**Original Article**

# DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences

CrossMark

*Shih-Chieh Lee*[*], *Chia-Hsiang Wang, Cheng-En Yen, Chieh Chang*

*Department of BioIndustry Technology, Dayeh University, Dacun, Changhua, Taiwan*

A R T I C L E   I N F O

A B S T R A C T

The major aim of made tea identification is to identify the variety and provenance of the tea plant. The present experiment used 113 tea plants [*Camellia sinensis* (L.) O. Kuntze] housed at the Tea Research and Extension Substation, from which 113 internal transcribed spacer 2 (ITS2) fragments, 104 *trnL* intron, and 98 *trnL-trnF* intergenic sequence region DNA sequences were successfully sequenced. The similarity of the ITS2 nucleotide sequences between tea plants housed at the Tea Research and Extension Substation was 0.379–0.994. In this polymerase chain reaction-amplified noncoding region, no varieties possessed identical sequences. Compared with the *trnL* intron and *trnL-trnF* intergenic sequence fragments of chloroplast cpDNA, the proportion of ITS2 nucleotide sequence variation was large and is more suitable for establishing a DNA barcode database to identify tea plant varieties. After establishing the database, 30 imported teas and 35 domestic made teas were used in this model system to explore the feasibility of using ITS2 sequences to identify the varieties and provenances of made teas. A phylogenetic tree was constructed using ITS2 sequences with the unweighted pair group method with arithmetic mean, which indicated that the same variety of tea plant is likely to be successfully categorized into one cluster, but contamination from other tea plants was also detected. This result provides molecular evidence that the similarity between important tea varieties in Taiwan remains high. We suggest a direct, wide collection of made tea and original samples of tea plants to establish an ITS2 sequence molecular barcode identification database to identify the varieties and provenances of tea plants. The DNA barcode comparison method can satisfy the need for a rapid, low-cost, frontline differentiation of the large amount of made teas from Taiwan and abroad, and can provide molecular evidence of their varieties and provenances.

Copyright © 2016, Food and Drug Administration, Taiwan. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author. Department of BioIndustry Technology, Dayeh University, Number 168, University Road, Dacun, Changhua 51591, Taiwan.
E-mail address: slee@cloud.dyu.edu.tw (S.-C. Lee).

## 1.    Introduction

Tea [*Camellia sinensis* (L.) O. Kuntze] is the most popular nonalcoholic health drink in the world. Tea contains many vitamins and minerals, and brewed teas are alkaline drinks that contain many functional ingredients, including fluorine, caffeine, catechins, and γ-amino butyric acid [1—3], which are beneficial to the human body. Tea has many effects, including delaying aging; lowering blood lipids; preventing high blood pressure; lowering blood sugar and preventing diabetes; preventing dental caries; possessing antibacterial and antiviral, anticancer, and anticell mutation, antiradiation, and UV properties; and helping strengthen bones [4,5]. Tea originated from the Yunnan Province in southwest China and has five subspecies and two varieties [6,7]. Currently, only *C. sinensis* (L.) O. Kuntze and its cultivars are widely cultivated and utilized [8].

Tea is an important domestic cash crop in Taiwan. Since joining the World Trade Organization in 2002, Taiwan has opened its market to foreign produce imports. Foreign tea products often vary in quality, and pesticides are applied improperly. However, because of their low prices, unscrupulous traders use low-priced imported teas to mix with or pose as domestic high-priced teas, particularly competition teas, which results in the loss of consumer rights and interests and damages domestic tea reputation [9]. To prevent the use of low-priced imported teas to mix with or pose as domestic high-priced teas, it is necessary to develop a rapid and consistent tea identification technology. Currently, the technologies developed to identify tea variety and provenance can be divided into four categories: mass spectrometry, spectroscopy, separation technology, and sensor technology [10—12]. For example, isotope ratio mass spectrometry can be used to examine the stable isotope ratio in the sample, or a proteomics method can be used to determine the particular proteins during tea growth in Taiwan or the special chemical components in Taiwanese Oolong tea, and then further identify the correct production origin of the analyzed sample [13,14]. However, these physical and chemical examinations focusing on the property of teas require not only expensive instruments, but also experienced operators.

Three *Camellia* species, namely, *C. sinensis* L. (small leaves), *Camellia assamica* Masters (big leaves; also referred to as *C. sinensis* var. assamica), and *Camellia assamica* ssp. lasiocalyx (intermediate leaves), referred to as China, Assam, and Cambod varieties, respectively, are the sources of almost all the teas produced in the world [15]. Most of the commercially grown tea plant varieties in Taiwan include the small-leaf variety, which was previously imported from mainland China and is used for making green tea and some fermented teas, and the large-leaf variety, which was imported from India and is employed for making black tea [16,17]. There are also new varieties (large-, medium-, and small-leaf varieties) obtained through crossbreeding and seed selection using imported varieties at the Taiwan Tea Research and Extension Substation (TRES) and the native large-leaf wild tea from Taiwan (*C. sinensis*). Moreover, tea plants are perennial cross-pollinated plants. All these phenomena make the identification harder. Currently, no one can identify how many tea products are sold commercially and how many plant species can be used for brewing.

The variety of tea can be identified by morphological characteristics, but identification is susceptible to environmental effects and plant age, and it is not easy to determine the genetic relationships between varieties. However, molecular traits can directly demonstrate genetic differences without being susceptible to environmental influences. Many methods have been established to distinguish different varieties. For example, random amplified polymorphic DNA and inter-simple sequence repeat have been used to study the tea plant varieties of Taiwan and evaluate the genetic divergence of the original species [18—20]. In addition, amplified fragment length polymorphism was used in a genetic diversity analysis of 45 Oolong tea varieties from Wu Yi Shan City, Anxi County, Chaoan County of Guangdong, and Taiwan [21]. Although all these methods are robust, development of new identification methods is urgently required.

In 2003, Canadian scholar Hebert and coworkers [22] initially proposed using DNA barcode technology in species identification; the most commonly used DNA barcode molecular marker in animals is cytochrome c oxidase subunit 1 mitochondrial region. Researchers have also proposed many sequences for application in plants. Ribosomal internal transcribed spacer (ITS) combined with a plastid DNA marker was able to discriminate 69.9—79.1% of species of seed plants [23]; also a combination of ITS2 + psbA—trnH barcodes was used to identify herbal materials [24]. However, whether these sequences are suitable for inner-species identification, as in cases of tea products, needs to be further investigated. Nevertheless, with the development of modern molecular biology, DNA markers can be applied to detect and identify the materials used in tea bags [25]. Stoeckle et al [26] used the *rbcL* and *matK* genes as a DNA barcode technology to study 146 merchandised samples of bagged tea leaves and observed that using the DNA barcode the plant materials contained in the tea bags could be identified effectively. To ensure the safety and effectiveness of the product, Li et al [27] systematically collected 177 representative herbal tea samples and performed a DNA barcode study focusing on 44 species of the plants contained in the raw materials. The authors observed four DNA barcode genes that produced 608 sequences, including 173 of the *rbcL* gene, 150 of the *matK* gene, 160 of *psbA*—*trnH* intergenic spacer, and 125 ITS sequences. This study confirmed not only that these sequences can be successfully used to authenticate the raw materials used in herbal teas and identify samples of dry merchandise, but also the value of a DNA barcode in the standardization and quality control of merchandise.

Polymerase chain reaction (PCR) combined with the knowledge of single-nucleotide polymorphism (SNP) sites has been developed to verify the correctness of medicinal plant varieties [28,29]. Some researchers have reported that using SNP molecular markers combined with high-throughput genotyping can establish the fingerprints of specific gene fragments, which can be used rapidly and effectively in the quality control required by the tea industry for identification [30]. Taking advantage of these methods, the present study used 113 original tea plant varieties housed by the TRES, and collected domestic and imported

made tea samples, which are produced as four major commercially available varieties. We investigated the feasibility of using ribosomal ITS2, the intron of *trnL*, and the intergenic sequences (IGS) of the noncoding region of chloroplast *trnL-trnF* and their SNPs as the core molecular markers of the DNA barcode. The barcode was also used to identify the tea plant varieties cultivated in Taiwan and to identify Taiwanese and imported tea varieties and their geographical origins. We wish that our results can be applied to identify the sources and differentiate between varieties of tea. This will not only guarantee the rights of the domestic farmers, but also provide the assurance for buyers.

**Table 1 – List of the 113 varieties of tea germplasms used for phylogenetic tree construction in the present study.[a,b,c,d,e,f,g]**

| Code | Variety or Accession Name | Code | Variety or Accession Name | Code | Variety or Accession Name | Code | Variety or Accession Name |
|---|---|---|---|---|---|---|---|
| 1 | TTES No.1 (I)(2,3,4)(M) | 31 | Mei Jan (II)(PF)(S) | 66 | Buh Jy Chuen (II)(PF)(S) | 98 | Mei Yuan (I)(L)g |
| 2 | TTES No.2 (I)(2,3,4)(M) | 33 | Gau Lu (II)(PF)(S)f | 68 | Jin Guei (II)(PF)(S) | 99 | Min Ghai (I)(L)g |
| 3 | TTES No.3 (I)(3,4)(M) | 34 | Shiang Eel (II)(PF)(S) | 69 | Taur Ren Wu (II)(PF)(S) | 100 | Shuei Jing (I)(L)g |
| 4 | TTES No.4 (I)(3,4)(M) | 35 | Herng Jeh Dah Yeh (II)(PF)(S) | 70 | Taur Ren Joong (II)(PF)(S) | 101 | Le Ye (I)(L)g |
| 5 | TTES No.5 (I)(1,2,5)(M) | 36 | Yan Chuan (II)(PF)(S) | 71 | Niou Shyy Wu (II)(PF)(S) | 102 | Yung Kang (I)(L)g |
| 6 | TTES No.6 (I)(1,2,3)(M) | 37 | Shiaau Tsu Keng (II)(PF)(S) | 72 | Hunan (II)(PF)(S) | 104 | Fu Yun No. 7 (I)(L) |
| 7 | TTES No.7 (I)(3)(L) | 38 | Tian Gong (II)(PF)(S) | 73 | Hannkoou (II)(PF)(S)f | 106 | Tzao Joong (II)(PF)(S) |
| 8 | TTES No.8 (I)(3)(L)f | 39 | Bor Yeh (II)(PF)(S) | 74 | Horng Woei Tzy (II)(PF)(S) | 107 | Chin Shin Oolong (II)(PF)(S)f |
| 9 | TTES No.9 (I)(1,3)(L) | 40 | Guey Hua (II)(PF)(S) | 75 | Mhau Eei (II)(PF)(S) | 108 | Hwang Shin Oolong (II)(PF)(S) |
| 10 | TTES No.10 (I)(1,3)(L) | 41 | Gan Tzy (II)(NF)(1)(S)f | 76 | Hwang Gan (II)(PF)(S) | 109 | Bair Shin Oolong (II)(PF)(S) |
| 11 | TTES No.11 (I)(1,3)(L) | 42 | Gan Tzy (Hwang) (II)(NF)(1)(S) | 78 | Shang Yuan (II)(PF)(S)f | 110 | Horng Shin Oolong (II)(PF)(S) |
| 12 | TTES No.12 (Chinshuan) (I)(2,5)(S)f | 43 | Hwang Jy (II)(NF)(1)(S) | 79 | Rui Sui Da Yeh Wu (II)(PF)(S) | 111 | Dah Yeh Oolong(1) (II)(PF)(S) |
| 13 | TTES No.13 (Tzuiyu) (I)(2,5)(S)f | 44 | San Cha Jy Lan (II)(NF)(1)(S) | 80 | Wu Guu (II)(PF)(S) | 112 | Chin Shin Dah Pan (II)(PF)(S)f |
| 14 | TTES No.14 (I)(2,5)(S)f | 45 | Wen Sun JY Lan (II)(NF)(1)(S) | 81 | Niou Buu (II)(PF)(S) | 113 | Horng Shin Dah Pan (II)(PF)(S) |
| 15 | TTES No.15 (I)(2,6)(S) | 46 | Jy Lan (II)(NF)(1)(S) | 82 | Darjeeling (III)(F)(3)(L) | 114 | Dah Pan (II)(PF)(S) |
| 16 | TTES No.16 (I)(5,7)(S) | 47 | Dann Shoei Chin Shin (II)(NF)(1)(S) | 83 | Burma (III)(F)(3)(L)f | 115 | Mau Tzy (II)(PF)(S) |
| 17 | TTES No.17 (I)(2,8)(S)f | 48 | Dah Hwu Woei (II)(NF)(1)(S) | 84 | Sun Cha (I)(L)f,g | 116 | Wu Jin (II)(PF)(S) |
| 18 | TTES No.18 (I)(3)(L)f | 49 | Wen Sun Dah Yeh Wu (II)(NF)(1)(S) | 85 | Chyh Ya Sun Cha (I)(L)g | 117 | Bair Mau Hour (II)(PF)(S)f |
| 19 | TTES No.19 (I)(2,5)(S)f | 50 | Lin Koou Dah Yeh Wu (II)(NF)(1)(S) | 86 | Shan (III)(F)(3)(L)f | 118 | Dah Nan Bair Mau Hour (II)(PF)(S) |
| 20 | TTES No.20 (I)(2,5)(S)f | 51 | Bair Yeh (II)(NF)(1)(S) | 87 | Manipuri (III)(F)(3)(L) | 119 | Tuh Tzy Keng Bair Mau Hour (II)(PF)(S) |
| 21 | Harn Shiaw (II)(PF)(S) | 52 | Shyr Cha (II)(NF)(1)(S) | 88 | Kyang (III)(F)(3)(L) | 121 | Fwu Jou (II)(PF)(S) |
| 22 | Ying Jy Horng Shin (II)(PF)(S)f | 55 | Yabukita (II)(NF)(1)(S)f | 89 | Jaipuri (III)(F)(3)(L) | 122 | Bian Joong Oolong (II)(PF)(S) |
| 23 | Ying Jy Tzao Joong (II)(PF)(S) | 56 | Jwu Yeh (II)(NF)(1)(S) | 90 | Assam indigenous (III)(F)(3)(L)f | 123 | Dah Yeh Oolong(2) (II)(PF)(S) |
| 24 | Tiee Guan In (II)(PF)(S)f | 59 | Hei Miann Tzao Joong (II)(PF)(S) | 91 | Keemon (III)(F)(3)(L) | | |
| 25 | Sheau Yeh Tiee Guan In (II)(PF)(S)f | 60 | Chin Shin Tzao Joong (II)(PF)(S) | 92 | Fkk (III)(F)(3)(L) | | |
| 26 | Bair Shin Wuu Yi (II)(PF)(S) | 61 | Keelung Bair Joong (II)(PF)(S) | 93 | Lai Tou (I)(L)g | | |
| 27 | Horng Shin Wuu Yi (II)(PF)(S) | 62 | Tsyh Joong (II)(PF)(S) | 94 | Long Tou (I)(L)g | | |
| 28 | Wuu Yi (II)(PF)(S)f | 63 | Shen Mann Joong (II)(PF)(S) | 95 | Nan Fong (I)(L)g | | |
| 29 | Dah Terng (II)(PF)(S) | 64 | Woan Joong (II)(PF)(S) | 96 | Fong Huang (I)(L)g | | |
| 30 | Feng Tzy Lin (II)(PF)(S) | 65 | Shy Jih Chuen (II)(PF)(S)f | 97 | De Hua She (I)(L)g | | |

[a] Species: I, *Camellia sinensis*; II, *Camellia sinensis* var. sinensis; III, *Camellia sinensis* var. assamica.

[b] Suitability: NF, nonfermented; PF, partially fermented; F, fermented.

[c] Tea type: 1, green Tea; 2, Oolong Tea; 3, red tea; 4, Mee Tea; 5, Pou Chong Tea; 6, White Tea; 7, Long Jing Tea; 8, Shou Mei Tea.

[d] Leaf type: L, large; M, medium; S, small.

[e] TTES=Tea Research and Extension Substation, Executive Yuan, Taiwan ROC.

[f] Popular or distinguishing tea variety in Taiwan.

[g] Native Taiwanese wild tea.

## Table 2 — A list of the 30 imported and 35 domestic tea samples used in present study.[a–f]

| Code | Geographical Origin | Variety or Accession Name | Code | Geographical Origin | Variety or Accession Name |
|---|---|---|---|---|---|
| Imported from overseas | | | | Domestic in Taiwan | |
| TJK58 | Yen Bai , Yenbai (N. Vietnam) | N/A (Oolong)[(S)] | PJK0075-0077 | Jhushan, Nantou | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK61 | South part of India | Assam indigenous [(III)(F)(3)(L)f] | PJK130-131,143 | Mingjian, Nantou | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK62 | Luxi, Yunnan (China) | N/A (Red Tea) | PJK190-191 | Lugu, Nantou | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK64 | Sri Lanka | Assam indigenous [(III)(F)(3)(L)f] | PJK402-404 | Alishan, Chiayi | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK65 | Moc Chau, Sonla (N. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | PJK448, 450-451 | Taitung | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK66 | Bao Loc, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | PJK452-454 | Taitung | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK67 | Ningde, Fujian (China) | Fu Bei Zhong | PJK550,552 | Pinglin, New Taipei | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK68 | Yaan, Sichuan (China) | Fu Ding Da Bai | PJK570,574 | Pinglin, New Taipei | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK69 | North Vietnam | N/A[(M)] | PJK907-909 | Yilan | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK70 | Simao, Yunnan (China) | N/A[(L)] | PJK946-948 | Hualien | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK71 | Nghe An (M. Vietnam) | N/A (Green Tea) | PJK1018 | Yilan | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK82 | Bao Loc, Lamdong (S. Vietnam) | TTES No.13 (Tzuiyu) [(I)(2,5)(S)f] | PJK1034,1036,1038 | Hualien | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK83 | Bao Loc, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | PJK1069-1071 | Lishan, Taichung | Chin Shin Oolong [(II)(PF)(S)f] |
| TJK84 | Bao Loc, Lamdong (S. Vietnam) | Shy Jih Chuen [(II)(PF)(S)f] | PJK1136 | Alishan, Chiayi | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] |
| TJK90 | Mae Salong (Thailand) | Chin Shin Oolong [(II)(PF)(S)f] | | | |
| TJK91-92 | Kota Medan, Sumatra (Indonesia) | Chin Shin Oolong [(II)(PF)(S)f] | | | |
| TJK93 | Da Lat, Lamdong (S. Vietnam) | Chin Shin Oolong [(II)(PF)(S)f] | | | |
| TJK94-95 | Bao Loc, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | | | |
| TJK96 | Da Lat, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | | | |
| TJK97-98 | Moc Chau, Sonla (N. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | | | |
| TJK99 | Bao Loc, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)f] | | | |
| TJK103 | Sumatra (Indonesia) | Chin Shin Oolong [(II)(PF)(S)f] | | | |
| TJK104 | Bao Loc, Lamdong (S. Vietnam) | TTES No.12 (Chinhsuan) [(I)(2,5)f] | | | |
| TJK105 | Bao Loc, Lamdong (S. Vietnam) | Shy Jih Chuen [(II)(PF)(S)f] | | | |
| TJK107 | Bao Loc, Lamdong (S. Vietnam) | TTES No.13 (Tzuiyu) [(I)(2,5)(S)f] | | | |
| TJK110 | South Vietnam | TTES No.12 (Chinhsuan) [(I)(2,5)(S)] | | | |
| TJK113 | Sumatra (Indonesia) | TTES No.12 (Chinhsuan) [(I)(2,5)(S)] | | | |

[a] Species: I, *Camellia sinensis*; II, *Camellia sinensis* var. sinensis; III, *Camellia sinensis* var. assamica.
[b] Suitability: NF, nonfermented; PF, partially fermented; F, fermented.
[c] Tea type: 1, green Tea; 2, Oolong Tea; 3, red tea; 4, Mee Tea; 5, Pou Chong Tea; 6, White Tea; 7, Long Jing Tea; 8, Shou Mei Tea.
[d] Leaf type: L, large; M, medium; S, small.
[e] TTES = Tea Research and Extension Substation, Executive Yuan, Taiwan ROC.
[f] Popular or distinguishing tea variety in Taiwan.

## 2. Methods

### 2.1. Sample sources and their treatment

Different varieties of tea plant [*C. sinensis* (L.) O. Kuntze] samples housed at the Taitung Substation of Taiwan TRES were provided by Mr Sang-Shun Wu, who also assisted in identifying the tea plants. An apical bud and the two youngest leaves of a single strain of each variety were collected, and then the leaves were fixed, dried, and processed without fermentation [31] into green tea samples for further analysis (sample preparation time was March 2008; Table 1). The process of making tea does not seem to affect DNA extraction, as evident from the work of Stoeckle et al [26], who collected 146 products from 33 manufacturers of 17 countries, and did not have any problem in DNA extraction and successful PCR amplification of > 90% (131/146) of samples. In addition, 35 Taiwanese domestic made tea (sample number starting with PJK) and 30 imported made tea samples (sample number starting with TJK) for the experiments were prepared and provided by Dr Chia-Chang Wu at the Tungding Substation of Taiwan TRES (sample preparation time was November 2011; Table 2). For each tea sample, 20 g was used for further experimental study, and another 10 g was placed in an aluminum bag, which was vacuumed and sealed before being stored in the specimen room of the Department of BioIndustry Technology of Dayeh University, Changhua, Taiwan.

### 2.2. DNA extraction

Genomic DNA extraction [32] was based on the cetyl trimethyl ammonium bromide (CTAB) method with slight modifications. Briefly, 50 mg of the dry tea sample was transferred into a 2 mL Eppendorf tube, and liquid nitrogen was added to the tube. A pestle was used to grind the tea sample into powder. Then, 0.5 mL of prewarmed extraction buffer (100mM TrisHCl, pH 8.0; 20mM EDTA; 1M NaCl; 1% CTAB; and 1% PVP-40) was added to the powder. The tube was left in a 65°C thermostatic water bath for 20 minutes with gentle shaking. An equal volume of chloroform:isoamyl alcohol (24:1) was then added to the sample, and the sample was centrifuged at 11,000 $g$ for 20 minutes at 4°C. The supernatant was transferred to a new Eppendorf tube containing 1 mL of precipitation buffer (50mM TrisHCl, pH 8.0; 10mM EDTA; 40mM NaCl; and 1% CTAB), and the tube was left at room temperature for 1 hour. The sample was centrifuged again at 11,000 $g$ for 15 minutes at 4°C. The supernatant was gently decanted, and the pellet was gently resuspended in 175 μL of 1.2M NaCl (containing 10 mg/mL RNase A). The sample was incubated at 37°C for 30 minutes in a dry bath, and then 75 μL of chloroform:isoamyl alcohol (24:1) was added to the sample. The aqueous phase was transferred to a new tube, and 1/10 of the recovered volume of 3M sodium acetate and 2 volumes of 95% ethanol were added to the tube to precipitate DNA. After centrifugation at 11,000 $g$ for 20 minutes at 4°C, a DNA pellet was obtained. The pellet was washed with 0.5 mL of 70% ethanol, and the tube was then left in a fume hood to allow the ethanol to dry completely. The DNA was then dissolved in 25–50 μL of Tris-EDTA buffer solution. The extracted DNA was stored at −20°C for further experimental analysis.

### 2.3. PCR amplification

The primer pair used in the PCR amplification reactions of the ribosomal noncoding region fragment sequence ITS2 of the tea samples was BEL-1/BEL-3 [33]. This is a mixed primer pair, designed as follows: forward primers, BEL-1: 5′-GGDGCGGA-KAHTGGCCYCCCGTGC-3′ (where D represents A, G, or T; K represents G or T; H represents A, C, or T; and Y represents T or C); reverse primers, BEL-3: 5′-GACGCTTCTCCAGACTACAAT-3′. The PCR solution with a volume of 50 μL contained 2 μL of DNA template (40–80 ng), 5 μL of 10× PCR reaction buffer, 3 μL of 25mM MgCl₂, 3 μL of 2.5mM Deoxynucleotide (dNTP), 1μL of 10μM forward primer, 1μL of 10μM reverse primer, 0.5 μL (5 units) of Taq DNA polymerase (Geneaid Biotech Ltd, Taipei, Taiwan), 3 μL of Dimethyl sulfoxide (DMSO), and 31.5 μL of sterile distilled water. The following reaction conditions were used: template DNA-> denatured, 96°C for 12 minutes followed by 36 cycles of reaction (95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1.5 minutes). The last cycle included a 10-minute extension at 72°C.

The following design of two universal primer pairs was used. The c, d pair was used for the PCR amplification of the chloroplast *trnL* intron of the tea samples, and the e, f pair was used for the PCR amplification of *trnL*-*trnF* IGS [34]; forward primers, c: 5′-CGAAATCGGTAGACGCTACG-3; e: 5′-GGTTCAAGTCCCTC-TATCCC-3′; reverse primers, d: 5′-GGGGATAGAGGGGACTT GAAC-3′ and f: 5′-ATTGAACTGGTGACACGAG-3′. The total volume of 25 μL PCR solution included 1.5 μL of DNA template (10 ng/μL), 2.8 μL of 10× PCR reaction buffer, 0.3 μL of 25mM MgCl₂, 2.0 μL of 2.5mM dNTP, 0.5μL of 10μM forward primer, 0.5 μL of 10μM reverse primer, 0.4 μL (5 units) of Taq DNA polymerase, 2.0 μL of DMSO, and 15.0 μL of sterile distilled water. The following reaction conditions were used: template DNA-> denatured, 95°C for 5 minutes followed by 35 cycles of reaction (95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 45 seconds]; the last cycle included a 10-minute extension at 72°C.

### 2.4. DNA sequencing

Five microliters of PCR product, 0.5 μL 100 bp DNA ladder, and 1 μL of Seeing safe (Seeing Bioscience Co., Ltd, Taipei, Taiwan) were loaded onto 1.6% agarose gel for electrophoresis to examine the PCR products. ABI PRISM 377 DNA sequencer (Applied Biosystems Industries, Foster City, CA, USA) was used for sequencing. The sequences obtained were initially compiled using BioEdit version 7.0 (Tom Hall, http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information homepage (http://www.ncbi.nlm.nih.gov/blast/blast.cgi) was then used to compare these sequences with in-house sequences and GenBank database sequences. The correctness of the possible variations of all sequences was confirmed by bidirectional sequencing. The PCR products of each tea sample were sequenced at least three times from 5′ and 3′ to determine the locations of ribosomal ITS2, chloroplast *trnL* intron, and *trnL*-*trnF* IGS fragment sequences.

### 2.5. Data analysis

The sequences of the successfully sequenced fragments were initially aligned and compared using BioEdit and MEGA 6

([http://www.megasoftware.net](http://www.megasoftware.net)) software. BioEdit was used to align the ITS2 sequences of the tea plant varieties in this study and to calculate their sequence similarity. A phylogenetic tree was constructed using the alignment, clustering, and hierarchy of the ITS2 fragment sequences of the tea plants. The clustering methods included unweighted pair group method with arithmetic mean (UPGMA), neighbor joining (NJ) method, minimum evolution (ME) method, maximum likelihood method, and maximum parsimony (MP) method. The bootstrap value was set at 500 repeated sampling to examine the reliability of the phylogenetic tree.

## 3. Results and discussion

### 3.1. Tracing the origins of Taiwanese tea plants and acquisition of domestic and foreign made teas for the experiments

Samples of the 113 tea plant varieties used in the present experiment were provided by the Taitung Substation of TRES, including 20 TRES breeding varieties (*C. sinensis*), 72 small-leaf varieties (*C. sinensis* var. sinensis), nine large-leaf varieties (*C. sinensis* var. assamica), and 12 Taiwanese native wild tea plants (*C. sinensis*). The tea plants listed in Table 1 were divided into five categories according to their suitability: (1) varieties selected and promoted for cultivation by the TRES with their Taiwan tea numbers 1–20 (TTES 1–20; code number in purple); (2) varieties that are suitable for making some fermented teas, with variety numbers 21–31, 33–40, 59–66, 68–76, 78–81, 106–119, and 121–123 (code number in orange); (3) varieties that are suitable for making green teas, with variety numbers 41–52 and 55–56 (code number in green); (4) varieties that are suitable for making black tea, with variety numbers 82–83 and 86–92 (code number in red); and (5) *Camellia* with variety numbers 84–85 and 93–102 (code number in blue; Table 1). Therefore, the database consisting of the core molecular markers of Taiwanese tea plant species can be used to authenticate domestic tea.

To test the constructed DNA barcode library, 30 imported and 35 domestically produced made teas were submitted for analysis. Table 2 lists the made teas used in this study, including the four varieties of Chin Shin Oolong, TTES No. 12 (Chinhsuan), Shy Jih Chuen, and TTES No. 13 (Tzuiyu), which are currently produced and sold in bulk in Taiwan. The imported made teas were purchased from 16 production sites in six countries, including Vietnam, Thailand, Indonesia, India, Sri Lanka, and China. Taiwanese domestic made tea samples were collected from several production sites in Taiwan, including Pinglin, New Taipei City, Lishan, Taichung city, Jhushan, Mingjian and Lugu, Natou County, Alishan, Chiayi County, Taitung County, Yilan County, and Hualien County. These regions were selected due to their significance in the market as major tea production sites, both domestically and internationally. In this report, the major source of domestic cultured tea library was established using tea plants supported by the Taitung Substation of TRES. From the above production sites, 30 more imported made tea and 35 more domestically produced made tea were supplemented to the database.

Most of the tested made teas could find their corresponding origins in the TRES. To our surprise, 19 out of 30 imported teas, especially those imported from Vietnam and Indonesia, could find their similarity to cultivars in Taiwan. Among them, 12 were similar to TRES breeding species, 10 were similar to TTES No. 12, and two were similar to TTES No. 13. In addition, five were Chin Shin Oolong species and two Shy Jih Chuen. These similarities raised our suspicion that specific varieties of tea cross bred by domestic farmers have been brought to foreign soil through unofficial channels, which directly impact the profitability of the farmers. In Taiwan, owing to tariff reduction and demand of soft drinks, more teas are imported from all over the world than produced domestically ([http://teaquality.ttes.gov.tw/loadfile/103-m.pdf](http://teaquality.ttes.gov.tw/loadfile/103-m.pdf)). These 30 samples may not really represent the bulk of imported made teas. However, these results still reflected that many imported teas imitated domestic made teas. Methods to distinguish domestic from imported teas are urgently required. It is a good start that Taiwan Food and Drug Administration ([http://www.fda.gov.tw/TC/newsContent.aspx?id=19946&chk=5b5c1079-e36d-4356-86bc-dbc0fc5cf0f8&param=pn&cid=4&cchk=f11420b2-cf8e-4d3a-beb5-66521b800453&key1=%E8%8C%B6%E8%91%89#.V2qu-bh942w](http://www.fda.gov.tw/TC/newsContent.aspx?id=19946&chk=5b5c1079-e36d-4356-86bc-dbc0fc5cf0f8&param=pn&cid=4&cchk=f11420b2-cf8e-4d3a-beb5-66521b800453&key1=%E8%8C%B6%E8%91%89#.V2qu-bh942w)) announced mandatory inspection of each batch of imported teas from July 31, 2015. That may make the Taiwan Food and Drug Administration as a good source of imported teas. This can serve as a good source for imported tea samples and allow further expansions of our database.

### 3.2. PCR amplification, sequencing, and analysis of the three noncoding regions of ribosomal ITS2, chloroplast trnL intron, and trnL-trnF IGS of tea plants

The primer pair BEL-1/BEL-3 is a mixture of primer pairs reported by Chiou et al [33] and were designed based on plant sequences from 120 species of 52 families. After testing, these primer pairs could successfully amplify the ITS2 of 300 medicinal herbs. In the present study, these primer pairs were successfully applied to the PCR amplification of the ITS2 fragments of made tea samples and the tea plant species
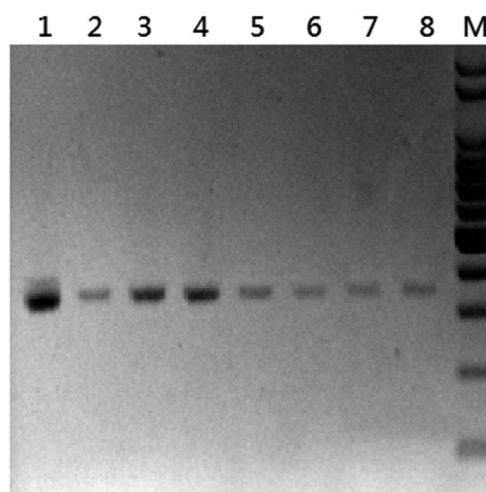


Figure 1 – Electrophoresis of the PCR amplification product of ITS2 fragment. ITS2 = internal transcribed spacer 2; PCR = polymerase chain reaction.
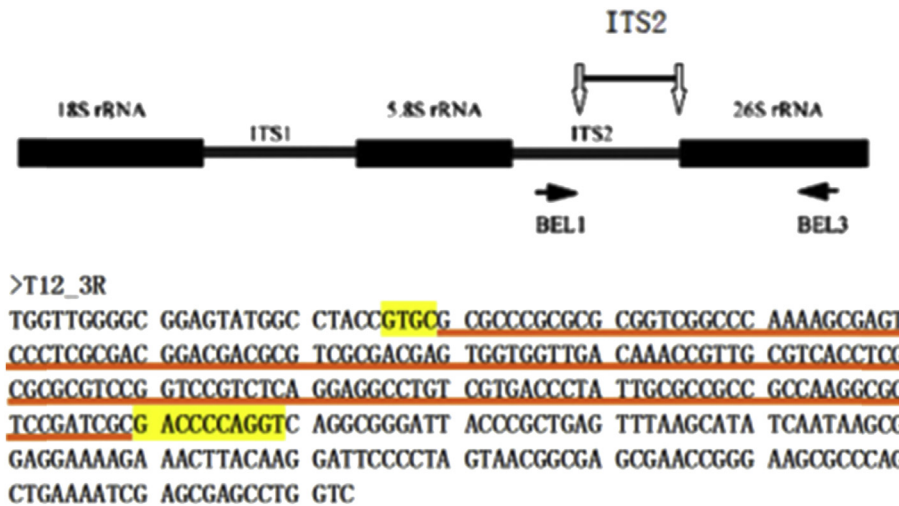
**Figure 2 — Schematic diagram of the position of ITS2 sequence fragment amplified using BEL-1/BEL-3 primer pair. ITS2 = internal transcribed spacer 2.**

housed by the TRES (Figure 1). Using the sequencing result of the TTES No. 12 (Chinhsuan) as an example (Figure 2), the ITS2 DNA fragment obtained through PCR amplification using the BEL-1/BEL-3 primer pairs was between the conserved sequence GTGC (yellow highlight) and the characteristic 5′ end sequence GACCCCAGGT (yellow highlight) of the 26S rRNA gene, which was a component of the ITS2 fragment.

All the sequences obtained by sample amplification and sequencing were first subjected to manual searching to determine the ITS2 fragment sequence before further analysis. The longest ITS2 sequence was 164 bp, from Taiwan tea No. 5, and the shortest was only 149 bp, from No. 88 Kyang and No. 114 Dah Pan. The ITS2 sequences of all tea samples were grouped according to their length. The long group was 160 (±1) bp and contained 101 varieties. The short group was 150 (±1) bp and contained only seven varieties. Only five varieties did not belong to any of these two groups: Nos. 4, 5, 30, 59, and 89. The average ITS2 sequence of these tea plant varieties was 159.3 bp. BioEdit was used to perform multiple sequence alignments of these 113 ITS2 sequences. The calculated similarity of ITS2 sequences of the Taiwanese tea plant varieties was between 0.379 and 0.994, and there were 149 variant sites, accounting for 88.1% of the 169 sites after multiple sequence alignment. The highest similarity occurred between TTES No. 14 and Dah Terng, and the largest difference was between Woan Joong and Hwang Gan. Within this noncoding region, no two varieties shared an identical sequence (Appendix 1).

Currently, 25 important and distinctive varieties have a wide cultivation area, and these varieties include TTES Nos. 8, 12, 13, 14, and 17; Assam indigenous (No. 90), Chin Shin Oolong (No. 107), Chin Shin Dah Pan (No. 112), Shy Jih Chuen (No. 65), Gan Tzy (No. 41), Tiee Guan In (No. 24), Wuu Yi (No. 28), and Ying Jy Horng Shin (No. 22); new cultivars TTES Nos. 18, 19, 20; and important varieties Yabukita (No. 55), Hannkoou (No. 73), Bair Mau Hour (No. 117), Sheau Yeh Tiee Guan In (No. 25), Shang Yuan (No. 78), Gau Lu (No. 33), Burma (No. 83), Shan (No. 86), and Sun Cha (No. 84; Table 1). In these 25 varieties, the length of their ITS2 sequences belongs to the long group with 160 (± 1) bp. Multiple sequence alignment and sequence

similarity analysis of the ITS2 sequences of these 25 important tea plant varieties were conducted, and the results indicated that their similarities ranged from the smallest of 0.529 (Yabukita No. 55 vs. TTES No. 19) to the largest of 0.984 (TTES No. 18 vs. TTES No. 8). Sequence similarity indicated that the ITS2 sequence had adequate polymorphism, and it not only could be used to identify the 25 distinctive tea plant varieties that are widely cultivated, but also had sufficient resolution to identify the 113 tea plant varieties housed at the TRES (Appendix 1). These results agree with those of Han et al [35] and Pang et al [36], who support that ITS2 can distinguish species, while others have different conclusion and prefer ITS1 [37] and ITS [38]. The discrepancy in conclusion may be the results of using primers and extracted DNA quality.

To have better resolution, chloroplast *trnL* intron sequences and *trnL-trnF* IGS were obtained through PCR amplification, sequencing, comparison of sequences with data in the National Center for Biotechnology Information database, and multiple sequence alignments using the BioEdit software. Sequencing determined the *trnL* intron sequences for 104 varieties with a sequence length between 510 bp and 515 bp, but the *trnL-trnF* IGS for 98 tea plant varieties with a sequence length between 387 bp and 394 bp. Owing to the high conservation of the *trnL* sequences, the similarity of the *trnL* intron sequences between each tea plant variety ranged between 0.948 and 1.000. Likewise, the similarity of the *trnL-trnF* IGS ranged between 0.979 and 1.000. These results indicate that the chloroplast *trnL* intron sequence and *trnL-trnF* IGS variation was more conserved compared with the ITS2 fragment, and their resolution was not sufficient for the molecular identification of all tea plant varieties studied here.

### 3.3. Using a phylogenetic tree to identify made tea varieties and provenances

To make the phylogenetic tree in the present experiment, in addition to the UPGMA, four other methods were also used, including the (1) NJ, (2) ME, (3) MP, and (4) maximum likelihood methods. The use of these different calculation methods
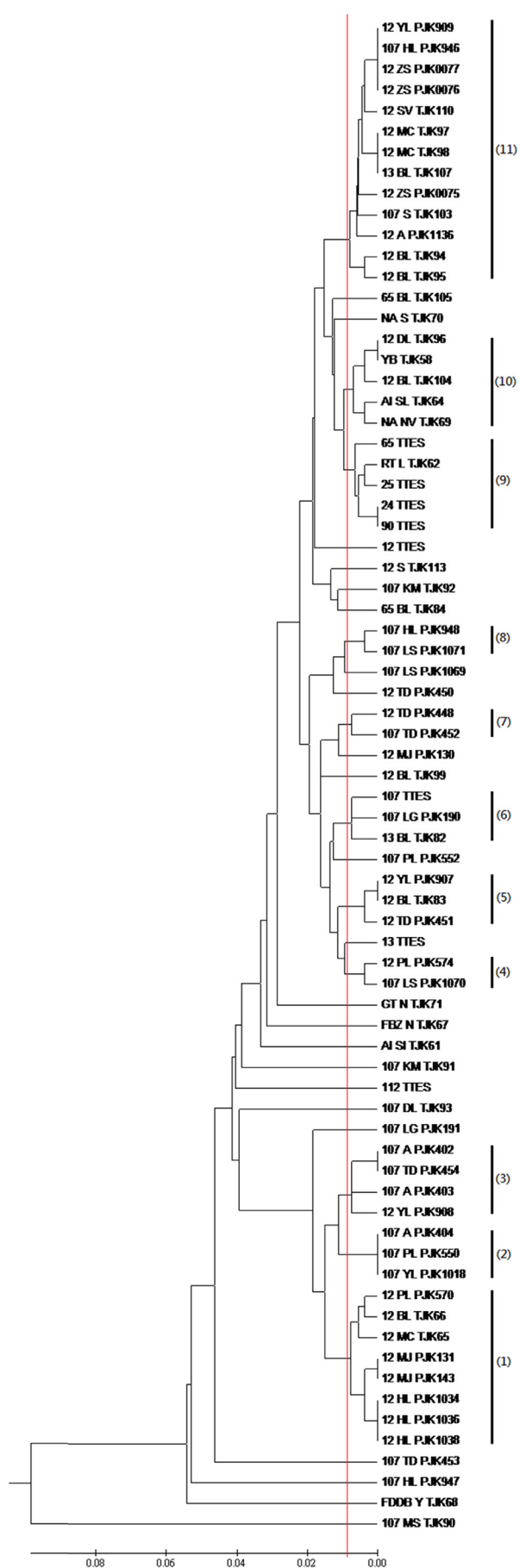
yielded results similar to that of the UPGMA (data not shown). The present study used a phylogenetic tree constructed using the ITS2 sequences of the 113 tea plant varieties to further investigate the affinity of the tea plant varieties (Appendix 2). An analysis using the NJ and ME methods showed that No. 76 Hwang Gan was the most prominent variety in the phylogenetic tree. Although this variety was grouped with No. 71 Niou Shyy Wu, their ITS2 sequence similarity was only 0.531. Agreeing with the methods listed earlier, the UPGMA also showed that Hwang Gan was an independent cluster. No. 85 Chyh Ya Sun Cha, No. 93 Lai Tou, and No. 100 Shuei Jing were always categorized into one cluster using the NJ, ME, and MP methods. These results were comparable with the category that was made by both leaf and bud morphology [39] and inter-simple sequence repeat molecular marker [40]. Another Taiwanese wild *Camellia* cluster can also be observed using the NJ, ME, and MP methods, and this cluster included No. 84 Sun Cha, No. 96 Fong Huang, No. 98 Mei Yuan, No. 99 Min Ghai, and No. 102 Yung Kang. These results indicate that the high proportion of genetic variation of Taiwanese tea plant varieties was contributed by Taiwanese wild *Camellia* and a small amount of other varieties. The observation that some important cultivated varieties had high similarities is consistent with the results of Huang et al [21], who performed a genetic diversity analysis on Oolong tea varieties in the three regions of Fujian, Taiwan, and Guangdong using amplified fragment length polymorphism, and observed that the genetic diversity within the population in Taiwan was the smallest.

The phylogenetic tree calculated with the UPGMA showed that the genetic divergence of several important tea plant varieties in Taiwan was small. Except for the foreign tea sample TJK90, which was collected from Mae Salong in Thailand, all the tea plant samples were categorized into one cluster when the genetic distance was approximately 0.055 (Figure 3). A genetic distance of ~0.009 (red line) appears to be a good boundary to determine the made tea sample cluster based on varieties because, at this time, important tea plants in Taiwan can be successfully divided into independent clusters according to their variety, such as the 11 made teas that belong to TTES No. 12 Chinhsuan [clusters (1) and (5) as shown in Figure 3] and the five made teas of No. 107 Chin Shin Oolong [clusters (2) and (8) as shown in Figure 3]. These methods were also applied to domestic made teas, as shown in Figure 3, such that some domestic made teas were clustered together. The present study also investigated the applicability of using a phylogenetic tree constructed with ITS2 sequences to identify made tea provenances. This method aimed to explore whether a phylogenetic tree could successfully distinguish the provenances of the tested tea samples and determine whether the tea was produced in Taiwan (sample number starting with PJK) or was a made tea from a foreign variety (sample number starting with TJK). The results indicate that among the 65 tested made tea samples, there was no regular pattern for classifying teas produced overseas or in

**Figure 3 – ITS2 phylogenetic tree of eight important, distinctive, and widely cultivated tea varieties of Taiwan, 35 local made tea samples, and 30 foreign made tea samples. ITS2 = internal transcribed spacer 2.**

Taiwan. This result indicates that a phylogenetic tree constructed with ITS2 sequences, a clustering method based on the genetic distance, allows us to correctly identify the variety but not the provenance, and therefore is not a viable method (Figure 3). One possible reason for this is that certain tea varieties have been traded between regions. As a result, the same tea varieties can be found in different regions. For example, the DNA sequence similarities among teas from Taiwan, Fujian, Guangdong, and Vietnam may be a result from the flow-out of Taiwanese tea plants; in particular, TTES 12 and TTES 13.

### 3.4. Using a DNA barcode of ITS2 sequence variations and SNPs to identify made tea varieties and provenances

The present study further examined ITS2 sequence variations of the made tea samples of identical varieties but different provenances to investigate the association between SNPs and provenance. The made teas of Chinhsuan varieties analyzed in the present experiment were the 18 tea samples from seven different production areas, including New Taipei City (Pinglin District), Nantou County (Mingjian Township and Jhushan Township), Chiayi County (Alishan), Taitung County, Yilan County, and Hualien County, and 1 tea plant sample of the Chinhsuan variety preserved by the TRES. An ITS2 sequence similarity analysis was performed on these 19 samples, and the results indicated that the made tea samples of the Chinhsuan variety had sequence differences. In general, the ITS2 sequence similarity of Chinhsuan made tea samples ranged from 0.910 to 1. The obtained ITS2 sequences through sequencing of the Chinhsuan made tea samples from the seven production areas were all different from the determined sequence of the Chinhsuan variety tea plant housed by the TRES (Table 3). A further analysis of the Chinhsuan made tea samples showed that the site and frequency of SNPs of the ITS2 sequences of samples from different production areas were also slightly different (Table 4). For example, the three tea samples PJK0075–0077 planted in Nantou County (Jhushan Township) had only two SNPs, and the frequency was 1.3% (2/160). The samples with the highest frequency were the two tea samples (PJK570 and PJK574) produced in Pinglin, and the frequency was 9.4% (15/160). The results of the SNP analysis of the made teas of identical varieties from different production areas indicated that only the ITS2 sequences of the three tea samples (PJK1034, PJK1036, and PJK1038) from Hualien were identical. In addition, for the tested made tea samples of the Chin Shin Oolong variety, we also observed that the site and frequency of SNPs of the ITS2 sequences of samples collected from different production areas were also slightly different. The tea samples (PJK1069–1071) produced at Lishan had the lowest frequency, which was 2.5% (4/160). The PJK452–PJK454 samples collected at Taitung had 23 SNPs, accounting for 14.4% (23/156) of the total 160 sites, and had the highest frequency of the two varieties of made teas from all production areas. The average SNP frequency of Chin Shin Oolong from all production areas in Taiwan was 8.8% (14.1/160), which was higher than the average frequency of 5.1% (8.1/160) of Chinhsuan from all production areas. However, the average SNP frequency of foreign Chinhsuan made teas

**Table 3 — Analysis of ITS2 sequence similarity of Chinhsuan made tea and tea varieties housed at the TRES.**

| Seq-> | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01. 12_TTES | ID | 0.981 | 0.975 | 0.968 | 0.943 | 0.900 | 0.900 | 0.950 | 0.950 | 0.962 | 0.900 | 0.956 | 0.956 | 0.881 | 0.975 | 0.893 | 0.893 | 0.893 | 0.343 |
| 02. 12_ZS_PJK0075 | 0.981 | ID | 0.993 | 0.987 | 0.950 | 0.906 | 0.906 | 0.956 | 0.956 | 0.968 | 0.906 | 0.962 | 0.962 | 0.887 | 0.993 | 0.900 | 0.900 | 0.900 | 0.343 |
| 03. 12_ZS_PJK0076 | 0.975 | 0.993 | ID | 0.993 | 0.956 | 0.912 | 0.912 | 0.962 | 0.962 | 0.975 | 0.912 | 0.968 | 0.968 | 0.893 | 1.000 | 0.906 | 0.906 | 0.906 | 0.343 |
| 04. 12_ZS_PJK0077 | 0.968 | 0.987 | 0.993 | ID | 0.962 | 0.918 | 0.918 | 0.956 | 0.956 | 0.968 | 0.918 | 0.962 | 0.962 | 0.900 | 0.993 | 0.912 | 0.912 | 0.912 | 0.337 |
| 05. 12_MJ_PJK130 | 0.943 | 0.950 | 0.956 | 0.962 | ID | 0.956 | 0.956 | 0.968 | 0.956 | 0.956 | 0.943 | 0.962 | 0.962 | 0.925 | 0.956 | 0.950 | 0.950 | 0.950 | 0.325 |
| 06. 12_MJ_PJK131 | 0.900 | 0.906 | 0.912 | 0.918 | 0.956 | ID | 1.000 | 0.937 | 0.925 | 0.925 | 0.987 | 0.918 | 0.931 | 0.956 | 0.912 | 0.993 | 0.993 | 0.993 | 0.300 |
| 07. 12_MJ_PJK143 | 0.900 | 0.906 | 0.912 | 0.918 | 0.956 | 1.000 | ID | 0.937 | 0.925 | 0.925 | 0.987 | 0.918 | 0.931 | 0.956 | 0.912 | 0.993 | 0.993 | 0.993 | 0.300 |
| 08. 12_TD_PJK448 | 0.950 | 0.956 | 0.962 | 0.956 | 0.968 | 0.937 | 0.937 | ID | 0.987 | 0.962 | 0.925 | 0.981 | 0.968 | 0.906 | 0.962 | 0.931 | 0.931 | 0.931 | 0.337 |
| 09. 12_TD_PJK450 | 0.950 | 0.956 | 0.962 | 0.956 | 0.956 | 0.925 | 0.925 | 0.987 | ID | 0.962 | 0.912 | 0.981 | 0.968 | 0.893 | 0.962 | 0.918 | 0.918 | 0.918 | 0.343 |
| 10. 12_TD_PJK451 | 0.962 | 0.968 | 0.975 | 0.968 | 0.956 | 0.925 | 0.925 | 0.962 | 0.962 | ID | 0.912 | 0.981 | 0.993 | 0.881 | 0.975 | 0.918 | 0.918 | 0.918 | 0.343 |
| 11. 12_PL_PJK570 | 0.900 | 0.906 | 0.912 | 0.918 | 0.943 | 0.987 | 0.987 | 0.925 | 0.912 | 0.912 | ID | 0.906 | 0.918 | 0.956 | 0.912 | 0.993 | 0.993 | 0.993 | 0.306 |
| 12. 12_PL_PJK574 | 0.956 | 0.962 | 0.968 | 0.962 | 0.962 | 0.918 | 0.918 | 0.981 | 0.981 | 0.981 | 0.906 | ID | 0.987 | 0.887 | 0.968 | 0.912 | 0.912 | 0.912 | 0.350 |
| 13. 12_YL_PJK907 | 0.956 | 0.962 | 0.968 | 0.962 | 0.962 | 0.931 | 0.931 | 0.968 | 0.968 | 0.993 | 0.918 | 0.987 | ID | 0.887 | 0.968 | 0.925 | 0.925 | 0.925 | 0.343 |
| 14. 12_YL_PJK908 | 0.881 | 0.887 | 0.893 | 0.900 | 0.925 | 0.956 | 0.956 | 0.906 | 0.893 | 0.881 | 0.956 | 0.887 | 0.887 | ID | 0.893 | 0.950 | 0.950 | 0.950 | 0.306 |
| 15. 12_YL_PJK909 | 0.975 | 0.993 | 1.000 | 0.993 | 0.956 | 0.912 | 0.912 | 0.962 | 0.962 | 0.975 | 0.912 | 0.968 | 0.968 | 0.893 | ID | 0.906 | 0.906 | 0.906 | 0.343 |
| 16. 12_HL_PJK1034 | 0.893 | 0.900 | 0.906 | 0.912 | 0.950 | 0.993 | 0.993 | 0.931 | 0.918 | 0.918 | 0.993 | 0.912 | 0.925 | 0.950 | 0.906 | ID | 1.000 | 1.000 | 0.306 |
| 17. 12_HL_PJK1036 | 0.893 | 0.900 | 0.906 | 0.912 | 0.950 | 0.993 | 0.993 | 0.931 | 0.918 | 0.918 | 0.993 | 0.912 | 0.925 | 0.950 | 0.906 | 1.000 | ID | 1.000 | 0.306 |
| 18. 12_HL_PJK1038 | 0.893 | 0.900 | 0.906 | 0.912 | 0.950 | 0.993 | 0.993 | 0.931 | 0.918 | 0.918 | 0.993 | 0.912 | 0.925 | 0.950 | 0.906 | 1.000 | 1.000 | ID | 0.306 |
| 19. 12_A_PJK1136 | 0.343 | 0.343 | 0.343 | 0.337 | 0.325 | 0.300 | 0.300 | 0.337 | 0.343 | 0.343 | 0.306 | 0.350 | 0.343 | 0.306 | 0.343 | 0.306 | 0.306 | 0.306 | ID |

ITS2 = internal transcribed spacer 2; TRES = Tea Research and Extension Substation.

| Table 4 – Analysis of ITS2 sequence SNPs of made teas from different production origins. | | | | | |
|---|---|---|---|---|---|
| **Imported from overseas** | | | **Domestic to Taiwan** | | |
| Code | Geographical origin | SNPs (%) | Code | Geographical origin | SNPs (%) |
| 12_MC_TJK65, 97–98 | Moc Chau, Sonla (N. Vietnam) | 10.0 (16/160) | 12_ZS_PJK0075–0077 | Jhushan, Nantou | 1.3 (2/160) |
| 12_BL_TJK66, 83, 94–95, 99, 104 | Bao Loc, Lamdong (S. Vietnam) | 13.1 (21/160) | 12_MJ_PJK130–131,143 | Mingjian, Nantou | 4.4 (7/160) |
| | Ave. | 11.6 (18.5/160) | 12_TD_PJK448, 450–451 | Taitung | 4.4 (7/160) |
| | | | 12_PL_PJK570, 574 | Pinglin, New Taipei | 9.4 (15/160) |
| | | | 12_YL_PJK907–909 | Yilan | 12.5 (20/160) |
| | | | 12_HL_PJK1034, 1036, 1038 | Hualien | 0.0 (0/160) |
| | | | | Ave. | 5.1 (8.1/160) |
| | | | 107_LG_ PJK190–191 | Lugu, Nantou | 6.9 (11/160) |
| | | | 107_A_PJK402–404 | Alishan, Chiayi | 4.4 (7/160) |
| | | | 107_TD_PJK452–454 | Taitung | 14.4 (23/160) |
| | | | 107_PL_PJK550, 552 | Pinglin, New Taipei | 10.6 (17/160) |
| | | | 107_HL_PJK946–948 | Hualien | 8.1 (13/160) |
| | | | 107_LS_PJK1069–1071 | Lishan, Taichung | 7.8 (4/160) |
| | | | | Ave. | 8.8 (14.1/160) |

ITS2 = internal transcribed spacer 2; SNP = single-nucleotide polymorphism.

was 11.6% (18.5/160), which was not only approximately two times the average value of Chinhsuan made teas from all production areas in Taiwan, but also higher than the detected average value of Chin Shin Oolong from all production areas in Taiwan (Table 4).

Among the local and foreign made teas of both the Chinhsuan and the Chin Shin Oolong varieties, nucleotide variation in the ITS2 sequence was observed in all the made teas from different production areas; therefore, further multiple sequence alignment analyses of the ITS2 sequences of identical-variety local and imported teas were performed to locate unique SNP variations representative of particular production areas. Figure 4A shows the multiple ITS2 sequence alignment results of 31 made tea samples, including Chinhsuan variety housed at the TRES, eight local (from 7 production areas) samples, and 12 foreign (from 5 production areas) samples. Figure 4B shows the multiple ITS2 sequence alignment result of three made tea samples, including the Tzuiyu variety housed at the TRES and two foreign (from 2 production areas) teas. Figure 4C illustrates the multiple ITS2 sequence alignment result of 24 made tea samples, including the Chin Shin Oolong variety housed at the TRES, and 17 local (from 7 production areas) and six foreign (from 5 production areas) teas. After comparison, we observed that the length of all ITS2 sequences of these made teas was 160 bp; however, the sites and frequencies of SNPs observed in these three varieties were not identical. No. 13 Tzuiyu had only 14 sites, accounting for 8.8% of the total sites. No. 107 Chin Shin Oolong had the most SNPs (53), accounting for 33.1% of the total sites. No. 12 Chinhsuan had 32 SNPs, accounting for 20.0% of the total sites (Table 3). A further analysis individually examining SNP sites in these three varieties of Chinhsuan, Tzuiyu, and Chin Shin Oolong teas observed that 42 SNPs (belonging to 13 made tea samples) were the characteristic sites of made tea samples of particular production areas. SNPs can be used to trace the provenances of made tea samples of particular varieties. For the tea samples of the Chinhsuan variety, a G to A SNP variant at 112 (c. 112G>A) was observed in the TRES Chinhsuan tea

sample. For the TJK96 sample, which was a foreign tea from Da Lat, Lamdong (S. Vietnam), a G to C (c. 3G>C) SNP variant at 3 was observed. In addition, in the TJK90 tea sample from Mae Salong (Thailand), 11 production area-associated characteristic SNPs were observed. This was the largest number of SNPs found in any of the 13 made tea samples, in which production area-characteristic SNPs were observed (Table 4). Although we observed that the SNPs of ITS2 sequences could be used to characteristically label the provenances of the made tea samples analyzed, it is notable that if all samples in Figure 3 were assembled for analysis, compared with analyzing made tea samples from only one variety, the number of varieties and samples increased and the SNP sites observed decreased. For the 13 made tea samples, of which SNPs used to label provenance were previously observed, nine of these made tea samples lost their characteristic SNPs used to identify provenance; therefore, these samples are not listed (Table 4). A further examination of the remaining made tea samples with characteristic SNPs to identify provenance indicated that the variation sites of the two made tea samples, No. 12 and PJK453, shifted, although their variation features remained consistent (Table 5).

The present study successfully sequenced the ITS2 sequence of 113 tea plant varieties. In this PCR-amplified noncoding region fragment, no tea plant varieties share identical sequences. The molecular fingerprint library established using the ITS2 sequences can successfully identify all tea plant varieties housed at the tea plantation, and they are good core molecular marker targets. These results agree with the previous conclusion that ITS2 was successfully used for between- and within-species identification [28].

A chart suggesting the development of a database for comparing Taiwanese local teas and other types of tea is given in Figure 5. Briefly, the genomic DNA is extracted from the analyzed tea samples; BEL-1/BEL-3 primer pairs can be used to amplify the ITS2 fragment for sequencing. Then, the sequence can be used to perform a BLAST search in an open or in-house ITS2 sequence database of tea plant varieties and made tea
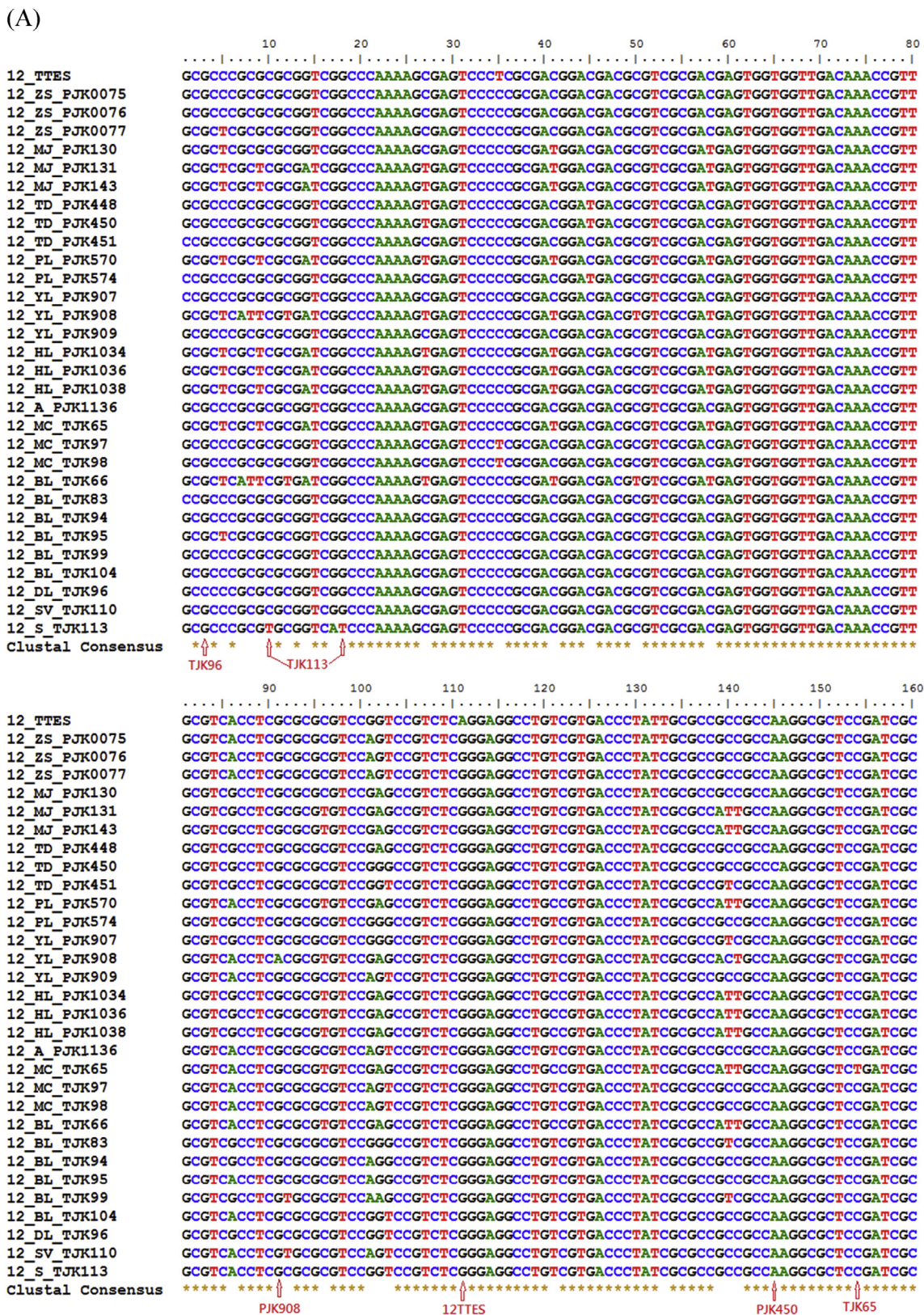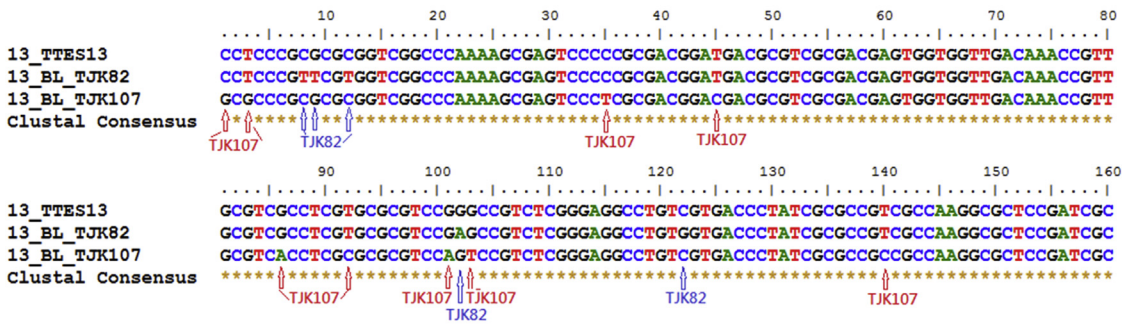
(A)



Figure 4 – Variations of ITS2 sequences and their SNP sites of the made tea samples of three different varieties.
ITS2 = internal transcribed spacer 2; SNP = single-nucleotide polymorphism; A = Adenosine; T = Thymine; C = Cytosine;
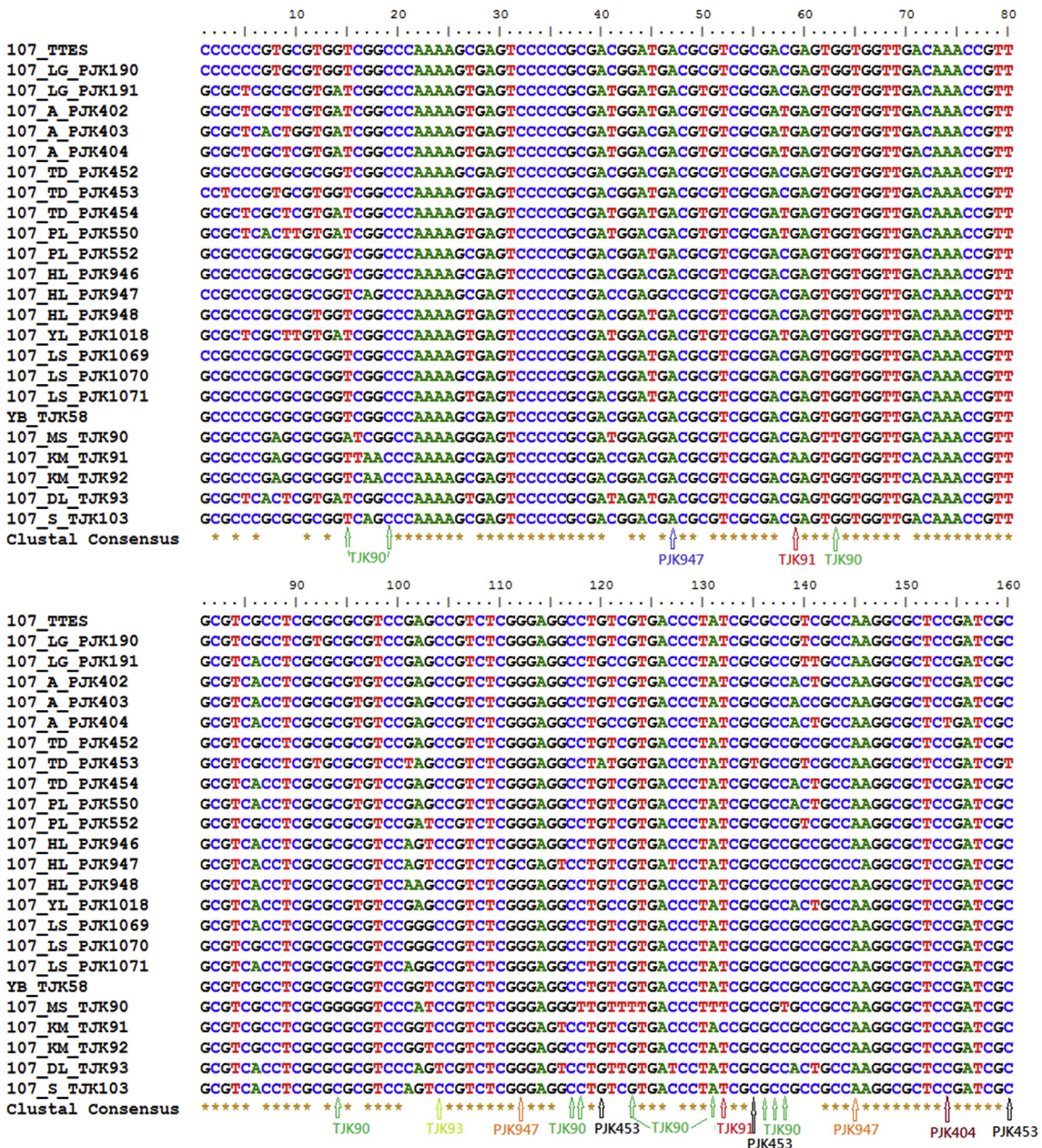G = Guanine.

(B)



(C)



**Figure 4 — (continued).**

**Table 5 – Characteristic SNPs of tea samples from a particular production origin.**

| Code | Geographical origin | Origin | SNPs |
|---|---|---|---|
| (1) TTES No. 12 (Chinhsuan) | | | |
| 12 | TTES | Domestic | c. 112G>A |
| TJK65 | Moc Chau, Sonla (N. Vietnam) | Imported | c. 154C>T |
| TJK96 | Da Lat, Lamdong (S. Vietnam) | Imported | c. 3G>C |
| TJK113 | Sumatra (Indonesia) | Imported | c. 10C>T, c. 18G>T |
| PJK450 | Taitung | Domestic | c. 145A>C |
| PJK908 | Yilan | Domestic | c. 92G>A |
| (2) TTES No. 13 (Tzuiyu) | | | |
| TJK82 | Bao Loc, Lamdong (S. Vietnam) | Imported | c. 8C>T, c. 9G>T, c. 12C>T, c. 102G>A, c. 122C>G |
| TJK107 | Bao Loc, Lamdong (S. Vietnam) | Imported | c. 1C>G, c. 3T>G, c. 35C>T, c. 45T>C, c. 86G>A, c. 92T>C, c. 101G>A, c. 103G>T, c. 140T>C |
| (3) Chin Shin Oolong | | | |
| TJK90 | Mae Salong (Thailand) | Imported | c. 15T>A, c. 19C>G, c. 63G>T, c. 94C>G, c. 118C>G, c. 119C>T, c. 123G>T, c. 131A>T, c. 136G>C, c. 137C>G, c. 138C>T |
| TJK91 | Kota Medan, Sumatra (Indonesia) | Imported | c. 59G>A, c. 132T>C |
| TJK93 | Da Lat, Lamdong (S. Vietnam) | Imported | c. 104C>T |
| PJK404 | Alishan, Chiayi | Domestic | c. 154C>T |
| PJK453 | Taitung | Domestic | c. 120G>A, c. 135C>T, c. 160C>T |
| PJK947 | Hualien | Domestic | c. 47A>C, c. 112G>C, c. 145A>C |
| | | | Subtotal: 42 |
| (4) All[a] | | | |
| 12 | TTES No. 12 | Domestic | c. 101G>A |
| TJK90 | Mae Salong (Thailand) | Imported | c. 15T>A, c. 63G>T, c. 94C>G, c. 96C>G, c. 117C>G, c. 118C>T, c. 123G>T, c. 137C>G, c. 138C>T |
| TJK91 | Kota Medan, Sumatra (Indonesia) | Imported | c. 132T>C |
| PJK453 | Taitung | Domestic | c. 120G>A, c. 135C>T, c. 160C>T |
| PJK908 | Yilan | Domestic | c. 91G>A |
| 25 | TTES No. 25 | Domestic | c. 90C>T |
| 112 | TTES N0. 112 | Domestic | c. 101G>C |
| TJK61 | Assam indigenous | Imported | c. 39G>C, c. 88C>T |
| TJK67 | Fu Bei Zhong | Imported | c. 119T>A |
| TJK68 | Fu Ding Da Bai | Imported | c. 82C>G, c. 92C>A, c. 94C>T |
| TJK71 | N/A (Green Tea) | Imported | c. 114A>G |

ITS2 = internal transcribed spacer 2; SNP = single-nucleotide polymorphism.
[a] All the samples used in Figure 3.

production origins. If an identical sequence is observed, then a further examination can determine whether this sequence is from a particular production area. If the answer is confirmed, then the variety and production area of the analyzed tea sample can be determined. For those samples in which no matched sequence was 100%, we suggest that a phylogenetic tree constructed using ITS2 sequences from 113 tea plant varieties housed at the TRES can be used as a model system with genetic distances used to create clusters while comparing sequences or SNPs to determine the variety of the sample. In addition to uploading the variety data obtained from the sequence analysis to the ITS2 sequences (ITS2 SEQs) database, the physical and chemical characteristics of the tea can further be used to determine the production area. It is true that the gas chromatography–mass method successfully distinguished sun-dried Pu-erh green tea from different tea mountains [41]. Coincidentally, near-infrared spectroscopy analysis was used to discriminate the geographical origins of

Oolong tea [42]. These methods can distinguish the origins of made teas; however, they did not mention whether those tea samples belong to the same cultivars. We suggest using the sequence ITS2 fragment first and, in case of confusion, a further analysis by gas chromatography–mass or near-infrared spectroscopy. These are particularly useful for those sequences for which a specific sequence has been observed in the database by the BLAST search, but which are labeled with more than one production area.

## 4.    Conclusion

The DNA barcode method can be used as a frontline inspection method to compare ITS2 sequences across databases, to identify made tea sample variety and provenance and fulfill the demand of inspecting a large amount of made teas.

PCR
(BEL-1/BEL-3)

↓

ITS2 SEQs Bank ←

↓

Phylogenic tree ⇌ Variety —

↓

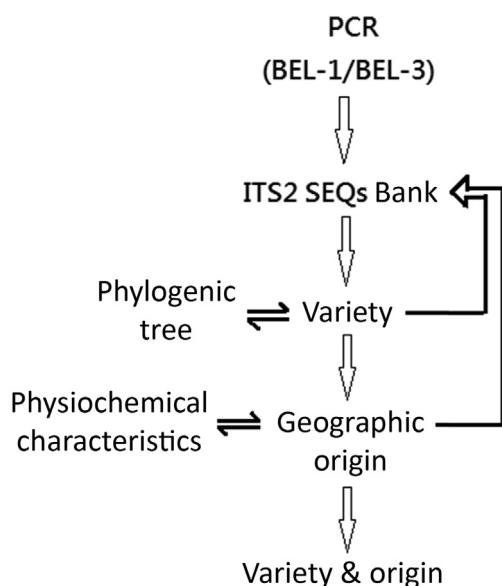Physiochemical characteristics ⇌ Geographic origin —

↓

Variety & origin

**Figure 5 – Flowchart of DNA barcode for made tea samples using ITS2 sequence database. ITS2 = internal transcribed spacer 2; PCR = polymerase chain reaction.**

## Conflicts of interest

All authors declare no conflicts of interest.

## Acknowledgments

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jfda.2016.06.008.

REFERENCES

[1] Yang CS, Landau JM. Effects of tea consumption on nutrition and health. J Nutr 2000;130:2409–12.

[2] Farzaneh V, Carvalho IS. A review of the health benefit potentials of herbal plant infusions and their mechanism of actions. Ind Crops Prod 2015;65:247–58.

[3] Suzuki T, Kumazoe M, Kim Y, Yamashita S, Nakahara K, Tsukamoto S, Sasaki M, Hagihara T, Tsurudome Y, Huang Y, Maeda-Yamamoto M, Shinoda Y, Yamaguchi W, Yamada K, Tachibana H. Green tea extract containing a highly absorbent catechin prevents diet-induced lipid metabolism disorder. Sci Rep 2013;3:2749.

[4] Hayat K, Iqbal H, Malik U, Bilal U, Mushtaq S. Tea and its consumption: benefits and risks. Crit Rev Food Sci Nutr 2015;55:939–54.

[5] Khan N, Mukhtar H. Tea and health: studies in humans. Curr Pharm Des 2013;19:6141–7.

[6] Wachira FN, Kamunya S, Karori S, Chalo R, Maritim T. The tea plants botanical aspects. In: Preedy V, editor. Tea in health and disease prevention. Chapter 1. London: Academic Press; 2012. p. 3–17.

[7] Yao MZ, Chen L. Tea germplasm and breeding in China. In: Chen L, Apostolides Z, Chen ZM, editors. Global tea breeding: achievements, challenges and perspectives [advanced topics in science and technology in China]. Chapter 2. Berlin: Springer Verlag; 2012. p. 13–68.

[8] Ho CT, Lin JK, Shahidi F. Tea and tea products: chemistry and health-promoting properties [nutraceutical science and technology]. Boca Raton, FL: CRC Press; 2008.

[9] Chang M. Make bogus Taiwanese tea nowhere to hide: from the foreign tea import to the discussion of labeling the production origin on the package. Taiwan Econ Res Monthly 2008;31:120–4 [in Chinese].

[10] Luykx DMAM, van Ruth SM. An overview of analytical methods for determining the geographical origin of food products. Food Chem 2008;107:897–911.

[11] Ye NS. A minireview of analytical methods for the geographical origin analysis of teas (Camellia sinensis). Crit Rev Food Sci Nutr 2012;52:775–80.

[12] Ma G, Zhang Y, Zhang J, Wang G, Chen L, Zhang M, Liu T, Liu X, Lu C. Determining the geographical origin of Chinese green tea by linear discriminant analysis of trace metals and rare earth elements: taking Dongting Biluochun as an example. Food Control 2016;59:714–20.

[13] Košir IJ, Kocjančič M, Ogrinc N, Kidrič J. Use of SNIF-NMR and IRMS in combination with chemometric methods for the determination of chaptalization and geographical origin of wines [the example of Slovenian wines]. Anal Chim Acta 2001;429:195–206.

[14] Du H, Liang Y, Pei K, Ma K. UV radiation-responsive proteins in rice leaves: a proteomic analysis. Plant Cell Physiol 2011;52:306–16.

[15] Chen L, Zhou ZX, Yang YJ. Genetic improvement and breeding of tea plant (Camellia sinensis) in China: from individual selection to hybridization and molecular breeding. Euphytica 2007;154:239–48.

[16] Su MH, Tsou CH, Hsieh CF. Morphological comparisons of Taiwan native wild tea plant (Camellia sinensis (L.) O. Kuntze forma formosensis Kitamura) and two closely related taxa using numerical methods. Taiwania 2007;52:70–83.

[17] Chen IZ. Taiwan tea breeding. Seed Nursery (Taiwan) 2006;8:1–20 [in Chinese with English abstract].

[18] Lai JA, Yang WC, Hsiao JY. An assessment of genetic relationships in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers. Bot Bull Acad Sin 2001;42:93–100.

[19] Lin SY, Chen IZ, Tsai CM, Chen YL. Detection of genetic relationship in Taiwan tea variety (Camellia sinensis (L.) O Kuntze) with RAPD markers. J Chin Soc Hortic Sci 2005;51:357–66 [in Chinese with English abstract].

[20] Yao MZ, Chen L, Liang YR. Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. Plant Breeding 2008;127:166–72.

[21] Huang FP, Liang YR, Lu JL, Chen RB, Mamati GE, Sun QL. Evaluation of genetic diversity in oolong tea germplasms by AFLP fingerprinting. J Tea Sci 2004;24:183–9 [in Chinese with English abstract].

[22] Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 2003;270:313—21.

[23] China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci U S A 2011;108:19641—6.

[24] Chen S, Pang X, Song J, Shi L, Yao H, Han J, Leon C. A renaissance in herbal medicine identification: from morphology to DNA. Biotechnol Adv 2014;32:1237—44.

[25] Vancouver JF. What lies in the future of tea science in the era of genomics? J Tea Sci Res 2011;1:1—4.

[26] Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep 2011;1:42.

[27] Li M, Wong KL, Chan WH, Li JX, But PPH, Cao H, Shaw PC. Establishment of DNA barcodes for the identification of the botanical sources of the Chinese 'cooling' beverage. Food Control 2012;25:758—66.

[28] Baldwin BG. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Mol Phylogenet Evol 1992;1:3—16.

[29] Lee OR, Kim MK, Yang DC. Authentication of medicinal plants by SNP-based multiplex PCR. In: Sucher NJ, Hennell JR, Carles MC, editors. Plant DNA fingerprinting and barcoding: methods and protocols. New York: Humana Press; 2012. p. 135—47.

[30] Fang WP, Meinhardt LW, Tan HW, Zhou L, Mischke S, Zhang D. Varietal identification of tea (Camellia sinensis) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Hortic Res 2014;1:14035.

[31] Pan R, Chen HP, Zhang ML, Wang QH, Jiang Y, Liu X. Dissipation pattern, processing factors, and safety evaluation for dimethoate and its metabolite (Omethoate) in tea (Camellia Sinensis). PLoS One 2015;10:e0138309.

[32] Almakarem ASA, Heilman KL, Conger HL, Shtarkman YM, Rogers SO. Extraction of DNA from plant and fungus tissues in situ. BMC Res Notes 2012;5:266.

[33] Chiou SJ, Yen JH, Fang CL, Chen HL, Lin TY. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. Planta Med 2007;73:1421—6.

[34] Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 1991;17:1105—9.

[35] Han J, Zhu Y, Chen X, Liao B, Yao H, Song J, Chen S, Meng F. The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. BioMed Res Int 2013;2013:741476.

[36] Pang X, Shi L, Song J, Chen X, Chen S. Use of the potential DNA barcode ITS2 to identify herbal materials. J Nat Med 2013;67:571—5.

[37] Wang XC, Liu C, Huang L, Bengtsson-Palme J, Chen H, Zhang JH, Cai D, Li JQ. ITS1: a DNA barcode better than ITS2 in eukaryotes? Mol Ecol Resour 2015;15:573—86.

[38] Selvaraj D, Shanmughanandhan D, Sarma RK, Joseph JC, Srinivasan RV, Ramalingam S. DNA barcode ITS effectively distinguishes the medicinal plant Boerhavia diffusa from its adulterants. Genomics Proteomics Bioinformatics 2012;10:364—7.

[39] Wang LC, Ho HF. Survey and utilization of Taiwan wild tea plants. Proceedings of A Symposium on the Experimental Results of Special Crops. 1997. p. 71—80 [in Chinese with English abstract].

[40] Cai YR, Lin SF. Using ISSR DNA markers to identify varieties and evaluate genetic diversity of tea germplasm in Taiwan. Chin Agric Sci Bull 2007;23:380—5 [in Chinese with English abstract].

[41] Wu Y, Lv S, Wang C, Gao X, Li J, Meng Q. Comparative analysis of volatiles difference of Yunnan sun-dried Pu-erh green tea from different tea mountains: Jingmai and Wuliang mountain by chemical fingerprint similarity combined with principal component analysis and cluster analysis. Chem Cent J 2016;10:11.

[42] Yan SM, Liu JP, Xu L, Fu XS, Cui HF, Yun ZY, Yu XP, Ye ZH. Rapid discrimination of the geographical origins of an Oolong tea (Anxi-Tieguanyin) by near-infrared spectroscopy and partial least squares discriminant analysis. J Anal Meth Chem 2014;2014:704971.