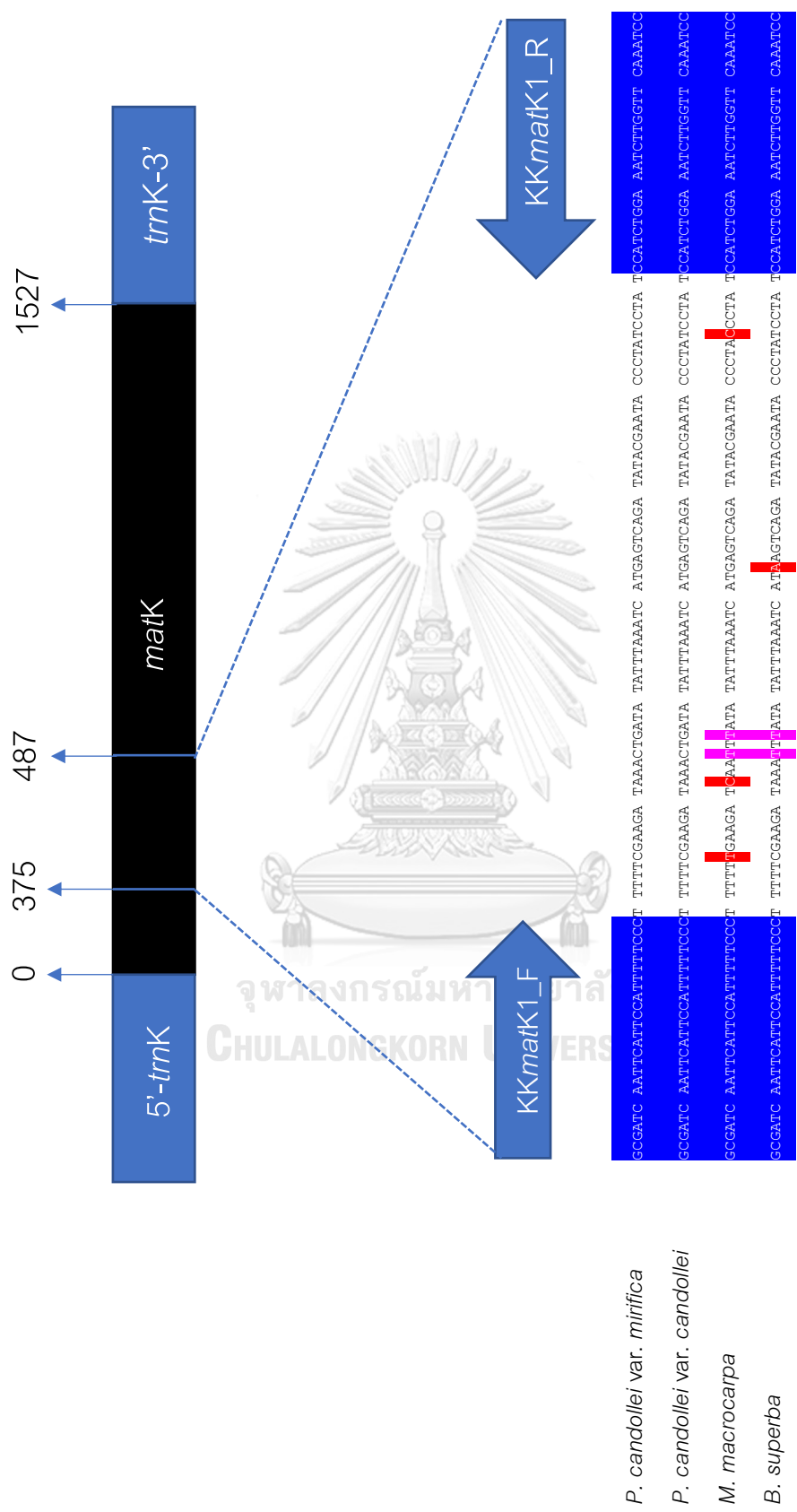
Figure 9 The schematic diagram of the amplification loci and primer pair and their polymorphisms KKmatK1 for real-time PCR
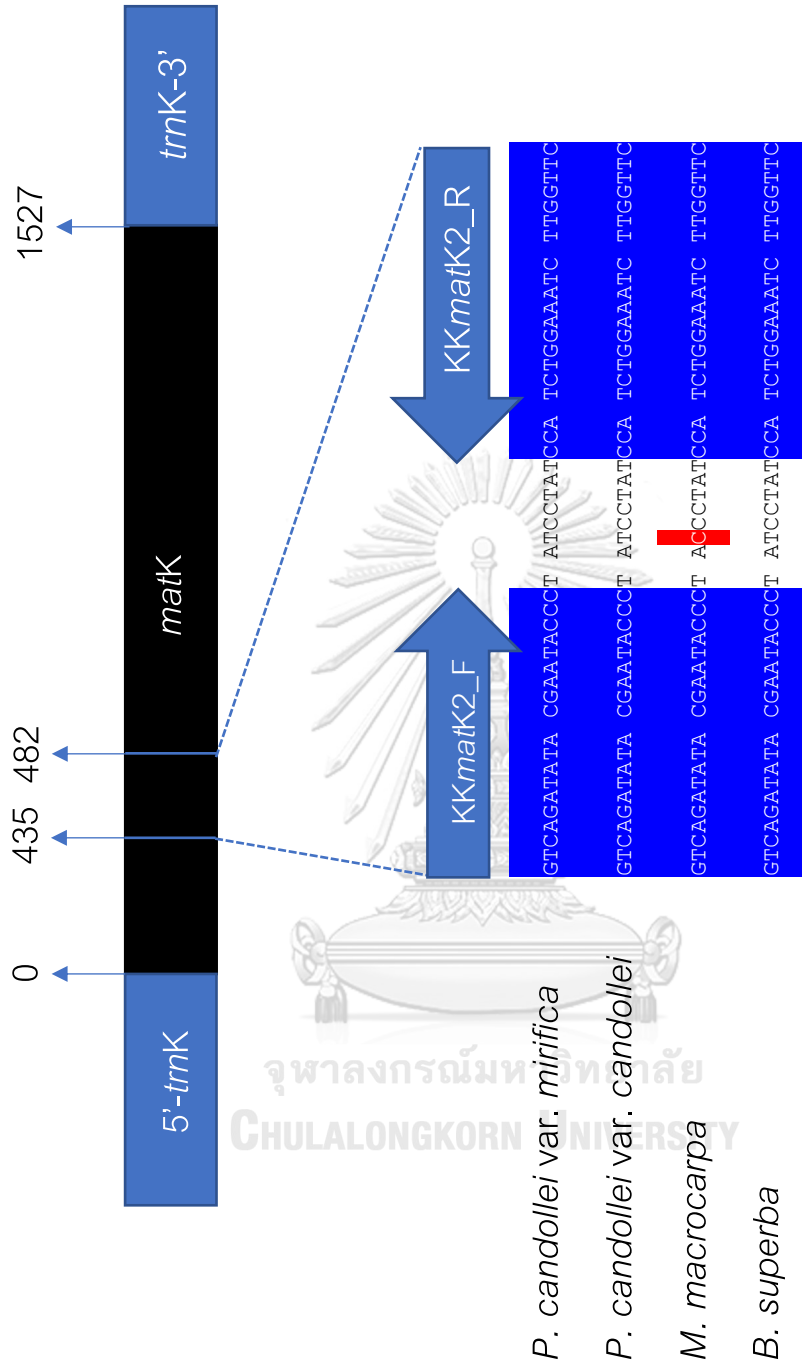
Figure 10 The schematic diagram of the amplification loci and primer pair and their polymorphisms KK*matK*2 for real-time PCR
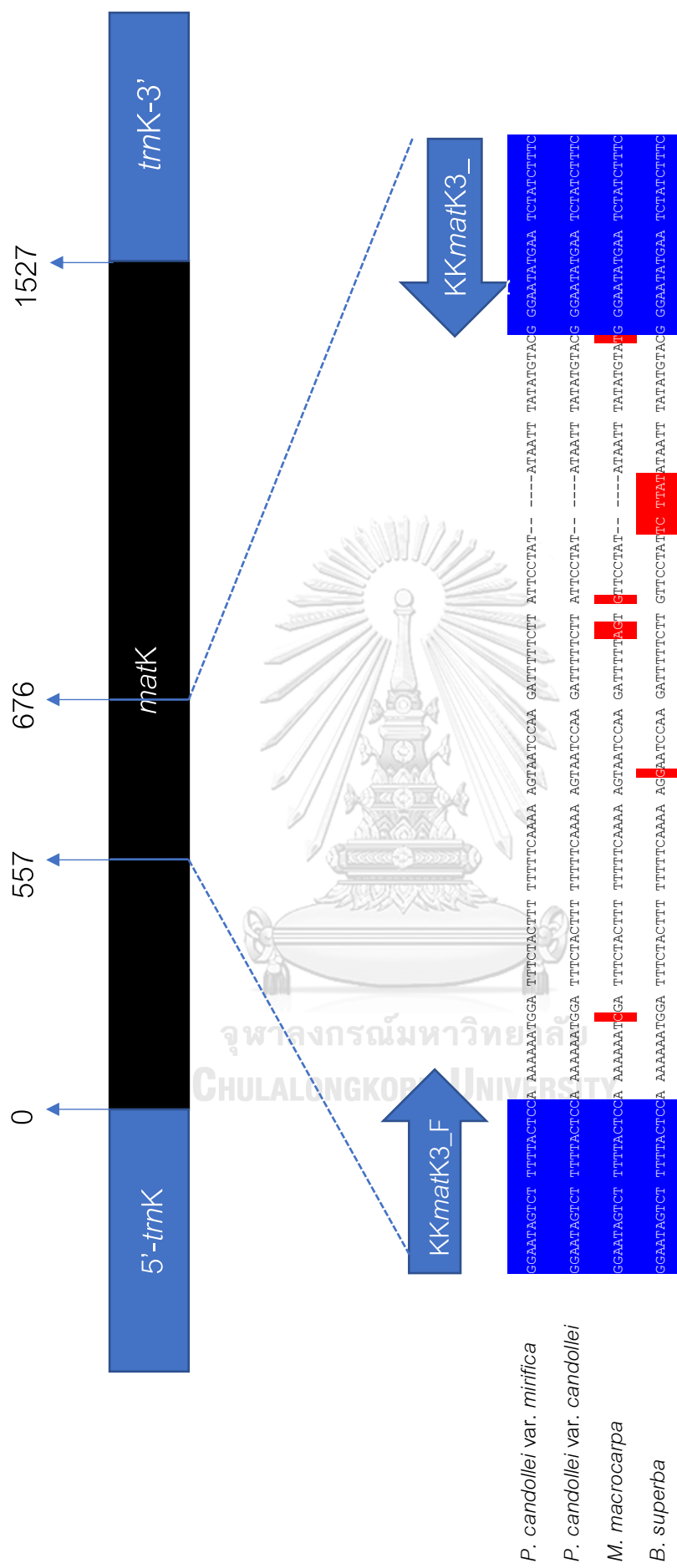
Figure 11 The schematic diagram of the amplification loci and primer pair and their polymorphisms KK*matK*3 for real-time PCR
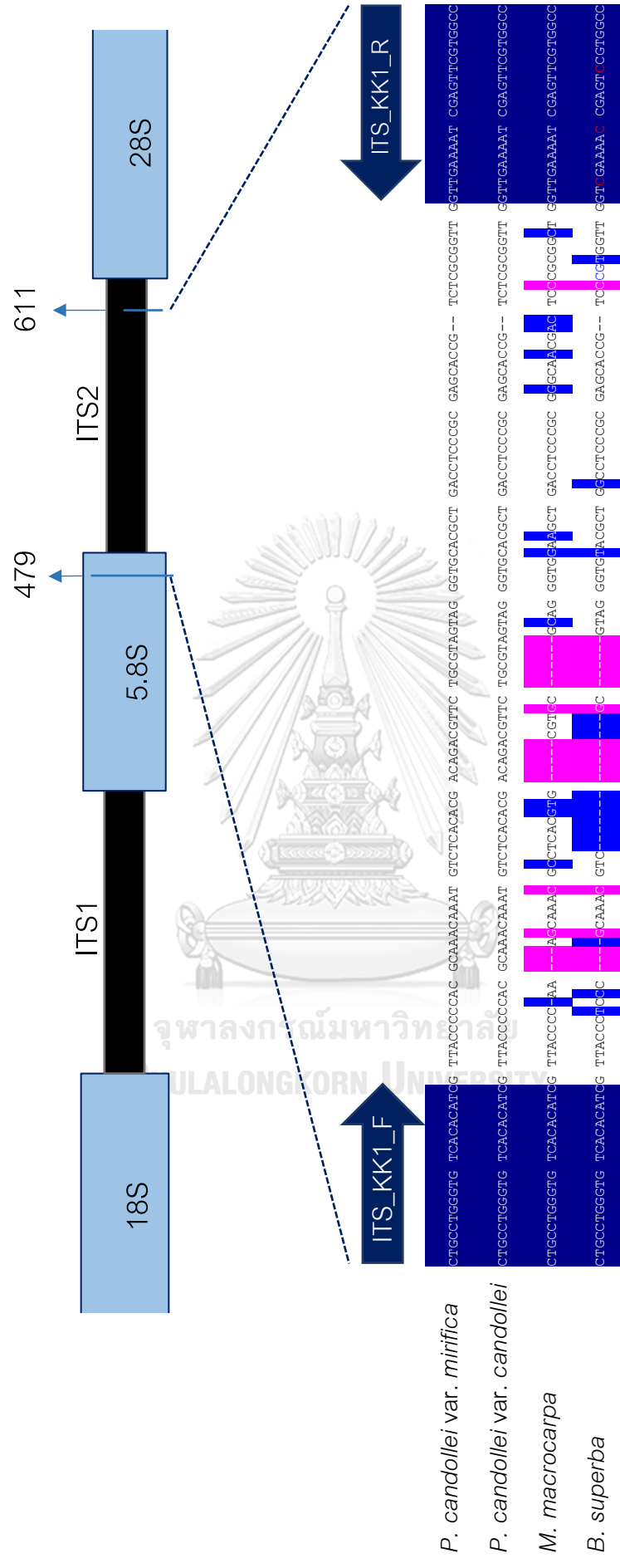
Figure 12 The schematic diagram of the amplification loci and primer pair and their polymorphisms KKmatK1 (A) ITS_KK1 (B) for real-time PCR

Table 9 The differences in polymorphism of three Kwao Khruea species revealed by PCR amplification using KK_*matK1* primer set

Position no. Kwao Khruea species	405	412	415	417	433	456
P. candollei var. *mirifica*	C	A	A	A	G	T
P. candollei var. *candollei*	C	A	A	A	G	T
M. macrocarpa	T	C	T	T	G	C
B. superba	C	A	T	T	A	T

Table 10 The differences in polymorphism of three Kwao Khruea species revealed by PCR amplification using KK_*matK2* primer set

Position no. Kwao Khruea species	456
P. candollei var. *mirifica*	C
P. candollei var. *candollei*	C
M. macrocarpa	T
B. superba	C

Table 11 The differences in polymorphism of three Kwao Khruea species revealed by PCR amplification using KK_*matK*3 primer set

Position no. / Kwao Khruea species	583	608	623	624	626	634	635	636	637	638	639	654
P. candollei var. *mirifica*	G	T	C	T	A	-	-	-	-	-	-	C
P. candollei var. *candollei*	G	T	C	T	A	-	-	-	-	-	-	C
M. macrocarpa	C	T	A	G	G	-	-	-	-	-	-	T
B. superba	G	G	C	T	A	T	C	T	T	A	T	C

Table 12 The differences in polymorphism of three Kwao Khruea species revealed by PCR amplification using ITS_KK1 primer set

Position no. Kwao Khruea species	506	507	508	510	511	512	513	514	519	521	523	524	525	526	527	528	529	530	531	532	533	534	535
P. candollei var. *mirifica*	C	C	A	G	C	A	A	A	T	T	T	C	A	C	A	C	G	T	G	C	G	T	A
P. candollei var. *candollei*	C	C	A	G	C	A	A	A	T	T	T	C	A	C	A	C	G	T	G	C	G	T	A
M. macrocarpa	C	-	A	-	-	-	A	G	C	C	T	C	A	C	G	T	G	-	-	-	-	-	C
B. superba	T	C	C	-	-	-	-	G	C	T	-	-	-	-	-	-	-	-	-	-	-	-	-

Position no. Kwao Khruea species	536	537	538	540	541	542	543	544	545	547	554	556	561	571	575	578	579	582	585	588	593	599	605
P. candollei var. *mirifica*	C	G	T	T	G	C	G	T	A	T	C	C	A	A	C	-	-	T	C	T	T	T	T
P. candollei var. *candollei*	C	G	T	T	G	C	G	T	A	T	C	C	A	A	C	-	-	T	C	T	T	T	T
M. macrocarpa	G	T	G	-	-	-	-	-	-	C	G	A	A	G	A	A	C	C	C	C	T	T	T
B. superba	-	-	G	-	-	-	-	-	-	T	T	C	G	A	C	-	-	C	T	T	C	C	C

The result of HRM analysis was shown in the melting curve patterns. Therefore, the use of KK*mat*K1 primer set was able to exclusively differentiate red Kwao Khruea out from black and white Kwao Khruea (**Figure 13**) and the use of KK*mat*K2 primer set was able to exclusively differentiate black Kwao Khruea out from white and red Kwao Khruea (**Figure 14**). However, KK*mat*K3 primer set (**Figure 15**) and ITS_KK1 primer set (**Figure 16**) was able to differentiate all three Kwao Khruea plant in a single analysis.
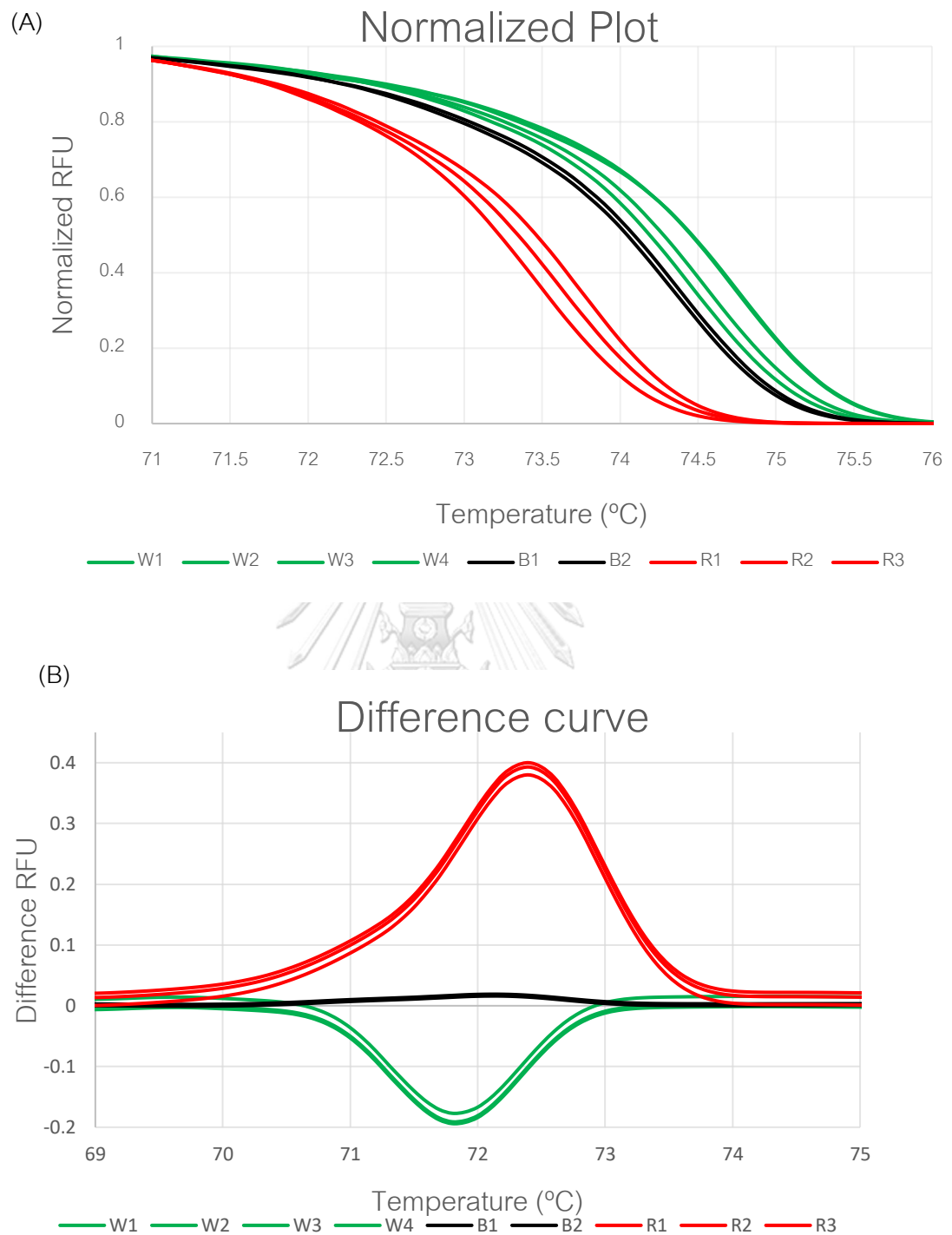
Figure 13 DNA barcoding coupled with HRM analysis of KK*mat*K1 primer set to differentiate two varieties of white (green line), red Kwao Khruea (red line) and black Kwao Khruea (black line), shown in terms of normalized plot (A) and difference plot (B)

(A)

(B)
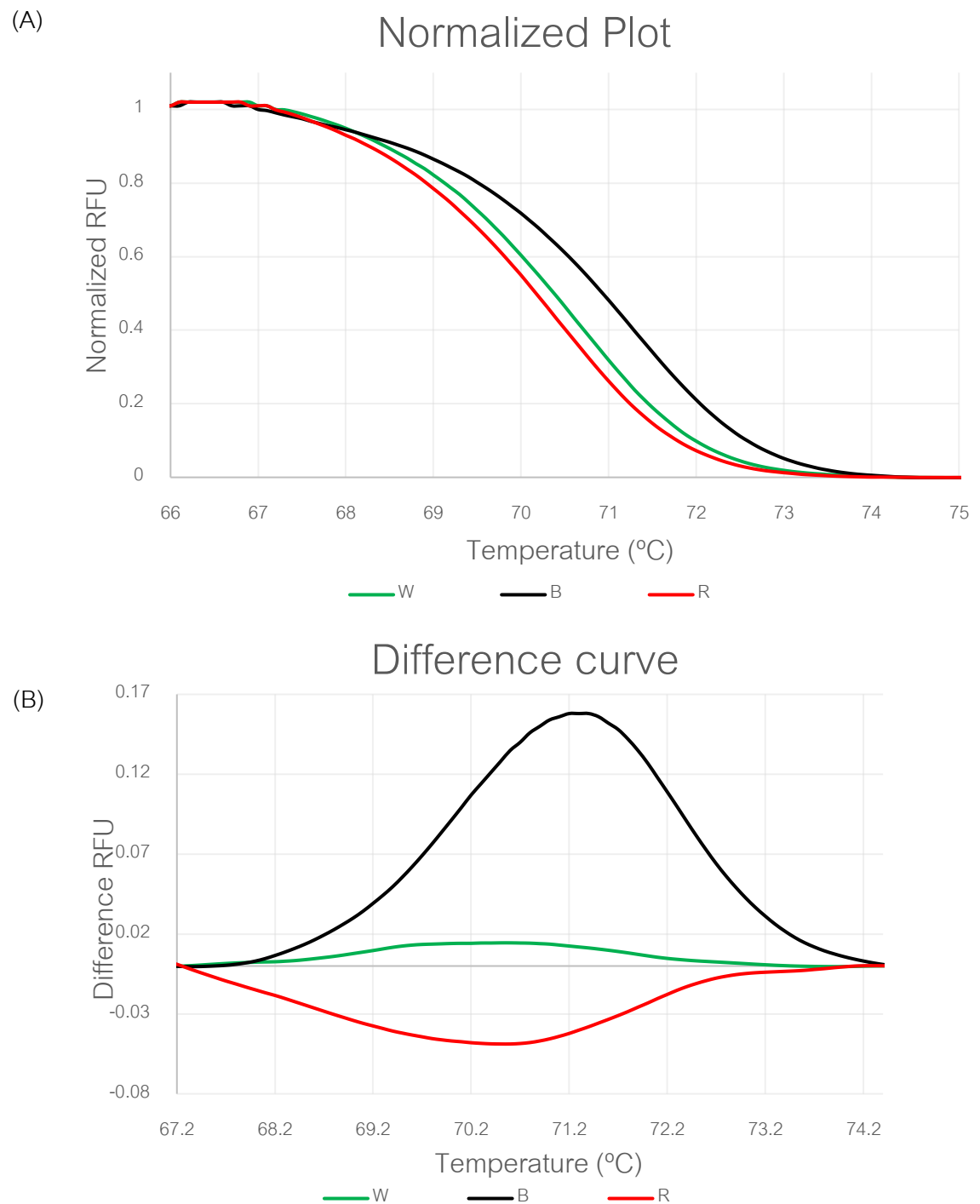
Figure 14 DNA barcoding coupled with HRM analysis of KK*mat*K2 primer set to differentiate two varieties of white (green line), red Kwao Khruea (red line) and black Kwao Khruea (black line), shown in terms of normalized plot (A) and difference plot (B)
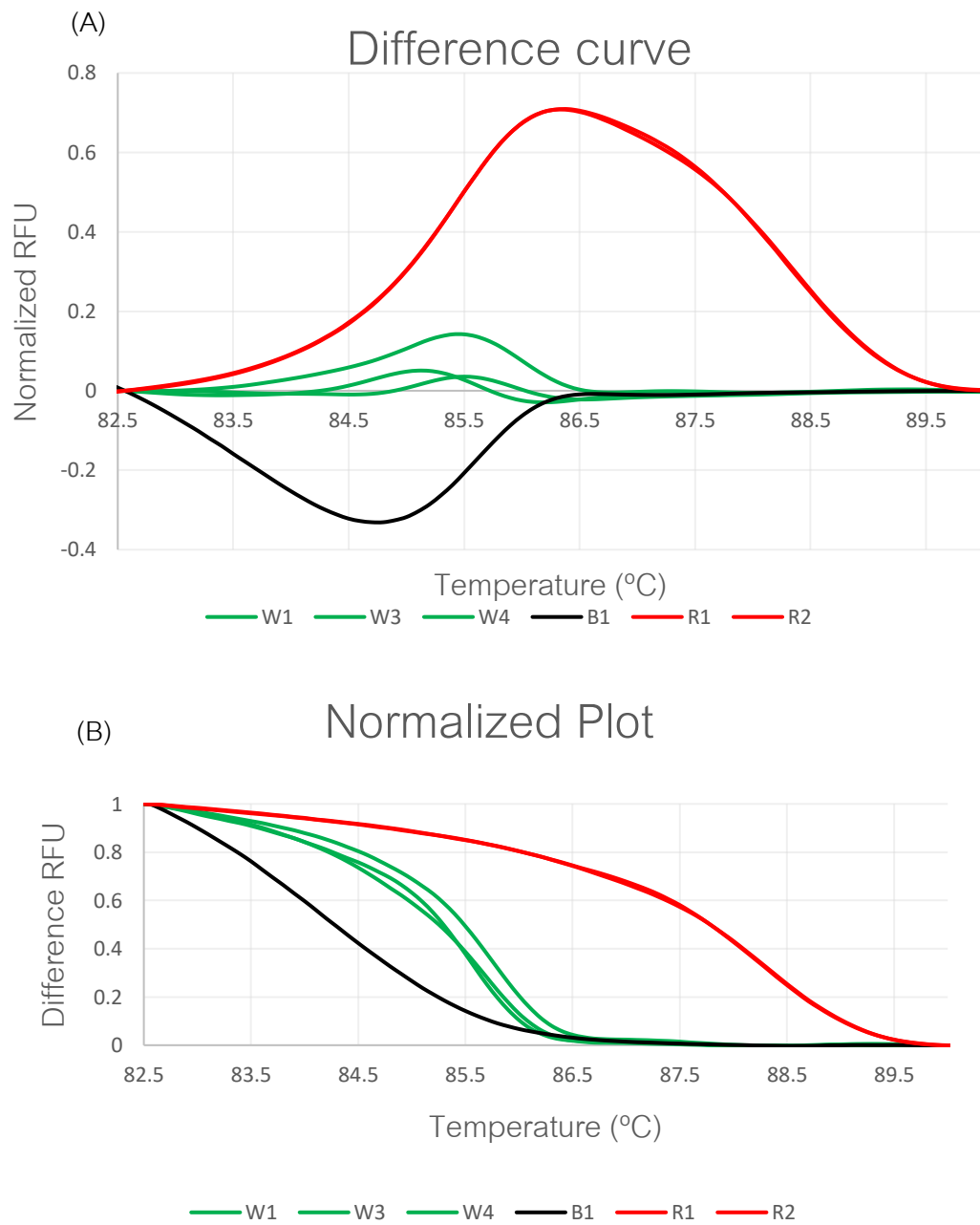
Figure 15 DNA barcoding coupled with HRM analysis of KK*mat*K3 primer set to differentiate two varieties of white (green line), red Kwao Khruea (red line) and black Kwao Khruea (black line), shown in terms of normalized plot (A) and difference plot (B)

(A)

(B)
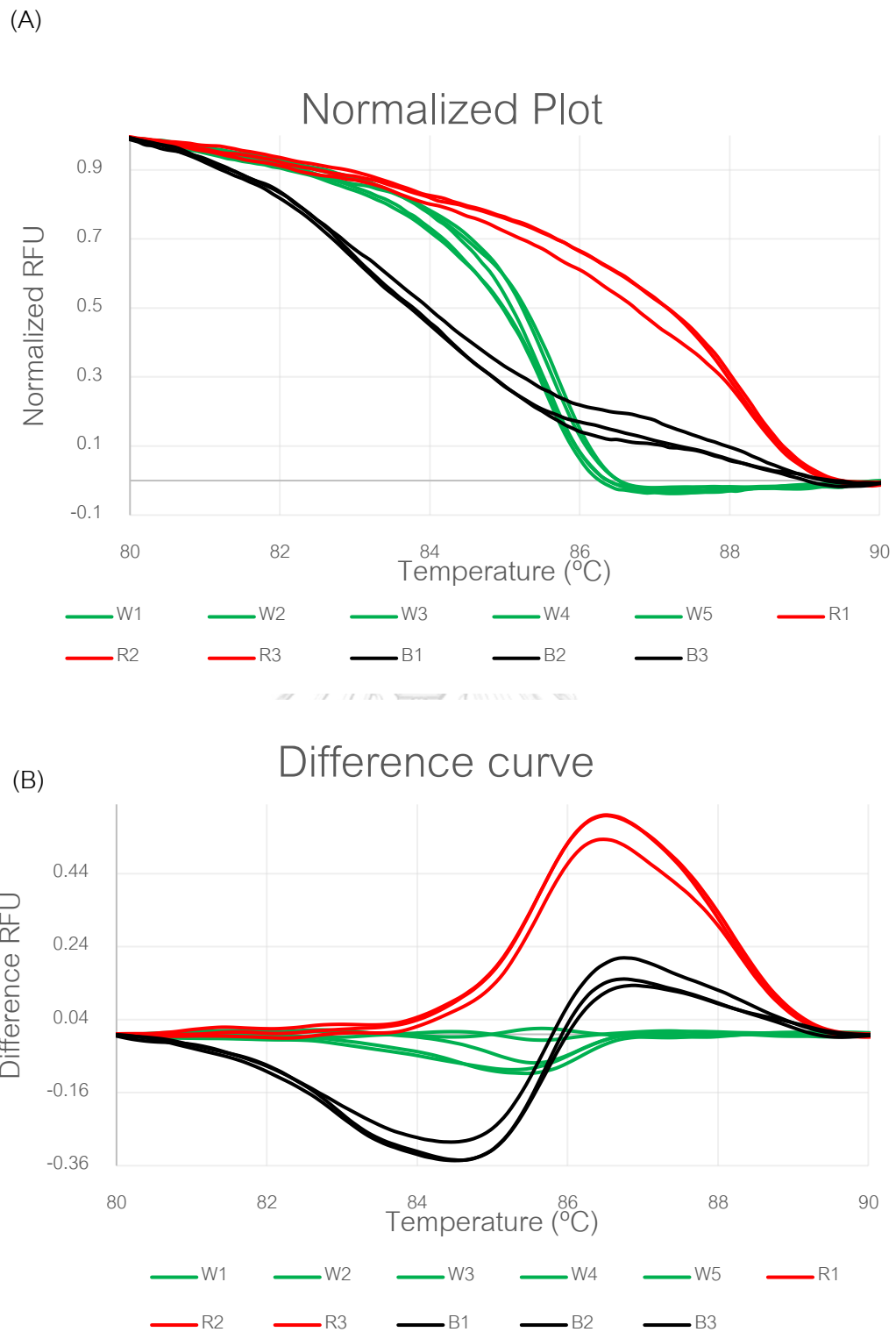
Figure 16 DNA barcoding coupled with HRM analysis of ITS_KK1 primer set to differentiate two varieties of white (green line), red Kwao Khruea (red line) and black Kwao Khruea (black line), shown in terms of normalized plot (A) and difference plot (B)

CHAPTER V
DISCUSSIONS

There has been an extensive confusion in the use of herbs with similar common names including Kwao Khruea. In Thailand, three different Kwao Khruea species have been used in general health promotion and traditional medicine. All of the three species share similar morphology since they are members of the same family. Having both shared vernacular names and some similar characteristics makes it simply misidentify these three Kwao Khruea species. Adverse reactions from the use of misidentified Kwao Khruea have possibly occurred as each Kwao Khruea plant accumulates different chemicals with diverse biological properties. Accordingly, the use of correct herbs at the first step for suitable therapeutic effects with minimal toxicity should be of concern.

Several techniques and methods including plant taxonomy, microscopy, chemical and DNA examination, have been used to help differentiate plant species. In this study, the generation of DNA core barcodes of these three plants was performed. To simplify the differential analysis of the three plants, high resolution melting (HRM) analysis of specific DNA sequences was also carried out. In conclusion, a DNA barcode coupled with HRM (Bar-HRM) analysis for differentiation of the three Kwao Khruea plants was accomplished. Herbal products containing Kwao Khruea as an ingredient were investigated by Bar-HRM analysis and found that some products were questioned to be contained with the incorrect species of Kwao Khruea or other Kwao Khruea contamination.

5.1 The benefit using DNA technique for species differentiate

In the early stages of plant identification analysis, several methods were used to make it possible to differentiate plants. Morphological and microscopic analyses were the two uncomplicated techniques that required particular expertise. Chemical analysis was considered as an approach to help identify plant species. Although specific

chemicals were marked to be representative compounds for identifying species, they were possibly accumulated in another species resulting in misinterpretation.

In the recent decade, the DNA technique was accepted for differentiating the species due to the characteristics of the DNA including unique sequence information of individual species and its stability in various conditions. DNA analysis as one of the basic techniques is simple and applied for investigating all living organisms. Thus, this technique is generally accessed and adapted to detect the unknown plant species (Butler, 2017). The potentiality to obtain an accurate, precise, sensitive and swift DNA analysis is elevated since high technology has been constantly developed. The real-time PCR is one of the developed technologies that allow the investigation of the gene or DNA of interest with high sensitivity and accuracy. It is reported that a small amount of DNA encoding an allergen Lup a 4 at picogram level was detected by the real-time PCR. In fact, the species (Villa, Costa, Gondar, Oliveira, & Mafra, 2018).

5.2 Core DNA barcoding analysis of the three Kwao Khruea herbs

The Consortium for the Barcode of Life (CBOL) is an international initiative devoted to developing DNA barcoding as a global standard for the identification of biological species. CBOL has suggested that using multi-loci of DNA sequences from various compartments, in the case of plants using nuclear and mitochondrial DNA, would make DNA barcoding more reliable. As a promising technique for plant identification, DNA barcoding for the three Kwao Khruea herbs was performed. In this study, three DNA regions from chloroplast including *mat*K, *rbc*L and *psb*A-*trn*H transgenic spacer and one nuclear DNA region, ITS, were found to be suitable for creating DNA barcodes of the three Kwao Khruea herbs as previously recommended (Little, 2014).

Primer sets for DNA barcodes in this study were taken from other research groups and redesigned. This study has shown the completion of DNA barcodes from the four regions of the three Kwao Khruea herbs that have never been reported. All obtained sequence information has been placed and available in the DNA database

(NCBI) that is possibly useful for other researchers. By considering these DNA sequences in detail, it was noteworthy that certain regions of DNA barcodes of the three Kwao Khruea plants exhibited nucleotide signatures which were appropriate for further HRM analysis experiments.

5.3 Bar-HRM analysis for differentiation of the Kwao Khruea herbal products

      Several techniques have been used to analyze PCR amplicons. For example, an analysis of restriction fragment length polymorphism (RFLP) has been used to preliminarily examine the validation of the PCR products prior to DNA sequencing. RFLP is one of the techniques used to identify plant species because it can differentiate PCR products generated from DNA regions of interest of each plant. All three Kwao Khruea plants were previously investigated by PCR-RFLP; however, the RFLP technique was a multi-step and laborious technique (Wiriyakarun et al., 2013). There was another report on the differentiation of the three Kwao Khruesa species by multiplex PCR in admixture (Wiriyakarun, Zhu, Komatsu, & Sukrong, 2014). The result showed that multiplex PCR was effective to differentiate the three Kwao Khruea plants with species-specific probes labeled with different fluorescent dyes.

      The DNA strands are comprised of nucleic acids that are connected and stranded in pairs. This characteristic of DNA allows an investigation of the melting temperature at which the interconnecting bonds between two strands of the DNA are cleaved. Since the PCR amplicons from three different Kwao Khruea plants exhibited their nucleotide signatures in the DNA barcodes, an analysis of the melting temperature was selected as an additional detection technique with an improvement of sensitivity, rapidity, accuracy and simplicity.

      In this study, the partial sequences of *mat*K and ITS were considered as selected areas for Bar-HRM analysis. From the results, two primer sets, KK*mat*K1 and KK*mat*K2, were not able to differentiate all three Kwao Khruea plants at once. The other two sets of primers, KK*mat*K3 and ITS_KK1, were effectively discriminate all three Kwao Khruea in a single experiment. In addition to fresh herbs, herbal products containing

Kwao Khruea were examined by the Bar-HRM technique. Most of the tested herbal products were found to contain correct Kwao Khruea herb as indicated on the label. One herbal product, the herbal product with white Kwao Khruea No.6 (WC6), appeared to be a discrepancy from others. All of the primer sets used to test WC6 showed that the sample was likely to be contaminated with black Pueraria (**Figure 5**). This finding indicated that there was a possibility of an un- or intentional misidentification of Kwao Khruea plants. As Kwao Khruea herbs have been produced as a wide variety of herbal products, the agency in charge of consumer health; for example, FDA needs to help monitor the authenticity of the product for the efficacy and safety of consumers. The Bar-HRM analysis is highly recommended for an inspection of Kwao Khruea herbal products during the manufacturing or to monitor products in the market.

5.4 Advance technique for differentiation the ingredient of Kwao Khruea herbal products

Currently, there has continuously been the development of advanced technology in DNA analyses; for example, Next-generation sequencing (NGS). Most highly developed DNA analysis techniques allow for high throughput experiments in order to obtain readings from more sequences or more samples or very complex samples. The use of techniques for plant differentiation depends on the potentiality of an individual laboratory. This study showed that although the Bar-HRM analysis was not the most advanced technology, it was a sensitive, rapid and applicable technique for identifying the three Kwao Khruea plants in Thailand.
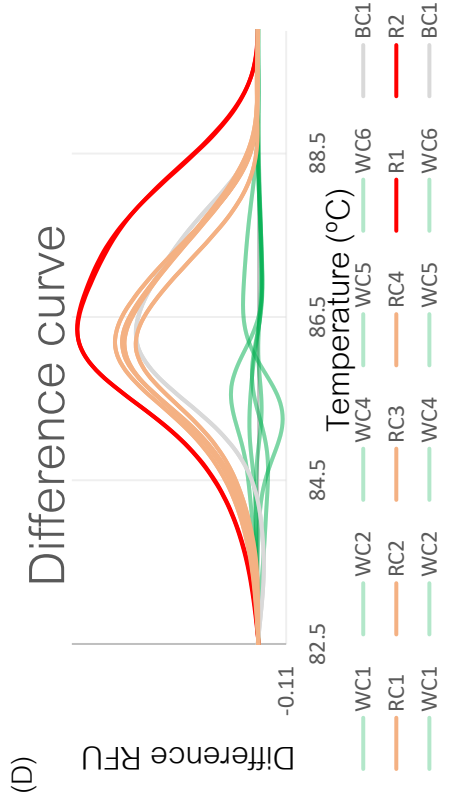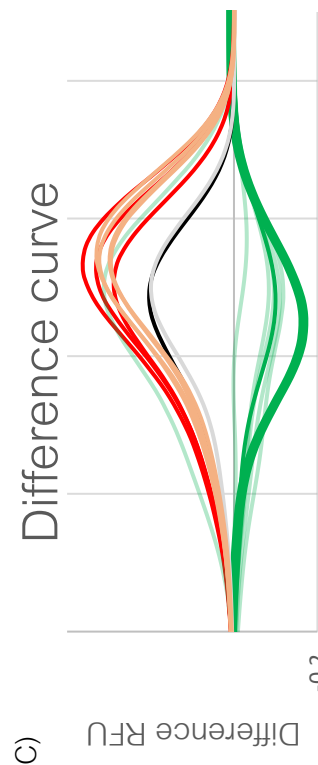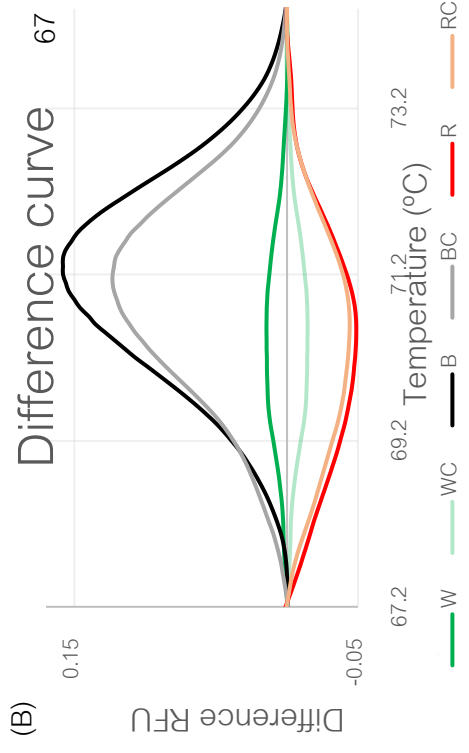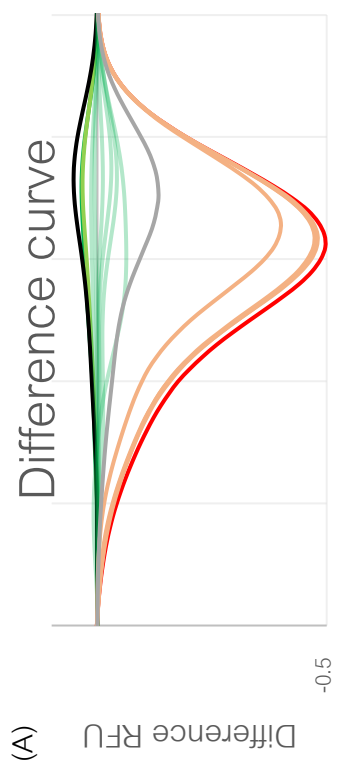
Figure 17 Merge of authentic and herbal product of DNA barcoding coupled with HRM analysis by KK*matK*1 (A), KK*matK*2 (B), KK*matK*3 (C) and ITS_KK1 (D) primer set to differentiate two varieties of white (green line), red Kwao Khruea (red line) and black Kwao Khruea (black line), shown in terms of difference plot

CHAPTER VI

CONCLUSIONS

Kwao Khruea herbs are one of the most useful herbal products, there are three species of Kwao Khruea that Thai medical used, white Kwao Khruea (*P. candollei*), black Kwao Khruea (*M. macrocarpa*) and red Kwao Khruea (*B. superba*). Different species of Kwao Khruea have different medical properties. The right use would bring the most benefit and the right result that suits customers' needs. Wrong use brings the wrong result and may provide other side effects to the customer. To make the confidence for the costumer, using DNA technology to detect individual species is offered as promising method.

In this study, we investigate the DNA sequence of four DNA core barcode area, *mat*K, *rbc*L, *psb*A-*trn*H transgenic spacer and ITS, for differentiation of plants samples and deposit the sequences data to Genbank already. The four areas of DNA barcode of three Kwao Khruea species were generated and analyzed by finding the nucleotide signature that is the uniqueness of individual Kwao Khruea species then applied with HRM analysis technique to differentiate both Kwao Khruea fresh herb and herbal products. Bar-HRM was the technique that differentiates the species by melting temperature of the different ratio of nucleotide in the sequences for rapid differentiation. DNA of *mat*K and ITS were selected to represent chloroplast DNA and nuclear DNA respectively. KK*mat*K1 primer set can differentiate red Kwao Khruea out of white and black Kwao Khruea. KK*mat*K2 primer set can differentiate black Kwao Khruea out of white and red Kwao Khruea. KK*mat*K3 primer set can differentiate all three species of Kwao Khruea individually. ITS_KK1 primer set can differentiate all three species of Kwao Khruea individually. However, when used KK*mat*K3 primer set detected Kwao Khruea product WC6 it gives the confusing result that wasn't consistent with other results from other primer sets. Otherwise, using DNA barcoding was successfully differentiate among the three species of Kwao Khruea plants. This technique can be applied to

identify the other crude drug in the market to be confirmed to the customer that all Kwao Khruea products are the right species as its labeling.

For further study, collecting more species in the same genus of each Kwao Khruea plant is recommended to test the specificity of the Bar-HRM technique. The greater sample numbers permit the more reliability of using this technique. To transform this technique into a user-friendly application, a development of a test kit set for identifying Kwao khruea species is proposed. The idea of the kit set based on the Bar-HRM technique is to include the Kwao Khruea species. This technique would be ofbenefit for both customers and producers to examine Kwao Khruea raw materials or herbal products.

Rapid differentiation of Kwao Khruea species can improve and control the quality of the Kwao Khruea products from agricultural to manufacturing purposes. This Bar-HRM analysis is applicable not only for Kwao Khruea herbal products but also for other herbal products to perform good quality control and safety surveillance for consumers.

APPENDIX

SEQUENCE ALIGNMENTS OF KWAO KHRUEA PLANTS

```
                     10        20        30        40        50        60        70        80        90       100
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su    1   ATGGAGGAATATCGAATATATTTAGAACTCCATAGATCTCGCCACCAGGACATCTTATACCCGCTTTTTTTTCGGGAATATATTTATGGACTCGTTTATG
B_superb    1   --------------------------------------------------------------------------------------------------
P_v_miri    1   ..............GC.........................A..........C.C......A..........................A.C.....
P_v_miri    1   ..............GC.........................A..........C.C......A..........................A.C.....
P_v_miri    1   ..............GC.........................A..........C.C......A..........................A.C.....
P_v_cand    1   ..............GC.........................A..........C.C......A..........................A.C.....
P_v_cand    1   ..............GC.........................A..........C.C......A..........................A.C.....
Pcv_miri    1   --------------------------------------------------------------------------------------------------
Pcv_cand    1   --------------------------------------------------------------------------------------------------
Mucuna_m    1   ...........T.G......G...................................C.......A............G..............C.....
M_macroc    1   --------------------------------------------------------------------------------------------------

                    110       120       130       140       150       160       170       180       190       200
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su  101   GTCATGGGTCCATTTTTGTAGAAAAAAATGTAGGTTATAACAAAAAATTTAGTTTACTAATTGTAAAACGTTTAATTACTCGAATGTATCAACAGACTCA
B_superb    1   --------------------------------------------------------------------------------------------------
P_v_miri  101   .....................................T..............................AG...........................
P_v_miri  101   .....................................T...............................G...........................
P_v_miri  101   .....................................T...............................G...........................
P_v_cand  101   .....................................T...............................G...........................
P_v_cand  101   .....................................T...............................G...........................
Pcv_miri    1   --------------------------------------------------------------------------------------------------
Pcv_cand    1   --------------------------------------------------------------------------------------------------
Mucuna_m  101   ......A.G..............---...........TC................AG...........................G.........
M_macroc    1   --------------------------------------------------------------------------------------------------

                    210       220       230       240       250       260       270       280       290       300
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su  201   TTTGATCATTTTTGCTAATGATTCTAACAAAAATCCTTTTTGGGGTTATAATAATAATTTTTTATTCTCAAATAATATTCGAAGGTTTTGTTGTCGTCGTG
B_superb    1   --------------------------------------------------------------------------------------------------
P_v_miri  201   ...C.........A...C.....................A.........C...C..............A.........C..G.......
P_v_miri  201   ...C.........A...C.....................A.........C...C..............A...........G.......
P_v_miri  201   ...C.........A...C.....................A.........C...C..............A...........G.......
P_v_cand  201   ...C.........A...C.....................A.........C...C..............A...........G.......
P_v_cand  201   ...C.........A...C.....................A.........C...C..............A...........G.......
Pcv_miri    1   --------------------------------------------------------------------------------------------------
Pcv_cand    1   --------------------------------------------------------------------------------------------------
Mucuna_m  198   ...........G.T.................A.................C........GG........A.........A.......
M_macroc    1   --------------------------------------------------------------------------------------------------

                    310       320       330       340       350       360       370       380       390       400
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su  301   GAGATTCTATTTTCCCTACAATTATTTATTCCCTTCCTTAAAGGATTTAGAAATCGTAAAATCTTATAATAATTTGCGATCAATTCATTCCATTTTTCCCT
B_superb    1   --------------------------------------------------------------------------------------------------
P_v_miri  301   .........................T.........G...A.......................................
P_v_miri  301   .........................T.........G...A.......................................
P_v_miri  301   .........................T.........G...A.......................................
P_v_cand  301   .........................T.........G...A.......................................
P_v_cand  301   .........................T.........G...A.......................................
Pcv_miri    1   --------------------------------------------------------------------------------------------------
Pcv_cand    1   --------------------------------------------------------------------------------------------------
Mucuna_m  298   .........................T...T.........G..........................................
M_macroc    1   --------------------------------------------------------------------------------------------------

                    410       420       430       440       450       460       470       480       490       500
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su  401   TTTTCGAAGATAAATTTATATATTTAAATCATAAGTCAGATATACGAATACCCTATCCTATCCATCTGGAAATCTTGGTTCAAATCCTTCGACATTGGAT
B_superb    1   .........................................................................................
P_v_miri  401   ..............C.G...........G...................................T......
P_v_miri  401   ..............C.G...........G...................................T......
P_v_miri  401   ..............C.G...........G...................................T......
P_v_cand  401   ..............C.G..C........G...................................T......
P_v_cand  401   ..............C.G...........G...................................T......
Pcv_miri    1   .....................G...................................T......
Pcv_cand    1   .....................G...................................T......
Mucuna_m  398   ...T......C............G...............C..................T......
M_macroc    1   .-...................G..............C.................T......

                    510       520       530       540       550       560       570       580       590       600
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_su  501   AAAAGATGTTTCTTTCTTTCATTTATTAAGATGTGTTTTTTTATTACTATTGTAATGGAATAGTCTTTTTACTCCAAAAAAATGGATTTCTACTTTTTTT
B_superb   83   .........................................................................................
P_v_miri  501   ..........C................G...........A......................
P_v_miri  501   ..........C................G...........A..................A...........
P_v_miri  501   ..........C................G...........A......................
P_v_cand  501   ..........C................G...........A......................
P_v_cand  501   ..........C................G...........A......................
Pcv_miri   83   ..........C................G...........A......................
Pcv_cand   83   ..........C................G...........A......................
Mucuna_m  498   ..........C................G.....................C...........
M_macroc   83   ..........C................G.....................C...........
```

Figure S1 Sequence alignment of *mat*K of Kwao Khruea plants. The number above the

alignment represents the base numbers of sequence alignments. "." represents

nucleotide base that is identical to the nucleotide in the first-row. "-" represents the gap

Figure S1 Continued

Figure S1 Continued

```
                 10        20        30        40        50        60        70        80        90       100
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   1 ----------------------------------------------------------------------------------------------------
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro   1 ----------------------------------------------------------------------------------------------------
B1_ITS1   1 CCCCCGGCAATTTTACTCCCGAGGGGGTTGGGGGGGGATGATACCCCCACCCCCCCCGGGGGGGTCCCCTTAACCCTAAAGGGTTTGGGGCCCACTTTGT
B1_ITS4   1 ----------------------------------------------------------------------------------------------------
B_super   1 ----------------------------------------------------------------------------------------------------
R1_ITS1   1 ----------------------------------------------------------------------------------------------------
R1_ITS4   1 ----------------------------------------------------------------------------------------------------

                110       120       130       140       150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   1 ----------------------------------------------------------------------------------------------------
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro   1 ----------------------------------------------------------------------------------------------------
B1_ITS1 101 CTTCAAAGGATCTGAGGGTTCAGGGGTTTTCGCAAATTCCCCCACCCAAGTATTGCCTTTTTCGTTATGTTTTTTACATGAAGCAAAGCAAATCATTCTT
B1_ITS4   1 ----------------------------------------------------------------------------------------------------
B_super   1 ----------------------------------------------------------------------------------------------------
R1_ITS1   1 ----------------------------------------------------------------------------------------------------
R1_ITS4   1 ----------------------------------------------------------------------------------------------------

                210       220       230       240       250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   1 ----------------------------------------------------------------------------------------------------
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro   1 ----------------------------------------------------------------------------------------------------
B1_ITS1 201 GCCAATTCTTTTTGATTGTGTTGAACAAGTGCGGATGCCTTTCTCCGGCCCGAGATTGCACCGAAACATTTGGATTCCTTGCCCAGAAGCCCGGGTTTGT
B1_ITS4   1 ----------------------------------------------------------------------------------------------------
B_super   1 ----------------------------------------------------------------------------------------------------
R1_ITS1   1 ----------------------------------------------------------------------------------------------------
R1_ITS4   1 ----------------------------------------------------------------------------------------------------

                310       320       330       340       350       360       370       380       390       400
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   1 ----------------------------------------------------------------------------------------------------
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro   1 ----------------------------------------------------------------------------------------------------
B1_ITS1 301 TTGACGGAGGGCCGGCGACCGCCTTCCCGCCAACCCCCCGAGGGGGACGGGTCTTCGTTGAGACCGAACGCGCCCCGAGCCGCCCCGGGGGCTAGATGT
B1_ITS4   1 ----------------------------------------------------------------------------------------------------
B_super   1 ----------------------------------------------------------------------------------------------------
R1_ITS1   1 ----------------------------------------------------------------------------------------------------
R1_ITS4   1 ----------------------------------------------------------------------------------------------------

                410       420       430       440       450       460       470       480       490       500
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   1 ------TAACAAGGTTTCCGTAGGTGAACCTTCGTAAGGATCATTGTCGATGCCTCACAATCAGATTGACCCGCGAATTCGTTTTTCCATCTACCAGCA
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro   1 ------------------------------------------T.ACAAGG..TC..TA.GTGA..CTG.G..AG..T.ATT.TCGAT.CC.CG.ACAC.GATT.G
B1_ITS1 401 AAAAGCGTTCCCCGTCGAACCG.TTGT.C..G.CATT.ACAAGG..TC..TA.GTGA..CTG.G..AG..T.ATT.TCGAT.CC.CG.ACAACC..TT..
B1_ITS4   1 ........................................................................T.ATT.TCGAT.CC.CG.ACAACC..TT..
B_super   1 ------------------------------------------T.ACAAGG..TC..TA.GTGA..CTG.G..AG..T.ATT.TCGAT.CCCC--ACGA.C.ACT..
R1_ITS1   1 ----------------------------------------------------------------------------------------------C.ACT..
R1_ITS4   1 ------------------------.TTC.T.TCAACT.ACAAGG..TC..TA.GTGA..CTG.G..AG..T.ATT.TCGAT...CC--ACGA.C.ACT..

                510       520       530       540       550       560       570       580       590       600
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando  94 GGCAAGCAGGGTCGGGCCGGGCTTCTTCCGTCCGTCCTCCCCTT.CTTTGCCTGTTGCGTTGG.GCG.GGG..TGGACACAAAGCAAGCATCGTCTTGC
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro  65 AC.GGCG.ACCCGTTTTA.CATCTACGC..AC.G.GG.CGG.TCG.GGGC...GT.C.GTC.C.AA.GAC.ACCCCGT.C.CCCT.GG.GGTT.G.GG.A
B1_ITS1 501 AC.GGCG.ACCCGTTTTA.CATCT.CG...AC.G.GG.CGG.TCG...C...GT.C.GTC.C.AA.GAC.ACCCCGT.C.CCCT.GG.GGTT.G.GG.A
B1_ITS4  33 AC.GGCG.ACCCGTTTTA.CATCT.CG...AC.G.GG.CGG.TCG...C...GT.C.GTC.C.AA.GAC.ACCCCCGTC.CCCT.GG.GGTT.G.GG.A
B_super  63 AC.GGCG.ACCCGTTTT--------GT..AC...GGTCGGGG.........----C.GCGCCC----C..CCC.G-.C.CTGT.GG.GG.T.G.CG.G
R1_ITS1   8 AC.GGCG.ACCCGTTTT--------GT..AC...GGTCGGGG.........----C.GCGCCC----C..CCC.G-.C.CTGT.GG.GG.T.G.CG.G
R1_ITS4  75 AC.GGCG.ACCCGTTTT--------GT..AC...GGTCGGGG.........----C.GCGCCC----C..CCC.G-.C.CTGT.GG.GG.T.G.CG.G

                610       620       630       640       650       660       670       680       690       700
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando 194 CTGGCCTCCTCTCCTCCTTCCCAACAAACACAACCCCCGGCGCTTCGTGCGCCAAGGAACTCCAATCTGTTAAGTGCGACTCCCAACGGCCCGGAGACGG
W1_ITS-   1 ----------------------------------------------------------------------------------------------------
W1_ITS4   1 ----------------------------------------------------------------------------------------------------
W2_ITS-   1 ----------------------------------------------------------------------------------------------------
W2_ITS4   1 ----------------------------------------------------------------------------------------------------
M_macro 165 GGC.GTCG.C.CG.G.GCC.CCTC.CGT.GA.CA.AAACCCCGGCGCTTC.TGCGCC..GGAACTCGAAATTGT.CG.TGCAATCT.C..GG.CCCGGA.
B1_ITS1 601 GGC.GTCG.C.CG.G.GCC.CCTC.CGT.GA.CA.AA.C...GGCGCTTC.TGCGCC..GGAACTCGAAATTGT.CG.TGCAATCT.C..GG.CCCGGA.
B1_ITS4 133 GGC.GT----------------------------------------------------------------------------------------------
B_super 145 TCTCG.A-------CGGCCTCCTCTCGA.CA.CA.AA.C...GGCGCTTC.CGCGCC..GGAATTCGAAATTGC.CG.TGC.GTC-.C.GGGACCCGGA.
R1_ITS1  90 TCTCG.A-------CGGCCTCCTCTCGA.CA.CA.AA.C...GGCGCTTC.CGCGCC..GGAATTCGAAATTGC.CG.TGC.GTC-.C.GGGACCCGGA.
R1_ITS4 157 TCTCG.A-------CGGCCTCCTCTCGA.CA.CA.AA.C...GGCGCTTC.CGCGCC..GGAATTCGAAATvGC.CG.TGC.GTC-.C.GGGACCCGGA.
```

Figure S2 Partial sequence alignment of ITS of Kwao Khruea plants. The number above the alignment represents the base numbers of sequence alignments. "." represents nucleotide base that is identical to the nucleotide in the first-row. "-" represents the gap

```
                    710       720       730       740       750       760       770       780       790       800
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   294 TGCCCCGCGGGCGTTGTCACGACGCACATTATTATACAAAATGACTCTCGGCAACGGATATCTCGGCTCTTGCATCGATGAAGAACGTAGCGAAATGCGA
W1_ITS-   1   --------------------------------------------------------------------------------------------........
W1_ITS4   1   ---------------------------------------------------------------------------------------------.......
W2_ITS-   1   ----------------------------------------------------------------------------------------------------
W2_ITS4   1   --------------------------------------------------------------------------------------------........
M_macro   265 AC.G.GATCCCGCGGAC.TT.T.A.GACAC.CG....................................................................
B1_ITS1   701 AC.G.GATCCCGCGGAC.TT.T.A.GACAC.CG....................................................................
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   237 AC.GTGTTCCCGCGGGT.TC.T.A.GACAC.CG...T.G..C...........................................................
R1_ITS1   182 AC.GTGTTCCCGCGGGT.TC.T.A.GACAC.CG...T.G..C...........................................................
R1_ITS4   249 AC.GTGTTCCCGCGGGT.TC.T.A.GACAC.CG...T.G..C...........................................................
                    810       820       830       840       850       860       870       880       890       900
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   394 ....................................................................TG.........C.........T.......A
W1_ITS-   9   ....................................................................TG.........C.........T.......A
W1_ITS4   11  ....................................................................TG.........C.........T.......A
W2_ITS-   1   ----------------------------------------------------------------------TG.........C.........T.......A
W2_ITS4   8   ....................................................................TG.........C.........T.......A
M_macro   365 .............................................................................C.........T.......A
B1_ITS1   801 .............................................................................C.........T.......A
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   337 ..............................................................G.............C.........T.......A
R1_ITS1   282 ..............................................................G.............C.........T.......A
R1_ITS4   349 ..............................................................G.............C.........T.......A
                    910       920       930       940       950       960       970       980       990      1000
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   494 .....TTA.C...A.GCA.ACAAAT.T.TCAC.CG.CA...GTT.T.C.TA..A---...T.C.CG...A.......C.A..A...T-.T.........T
W1_ITS-   109 .....TTA.C...A.GCA.ACAAAT.T.TCAC.CG.CA...GTT.T.C.TA..A---...T.C.CG...A.......C.A..A...T-.T.........T
W1_ITS4   111 .....TTA.C...A.GCA.ACAAAT.T.TCAC.CG.CA...GTT.T.C.TA..A---...T.C.CG...A.......C.A..A...T-.T.........T
W2_ITS-   76  .....TTA.C...A.GCA.ACAAAT.T.TCAC.CG.CA...GTT.T.C.TA..A---...T.C.CG...A.......C.A..A...T-.T.........T
W2_ITS4   108 .....TTA.C...A.GCA.ACAAAT.T.TCAC.CG.CA...GTT.T.C.TA..A---...T.C.CG...A.......C.A..A...T-.T.........T
M_macro   465 .....TCG.C...AAA...AA------------------CGCC.CA.----.T...TGCG.A..GTGGAA.CTGA..T.CC-GCG.GC.AC.T.
B1_ITS1   901 .....TCG.C...AAA...AA------------------CGCC.CA.----.T...TGCG.A..GTGGAA.CTGA..T.CC-GCG.GC.AC.T.
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   437 .....TTA.C.T...GCA.AC------------------..C.C..A----...T.T.CG.............C.A..A...T-.C..T......T
R1_ITS1   382 .....TTA.C.T...GCA.AC------------------..C.C..A----...T.T.CG.............C.A..A...T-.C..T......T
R1_ITS4   449 .....TTA.C.T...GCA.AC------------------..C.C..A----...T.T.CG.............C.A..A...T-.C..T......T
                   1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   590 TG...AT.....T.GT...CGA.TG..C.GT..T..AA.......A..GCAA-CG...GA.ACCAA..A..CGC.GC.A.TCGGT.GGC.CC..A....
W1_ITS-   205 TG...AT.....T.GT...CGA.TG..C.GT..T..AA.......A..GCAA-CG...GA.ACCAA..A..CGC.GC.A.TCGGT.GGC.CC..A....
W1_ITS4   207 TG...AT.....T.GT...CGA.TG..C.GT..T..AA.......A..GCAA-CG...GA.ACCAA..A..CGC.GC.A.TCGGT.GGC.CC..A....
W2_ITS-   172 TG...AT.....T.GT...CGA.T...C.GT..T..AA.......A..GCAA-CG...GA.ACCAA..A..CGC.GC.A.TCGGT.GGC.CC..A....
W2_ITS4   204 TG...AT.....T.GT...CGA.T...C.GT..T..AA.......A..GCAA-CG...GA.ACCAA..A..CGC.GC.A.TCGGT.GGC.CC..A....
M_macro   536 .CGCGGCT.GTCGA.AAT...GTCCGTGGC..GGTGC.CC.C.A.AC.TGG--GGACGAGCG.CGC...A.A.C---.AT..CGCGCGACCCC.T.CG.
B1_ITS1   972 .CGCGGCT.GTCGA.AAT...GTCCGTGGC..GGTGC.CC.C.A.AC.TGG--GGACGAGCG.CGC...A.A.C---.AT..CGCGCGACCCC.T.CG.
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   508 .G...AC.......GT...CGA..GA.C......../..GA......A..GGCAA-CG...GA.ACCG....T.CGC---.A..CGGC.CG..GC..A...T
R1_ITS1   453 .G...AC.......GT...CGA..GA.C......../..GA......A..GGCAA-CG...GA.ACCG....T.CGC---.A..CGGC.CG..GC..A...T
R1_ITS4   520 .G...AC.......GT...CGA..GA.C......../..GA......A..GGCAA-CG...GA.ACCG....T.CGC---.A..CGGC.CG..GC..A...T
                   1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   688 T.C..GACCC.A.AGTGCCTCYTCT..G.GGA.ACG.T.TCAACG..A.CTC.GG-------------------------------------------
W1_ITS-   303 T.C..GACCC.A.AGTGCCTCYTCT..G.GGA.ACG.T.TCAACG..A.CTC.GG.AAGA..CGA.G..CT.C--GGT----------------------
W1_ITS4   305 T.C..GACCC.A.AGTGCCTCTTCT..G.GGA.ACG.T.TCAACG..A.CTC.GGT.AG..-G.GG.TCC.C--GG.T.---------------------
W2_ITS-   270 T.C..GACCC.A.AGTGCCTCTTCT..G.GGA.ACG.T.TCAACG..A.CTC.GGT.AG..G-G.GC..CC.GCTGAGT.T.AGCATATC..TAGGG.GG
W2_ITS4   302 T.C..GACCC.A.AGTGCCTCTTCT..G.GGA.ACG.T.TCAACG..A.CTC.GGT.AG..-GAGC..CC.G--GATT---------------------
M_macro   631 .TCGGA..CCCCGA.C.CACACGCG....CGGA...CT..G.ACGAGAC.TCAGG----------------------------------------------
B1_ITS1   1067 .TCGGA..CCCCGA.C.CACACGCG....CGGA...CT..G.ACGAGAC.TCAGG.CAGGCGG.GTACCCC.T..T---------------------
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   603 GT.....CC..GC.TC.CCG.A....TCGG.......T...A-----------------------------------------------------------
R1_ITS1   548 GT.....CC..GC.TC.CCG.A....TCGG.......T...A.G.ACTT.CT..TT.T.TGA.C..GCTCTT.C.GTC..TTGA.TCAGG.TTGTCTGC
R1_ITS4   615 GT.....CC..GC.TC.CCG.A....TCGG.......T...A.G-----------A.C..GCTCTT.C.GTC..TTGA.TCAGG.TTGTCTGC
                   1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P_cando   742 ----------------------------------------------------------------------------------------------------
W1_ITS-   378 ----------------------------------------------------------------------------------------------------
W1_ITS4   380 ----------------------------------------------------------------------------------------------------
W2_ITS-   369 .GG.T-----------------------------------------------------------------------------------------------
W2_ITS4   376 ----------------------------------------------------------------------------------------------------
M_macro   684 ----------------------------------------------------------------------------------------------------
B1_ITS1   1141 ----------------------------------------------------------------------------------------------------
B1_ITS4   138 ----------------------------------------------------------------------------------------------------
B_super   645 ----------------------------------------------------------------------------------------------------
R1_ITS1   648 TTCTAAC.GT.TGC.T.TGTA.AC.AT.TTTGAACCTTTTGA.CT.AAATC.GGTA..ACTA..CG.TGAACTTAA.CATATCAATAAGCGGAAGAAGGG
R1_ITS4   698 TTCTAAC.GT.TGC.T.TGTA.AC.AT.TTTGAACCTTTTGA.CT.AAATC.GGTA..ACTA..CG.TGAACTTAA.C-----------------------
              ...
P_cando   742 ---
W1_ITS-   378 ---
W1_ITS4   380 ---
W2_ITS-   373 ---
W2_ITS4   376 ---
M_macro   684 ---
B1_ITS1   1141 ---
B1_ITS4   138 ---
B_super   645 ---
R1_ITS1   748 AGG
R1_ITS4   775 ---
```

Figure S2 continued

Figure S3 Partial sequence alignment of *psb*A-*trn*H intergenic spacer of Kwao Khruea plants. The number above the alignment represents the base numbers of sequence alignments. "." represents nucleotide base that is identical to the nucleotide in the first-row. "-" represents the gap

Figure S4 Partial sequence alignment of *rbc*L of Kwao Khruea plants. The number above the alignment represents the base numbers of sequence alignments. "." represents nucleotide base that is identical to the nucleotide in the first-row. "-" represents the gap

```
                   510       520       530       540       550       560       570       580       590       600
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   501  TCCCCTATTAGGATGTACTATTAAACCTAAATTGGGGTTATCCGCTAAGAATTACGGTAGAGCAGTTTATGAATGTCTTCGTGGGGGACTTGATTTTACC
Butea_mo   501  ....................................................................................................
B_superb   296  ....................................................................................................
Mucuna_m   501  ....................................................................................................
Mucuna_m   501  ....................................................................................................
M_macroc   495  ....................................................................................................
P_mirifi   492  ......................................................................T.............................
P_mirifi   491  ......................................................................T.............................
P_mirifi   492  ......................................................................T.............................
P_montan   471  ......................................................................T.............................
P_montan   471  ......................................................................T.............................
Pv_cando   493  ......................................................................T.............................
Pv_mirif   493  ......................................................................T.............................

                   610       620       630       640       650       660       670       680       690       700
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   601  AAAGATGATGAAAATGTGAATTCCCAACCATTTATGCGTTGGAGAGACCGTTTCGTATTTTGTGCCGAAGCGATTTATAAAGCACAGGCTGAAACAGGTG
Butea_mo   601  ....................................................................................................
B_superb   396  ....................................................................................................
Mucuna_m   601  .............................T...................T...................AC..............................
Mucuna_m   601  .............................T...................T...................AC..............................
M_macroc   595  .............................T...................T...................AC..............................
P_mirifi   592  .................................................T...................C...............................
P_mirifi   591  .................................................T...................C...............................
P_mirifi   592  .................................................T...................C...............................
P_montan   571  .................................................T...........................T...T...................
P_montan   571  .................................................T...........................T...T...................
Pv_cando   593  .................................................T...................C...............................
Pv_mirif   593  .................................................T...................C...............................

                   710       720       730       740       750       760       770       780       790       800
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   701  AAATCAAAGGACATTACTTGAATGCAACTGCGGGTACATGTGAAGAGATGATGAAAAGAGCTGTATTTGCCAGAGAATTGGGCGTTCCTATCGTAATGCA
Butea_mo   701  ....................................................................................................
B_superb   496  ....................................................................................................
Mucuna_m   701  ..........G............A.......C.........A.......C.....................A.............................
Mucuna_m   701  ..........G............A.......C.........A.......C.....................A.............................
M_macroc   695  ..........G............A.......C.........A.......C.....................A.............................
P_mirifi   680  ----------------------------------------------------------------------------------------------------
P_mirifi   691  ..........G.........T...A...........................................A................................
P_mirifi   692  ..........G.........T...A...........................................A................................
P_montan   671  ..........G.....................C...A...............................A................................
P_montan   671  ..........G.....................C...A...............................A................................
Pv_cando   693  ..........G.........T...A...........................................A................................
Pv_mirif   693  ..........G.........T...A...........................................A................................

                   810       820       830       840       850       860       870       880       890       900
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   801  TGATTATTTAACAGGGGGGTTCACTGCAAATACTAGCTTAGCTCATTATTGCCGGGATAATGGTCTACTTCTTCATATACATCGTGCAATGCATGCAGTT
Butea_mo   801  ....................................................................................................
B_superb   596  ....................................................................................................
Mucuna_m   801  ..........A...................G...............T.....................................................
Mucuna_m   801  ..........A...................G...............T.....................................................
M_macroc   788
P_mirifi   680  ----------------------------------------------------------------------------------------------------
P_mirifi   723  ----------------------------------------------------------------------------------------------------
P_mirifi   724  ----------------------------------------------------------------------------------------------------
P_montan   771  ............G.....A...............G...........A....................................................
P_montan   771  ............G.....A...............G...........A....................................................
Pv_cando   793  ....................................................................................................
Pv_mirif   793  ....................................................................................................

                   910       920       930       940       950       960       970       980       990      1000
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   901  ATCGACAGACAAAAGAATCATGGTATGCACTTTCGTGTACTAGCTAAAGCGTTACGTTTATCTGGTGGAGATCATGTTCACGCCGGTACCGTAGTAGGTA
Butea_mo   901  ....................................................................................................
B_superb   658  ----------------------------------------------------------------------------------------------------
Mucuna_m   901  ......................................................................A......T.......................
Mucuna_m   901  ......................................................................A......T.......................
M_macroc   788
P_mirifi   680  ----------------------------------------------------------------------------------------------------
P_mirifi   723  ----------------------------------------------------------------------------------------------------
P_mirifi   724  ----------------------------------------------------------------------------------------------------
P_montan   871  ..........................................A.......G.................................................
P_montan   871  ..........................................A.......G.................................................
Pv_cando   831  ----------------------------------------------------------------------------------------------------
Pv_mirif   808  ----------------------------------------------------------------------------------------------------

                  1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo  1001  AACTTGAAGGGGAAAGAGAAATCACTTTAGGTTTTGTTGATTTACTACGTGATGATTTTATTGAAAAAGATCGAAGTCGCGGTATTTATTTCACTCAGGA
Butea_mo  1001  ....................................................................................................
B_superb   658  ----------------------------------------------------------------------------------------------------
Mucuna_m  1001  ......................................................G............A.................................
Mucuna_m  1001  ......................................................G............A.................................
M_macroc   788
P_mirifi   680  ----------------------------------------------------------------------------------------------------
P_mirifi   723  ----------------------------------------------------------------------------------------------------
P_mirifi   724  ----------------------------------------------------------------------------------------------------
P_montan   971  ...........................................................................T........................
P_montan   971  ...........................................................................T........................
Pv_cando   831  ----------------------------------------------------------------------------------------------------
Pv_mirif   808  ----------------------------------------------------------------------------------------------------
```
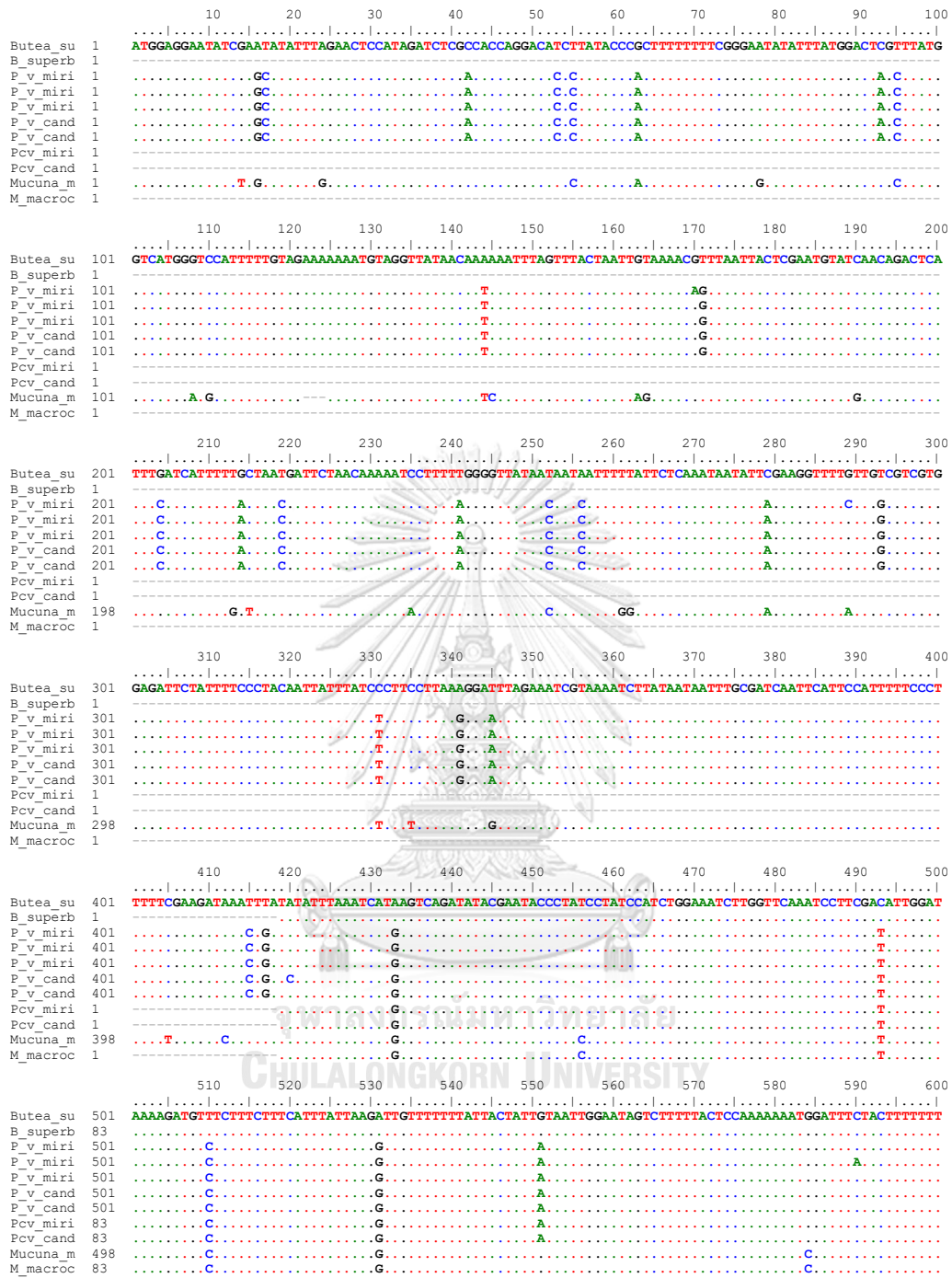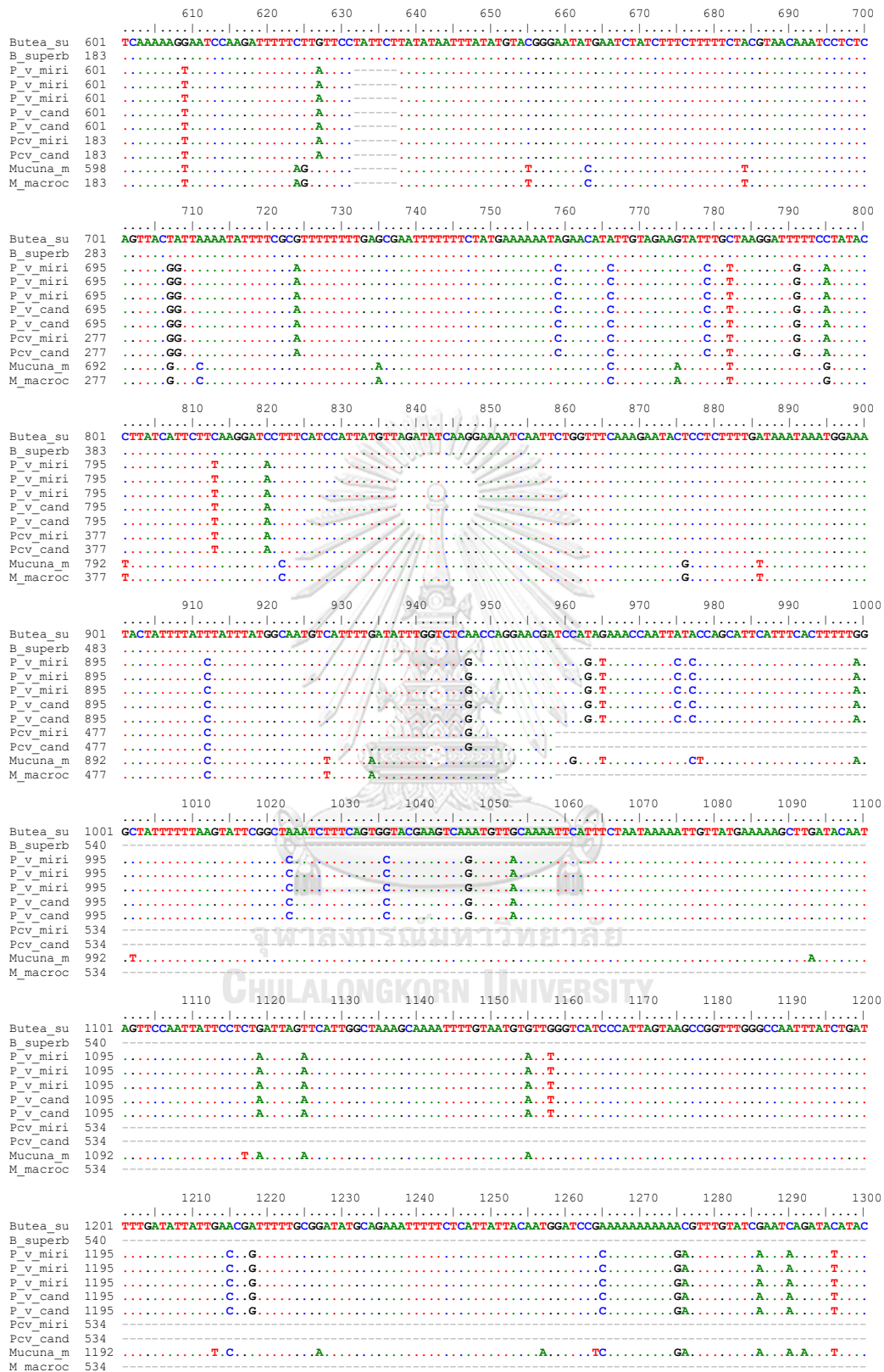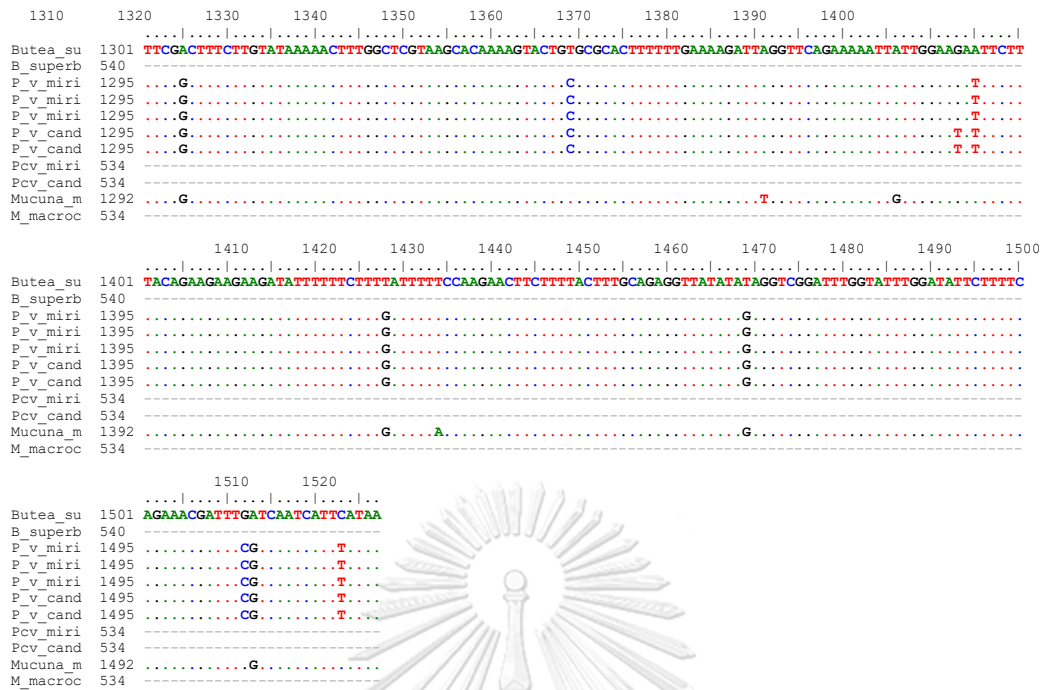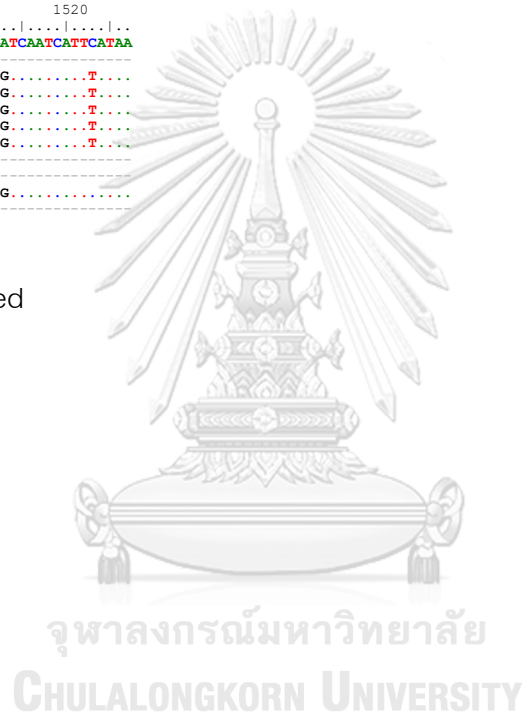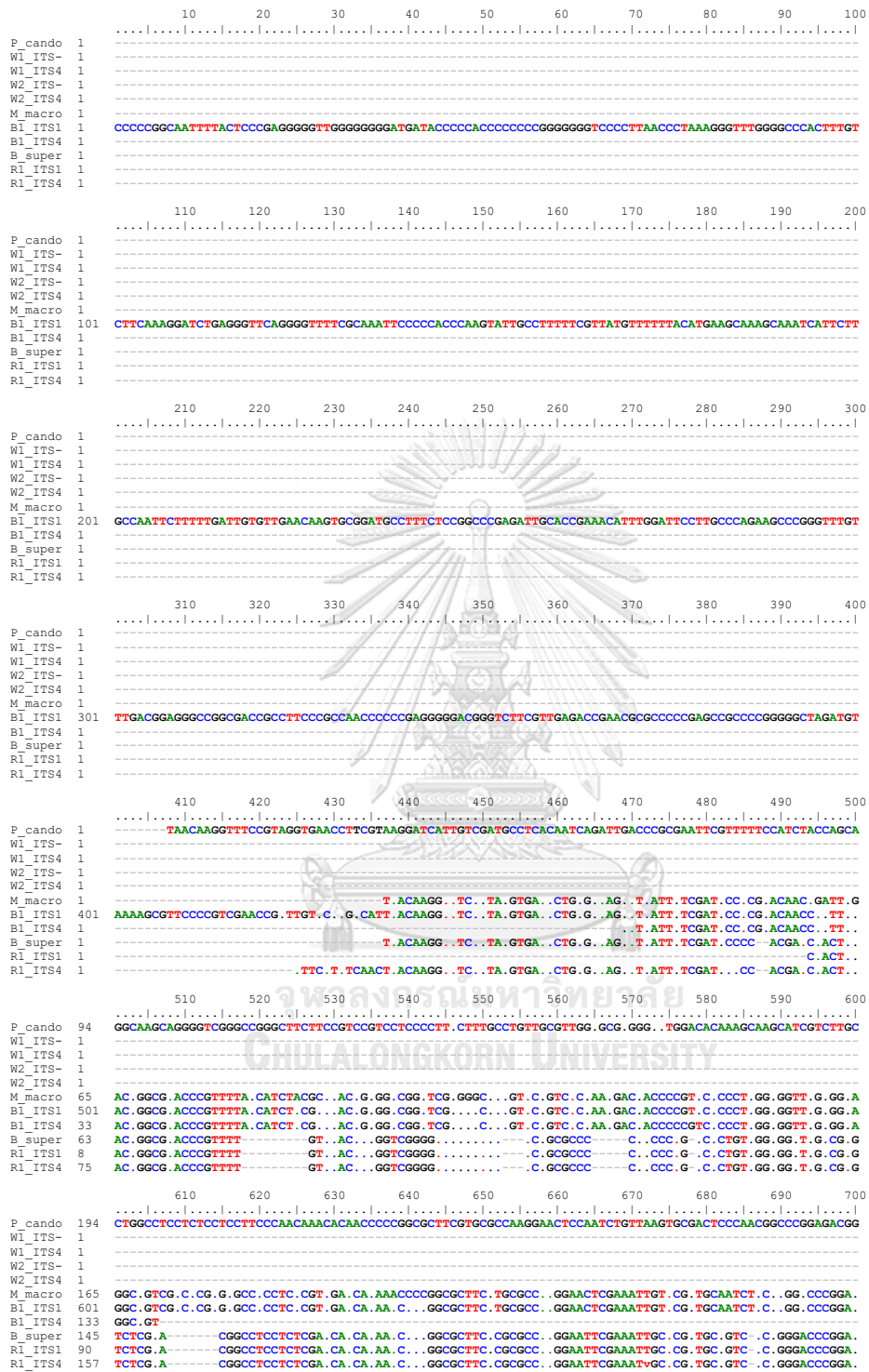
Figure S4 continued

```
                   1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   1101 TTGGGTTTCTCTACCAGGTGTTCTACCTGTTGCTTCGGGAGGTATTCACGTTTGGCATATGCCTGCTCTGACCGAGATTTTTGGGGATGATTCCGTACTC
Butea_mo   1101 ....................................................................................................
B_superb    658 ----------------------------------------------------------------------------------------------------
Mucuna_m   1101 .........C.....T........G..................................................C.....A...................
Mucuna_m   1101 .........C.....T........G..................................................C.....A...................
M_macroc    788 ----------------------------------------------------------------------------------------------------
P_mirifi    680 ----------------------------------------------------------------------------------------------------
P_mirifi    723 ----------------------------------------------------------------------------------------------------
P_mirifi    724 ----------------------------------------------------------------------------------------------------
P_montan   1071 .........................G................................................C.....A...................
P_montan   1071 .........................G................................................C.....A...................
Pv_cando    831 ----------------------------------------------------------------------------------------------------
Pv_mirif    808 ----------------------------------------------------------------------------------------------------

                   1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   1201 CAATTTGGCGGAGGAACTTTAGGACACCCTTGGGGAAATGCACCTGGTGCTGTAGCTAATCGAGTAGCTCTTGAAGCATGTGTACAAGCTCGAAATGAAG
Butea_mo   1201 ....................................................................................................
B_superb    658 ----------------------------------------------------------------------------------------------------
Mucuna_m   1201 .................................................C..................................................
Mucuna_m   1201 .................................................C..................................................
M_macroc    788 ----------------------------------------------------------------------------------------------------
P_mirifi    680 ----------------------------------------------------------------------------------------------------
P_mirifi    723 ----------------------------------------------------------------------------------------------------
P_mirifi    724 ----------------------------------------------------------------------------------------------------
P_montan   1171 .............................................A.............................G..G.....................
P_montan   1171 .............................................A.............................G..G.....................
Pv_cando    831 ----------------------------------------------------------------------------------------------------
Pv_mirif    808 ----------------------------------------------------------------------------------------------------

                   1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Butea_mo   1301 GACGTGATCTTGCTCGTGAAGGTAATGAAATTATCCGTGAGGCTAGCAAATGGAGTCCTGAATTAGCTGCTGCTTGTGAAGTATGGAAGGAGATCAAATT
Butea_mo   1301 ....................................................................................................
B_superb    658 ----------------------------------------------------------------------------------------------------
Mucuna_m   1301 ...............................................T....................................................
Mucuna_m   1301 ...............................................T....................................................
M_macroc    788 ----------------------------------------------------------------------------------------------------
P_mirifi    680 ----------------------------------------------------------------------------------------------------
P_mirifi    723 ----------------------------------------------------------------------------------------------------
P_mirifi    724 ----------------------------------------------------------------------------------------------------
P_montan   1271 ....C...............................................................................................
P_montan   1271 ....C...............................................................................................
Pv_cando    831 ----------------------------------------------------------------------------------------------------
Pv_mirif    808 ----------------------------------------------------------------------------------------------------

                   1410      1420
                ....|....|....|....|....|...
Butea_mo   1401 TGAATTCGATGCAGTGGATACTTTGTAA
Butea_mo   1401 ...........................
B_superb    658 ---------------------------
Mucuna_m   1401 C......CCA...A.............
Mucuna_m   1401 C......CCA...A.............
M_macroc    788 ---------------------------
P_mirifi    680 ---------------------------
P_mirifi    723 ---------------------------
P_mirifi    724 ---------------------------
P_montan   1371 ......---------------------
P_montan   1371 ......---------------------
Pv_cando    831 ---------------------------
Pv_mirif    808 ---------------------------
```
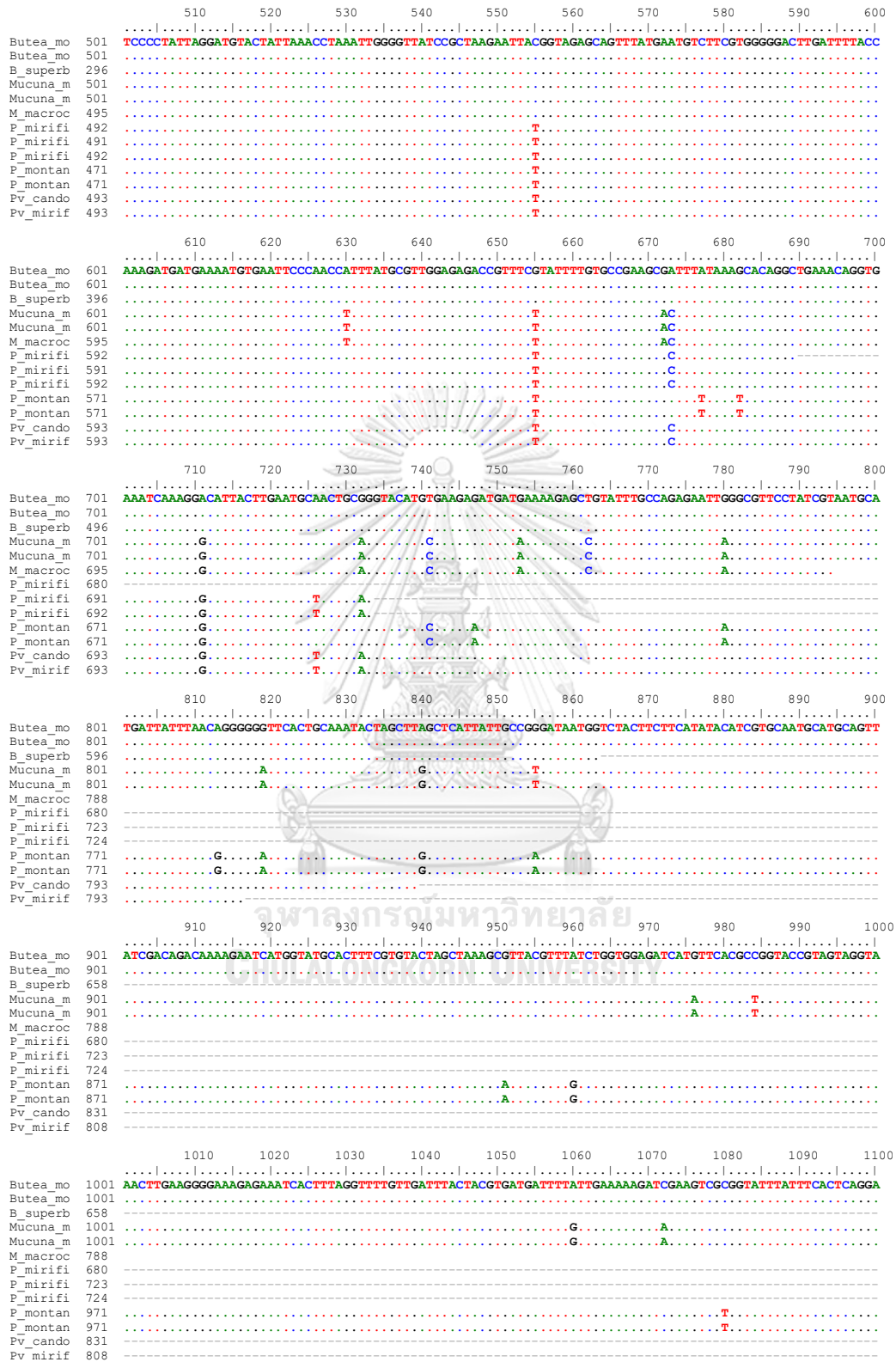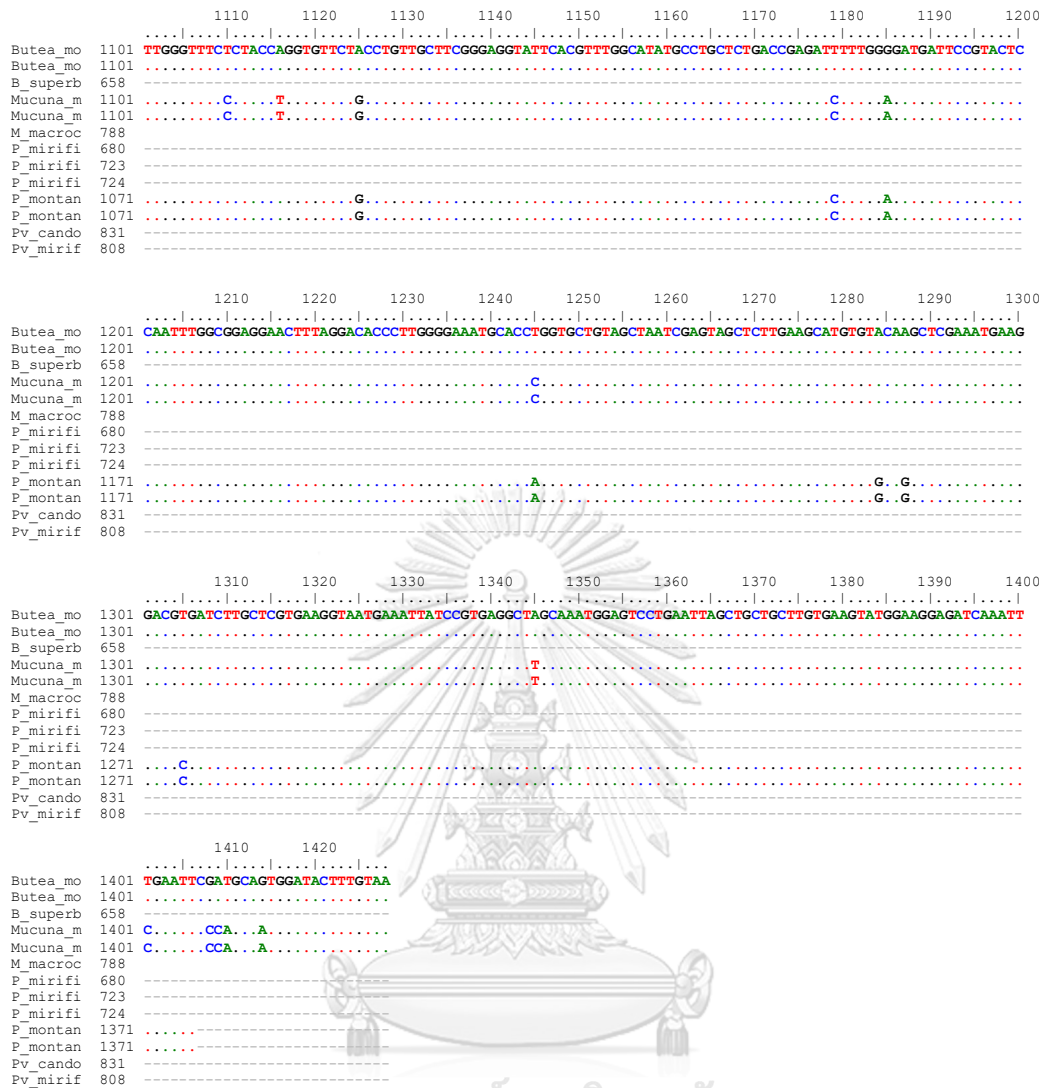
Figure S4 continued

REFERENCES

Anwar, P., Bendini, A., Gulfraz, M., Qureshi, R., Valli, E., Di Lecce, G., . . . Toschi, T. G. (2013). Characterization of olive oils obtained from wild olive trees (Olea ferruginea Royle) in Pakistan. *Food research international, 54*(2), 1965-1971.

Ari, S., & Arıkan, M. (2016). Next-Generation Sequencing: Advantages, Disadvantages, and Future. In (pp. 109-135).

Badea, G., Lăcătuşu, I., Badea, N., Ott, C., & Meghea, A. (2015). Use of various vegetable oils in designing photoprotective nanostructured formulations for UV protection and antioxidant activity. *Industrial Crops and Products, 67*, 18-24.

Butler, J. M. (2017). Recent activities in the United States involving the National Commission on Forensic Science and the Organization of Scientific Area Committees for Forensic Science. *Australian Journal of Forensic Sciences, 49*(5), 526-540. doi:10.1080/00450618.2016.1243153

Chandorn, T. (2016). กวาวเครือขาว. สำนักงานข้อมูลสมุนไพร คณะเภสัชศาสตร์.

Chansakaow, S., Ishikawa, T., Seki, H., Sekine, K., Okada, M., & Chaichantipyuth, C. (2000a). Identification of Deoxymiroestrol as the Actual Rejuvenating Principle of "Kwao Keur", Pueraria m irifica. The Known Miroestrol May Be an Artifact. *Journal of Natural Products, 63*(2), 173-175.

Chansakaow, S., Ishikawa, T., Seki, H., Sekine, K., Okada, M., & Chaichantipyuth, C. (2000b). Identification of Deoxymiroestrol as the Actual Rejuvenating Principle of "Kwao Keur", Pueraria mirifica. The Known Miroestrol May Be an Artifact. *Journal of Natural Products, 63*(2), 173-175. doi:10.1021/np990547v

Chen, Y. W., Song, S., Weng, R., Verma, P., Kugler, J. M., Buescher, M., . . . Cohen, S. M. (2014). Systematic Study of Drosophila MicroRNA Functions Using a Collection of Targeted Knockout Mutations. *Developmental Cell, 31*(6), 784-800. doi:10.1016/j.devcel.2014.11.029

Coward, L., Barnes, N. C., Setchell, K. D., & Barnes, S. (1993). Genistein, daidzein, and their. beta.-glycoside conjugates: antitumor isoflavones in soybean foods from American and Asian diets. *Journal of Agricultural and Food Chemistry, 41*(11),

1961-1967.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., . . . Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, 106*(31), 12794-12797. doi:10.1073/pnas.0905845106

Ingham, J. L., Tahara, S., & Dziedzic, S. Z. (1988). Coumestans from the roots of Pueraria mirifica. *Zeitschrift für Naturforschung C, 43*(1-2), 5-10.

Jing, L., & Zon, L. I. (2011). Zebrafish as a model for normal and malignant hematopoiesis. *Dis Model Mech, 4*(4), 433-438. doi:10.1242/dmm.006791

Kapiotis, S., Hermann, M., Held, I., Seelos, C., Ehringer, H., & Gmeiner, B. (1997). Genistein, the dietary-derived angiogenesis inhibitor, prevents LDL oxidation and protects endothelial cells from damage by atherogenic LDL. *Arteriosclerosis, Thrombosis, and Vascular Biology, 17*(11), 2868-2874.

Karako K, C. Y., Song P, and Tang W. . (2019). Super-aged society: Constructing an integrated information platform of self-recording lifelogs and medical records to support health care in Japan. *BioScience Trends, 13*(3), 278.

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

Kurdyukov, S., Faust, A., Trenkamp, S., Bär, S., Franke, R., Efremova, N., . . . Yephremov, A. (2006). Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain α-, ω-dicarboxylic fatty acids and formation of extracellular matrix. *Planta, 224*(2), 315-329.

Leblanc, V. G., & Marra, M. A. (2015). Next-generation sequencing approaches in cancer: Where have they brought us and wherewill they take us? *Cancers, 7*(3), 1925-1958. doi:10.3390/cancers7030869

Lee, D.-U., Park, H.-W., & Lee, S.-C. (2020). Effect of stigmasterol in liposome bilayer on

the stabilization of encapsulated ascorbic acid. *Korean Journal of Food Science and Technology, 52*(2), 200-203.

Lee, J. H., Kim, J. Y., Cho, S.-H., Jeong, J. H., Cho, S., Park, H. J., & Baek, S. Y. (2017). Determination of miroestrol and isomiroestrol from Pueraria mirifica (white Kwao Krua) in dietary supplements by LC–MS-MS and LC–Q-Orbitrap/MS. *Journal of chromatographic science, 55*(3), 214-221.

Little, D. P. (2014). A DNA mini-barcode for land plants. *Molecular Ecology Resources, 14*(3), 437-446. doi:10.1111/1755-0998.12194

Loera-Sánchez, M., Studer, B., & Kölliker, R. (2019). DNA-Based Assessment of Genetic Diversity in Grassland Plant Species: Challenges, Approaches, and Applications. doi:10.3929/ethz-b-000386451

Matsubara, K., Collins, M. A., & Neafsey, E. J. (1992). Mono-N-Methylation of 1, 2, 3, 4-Tetrahydro-β-Carbolines in Brain Cytosol: Absence of Indole Methylation. *Journal of neurochemistry, 59*(2), 505-510.

Md, F., & Abhijit, C. (2012). Pyrosequencing-An Alternative to Traditional Sanger Sequencing. *American Journal of Biochemistry and Biotechnology, 8*(1). doi:10.3844/ajbbsp.2012.14.20

Osathanunkul M., M. P., Boer H.D. . (2015). Bar-HRM for Authentication of Plant-Based Medicines: Evaluation of Three Medicinal Products Derived from Acanthaceae Species. *Plos one, 10*(5), 11.

Poschner, S., Maier-Salamon, A., Zehl, M., Wackerlig, J., Dobusch, D., Pachmann, B., . . . Jäger, W. (2017). The impacts of genistein and daidzein on estrogen conjugations in human breast cancer cells: a targeted metabolomics approach. *Frontiers in pharmacology, 8*, 699.

Rauf, A., Uddin, G., Arfan, M., & Muhammad, N. (2013). Chemical composition and biological screening of essential oils from Pistacia integerrima. *African Journal of Pharmacy and Pharmacology, 7*(20), 1220-1224.

Roengsumran, S., Petsom, A., Kuptiyanuwat, N., Vilaivan, T., Ngamrojnavanich, N., Chaichantipyuth, C., & Phuthong, S. (2001). Cytotoxic labdane diterpenoids from Croton oblongifolius. *Phytochemistry, 56*(1), 103-107.

Roengsumran, S., Petsom, A., Ngamrojanavanich, N., Ruksilp, T., Sittiwicheanwong, P., Khorphueng, P., . . . Chaichantipayuth, C. (2000). Flavonoid and Flavonoid glycoside from Butea superba Roxb. and their cAMP Phosphodiesterase Inhibitory Activity. *J Sci Res Chula Univ, 25*.

Selepe, M. A., & Van Heerden, F. R. (2013). Application of the Suzuki-Miyaura reaction in the synthesis of flavonoids. *Molecules, 18*(4), 4739-4765.

Shimokawa, S., Kumamoto, T., Ishikawa, T., Takashi, M., Higuchi, Y., Chaichantipyuth, C., & Chansakaow, S. (2013). Quantitative analysis of miroestrol and kwakhurin for standardisation of Thai miracle herb 'Kwao Keur'(Pueraria mirifica) and establishment of simple isolation procedure for highly estrogenic miroestrol and deoxymiroestrol. *Natural Product Research, 27*(4-5), 371-378.

Simko, I. (2016). High-Resolution DNA Melting Analysis in Plant Research. *Trends in Plant Science, 21*(6), 528-537. doi:10.1016/j.tplants.2016.01.004

Singh, B., & Singh, S. (2003). Antimicrobial activity of terpenoids from Trichodesma amplexicaule Roth. *Phytotherapy Research, 17*(7), 814-816.

Stallman, J. K., Funk, V. A., Price, J. P., & Knope, M. L. (2019). DNA barcodes fail to accurately differentiate species in Hawaiian plant lineages. *Botanical Journal of the Linnean Society, 190*(4), 374-388. doi:10.1093/botlinnean/boz024

Tocharus, J., Jamsuwan, S., Tocharus, C., Changtam, C., & Suksamrarn, A. (2012). Curcuminoid analogs inhibit nitric oxide production from LPS-activated microglial cells. *Journal of natural medicines, 66*(2), 400-405.

Villa, C., Costa, J., Gondar, C., Oliveira, M. B. P. P., & Mafra, I. (2018). Effect of food matrix and thermal processing on the performance of a normalised quantitative real-time PCR approach for lupine (Lupinus albus) detection as a potential allergenic food. *Food Chemistry, 262*, 251-259. doi:https://doi.org/10.1016/j.foodchem.2018.04.079

Wallich, N., Curtis, C. M., Drake, S. A., Gorachaud, Vishnupersaud, Treuttel, & Würtz. (1830). *Plantae Asiaticae rariores, or, Descriptions and figures of a select number of unpublished East Indian plants* (Vol. v.1). London :: Treuttel and Würtz.

Wangler, M. F., Yamamoto, S., Chao, H. T., Posey, J. E., Westerfield, M., Postlethwait, J., . . . Undiagnosed Diseases, N. (2017). Model organisms facilitate rare disease diagnosis and therapeutic research. *Genetics, 207*(1), 9-27. doi:10.1534/genetics.117.203067

Wilmot-Dear, C. (1992). A revision of Mucuna (Leguminosae: Phaseoleae) in Thailand, Indochina and the Malay Peninsula. *Kew Bulletin*, 203-245.

Wiriyakarun, S., Yodpetch, W., Komatsu, K., Zhu, S., Ruangrungsi, N., & Sukrong, S. (2013). Discrimination of the Thai rejuvenating herbs Pueraria candollei (White Kwao Khruea), Butea superba (Red Kwao Khruea), and Mucuna collettii (Black Kwao Khruea) using PCR-RFLP. *Journal of natural medicines, 67*(3), 562-570. doi:10.1007/s11418-012-0716-1

Wiriyakarun, S., Zhu, S., Komatsu, K., & Sukrong, S. (2014). The use of cycleave PCR for the differentiation of the rejuvenating herb species Pueraria candollei (White Kwao Khruea), Butea superba (Red Kwao Khruea), and Mucuna macrocarpa (Black Kwao Khruea), and the simultaneous detection of multiple DNA targets in a DNA admixture. *Natural product communications, 9*(1), 111-117. Retrieved from http://ezproxy.car.chula.ac.th/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=24660477&site=eds-live

Yodpetch, W. (2007). *Sequence analysis of matK gene of Pueraria Candollei and Butea Superba and the application of PCR-RFLP genetic marker for identification.* Chulalongkorn University,

Yusakul G, P. W., Udomsin O, Juengwatanatrakul T, and Chaichantipyuth C. (2011). Comparative analysis of the chemical consituents of two varieties of Pueraria candollei. *Fitoterpia., 81*(2), 207.

Zhao, Y., McIntosh, K. B., Rudra, D., Schawalder, S., Shore, D., & Warner, J. R. (2006). Fine-structure analysis of ribosomal protein gene transcription. *Molecular and Cellular Biology, 26*(13), 4853-4862. doi:10.1128/MCB.02367-05

VITA

NAME	Nonthapon Vejpanich
DATE OF BIRTH	26 Aug 1996
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Faculty of Pharmaceutical Sciences, Chulalongkorn University
HOME ADDRESS	1325/28 Ekkachai Road, Bangbon, Bangbon, Bangkok< Thailand 10150
PUBLICATION	-
AWARD RECEIVED	-

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY