TESTING THREE PROPOSED DNA REGIONS (*matK, rbcL* and *ITS2*) FOR IDENTIFICATION OF *CAMELLIA EUPHLEBIA* AND *CAMELLIA CHRYSANTHA*

Nguyen Van Viet[1], Pham Quang Chung[1], Do Quang Trung[1], Tran Viet Ha[1], Sounthone Douangmala[2]
[1]*Vietnam National University of Forestry*
[2]*Bolykhamxay College of Agriculture and Forestry, Vientiane, Laos*

SUMMARY

Camellia sp. is a yellow flower tea species that have high economic values and have been used as a nutritious beverage, medicine and an ornamental plant. The identification of these species are based on morphological characteristics and recently by molecular markers such as *mat*K, *ITS2*, *rbc*L. This paper shows that the nucleotide sequenc of three DNA regions (*matK*, *ITS2*, *rbcL*) can be used to identify *Camellia sp.* collected from different areas in Vietnam (Quang Ninh and Bac Giang).The results also show that those three locations can be amplified by directional PCR. In addition, the nucleotide differences among those DNA regions of the two Camellia species ranged from 0% to 0.82% suggesting the *Camellia chrysantha* (Ba Che, Quang Ninh) is a derivative of *Camellia euphlebia* (Son Dong, Bac Giang).

Keywords: *Camellia chrysantha, Camellia euphlebia*, DNA barcode, species identification, Yellow flower tea.

1. INTRODUCTION

The Yellow-flower tea plant belongs to the *Theaceae* family, *Camellia sp*. It has a diverse species of more than 300 and many different variants that have been reported all over the world. For example, 28 and 24 species of the yellow-flower tea plant were identified in China and Vietnam, respectively. In Vietnam, the *Camellia sp*. distributed in some areas such as Tam Dao national park (8 species) (Hakoda *et al.*, 2007), Cuc Phuong national park (2 species), Bac Giang (1 species), Ba Vi national park (1 species), and some other provinces such as Quang Ninh. Notably, the yellow-flower tea plant living in Ba Che (Quang Ninh) i originated from *C. euphlebia* (Ngo Thi Minh Duyen *et al.*, 2011).

The yellow-flower tea plant is of high economic value and has been used as medicinal plants because it contains some ingredients (Se, Ge, Mn, Mo, V, Zn and some other elements), which might play roles in many processes like health protection, anti-cancer, improvement of elasticity of blood vessels, regulation of cholesterol-activated enzymes, lower blood cholesterol, boosting the immune system (Luong Thinh Nghiep, 2000).

Currently, the number of individuals of the yellow-flower tea plant is becoming smaller and their distribution is going to be narrowed down due to eradicated exploitation as well as increasing stress conditions in habitats caused by climate change (Ninh T., 2007; Tran Ninh and Hakoda Naotoshi, 2010). These lead to the extinction of some medical plants and some endangered plant species affecting the sustainable supply of human pharmaceuticals. Moreover, studies on the biological, ecological and culturing characteristics of yellow-flower tea are still limited, incomplete and not synchronous. Preservation of these precious tea species has also been neglected, especially research on the development and application of this tea is almost nonexistent (Tran Ninh and Hakoda Naotoshi, 2010).

In Vietnam, the classification of *Camellia sp.* Is mostly based on morphological characteristics, which still have some problems and limitations, especially some species of *Camellia*that which has a similar morphology leading to difficulties for classification. Recently, some molecular markers (*matK*, *ITS2*, and *rbcL*) have been used to classify and identify some of *Camellia sp.* such as *C. sinensis, C. petelotii*, *C. yunnanensis*, *C. oleifera*, *C. taliensis*, *C. japonica*, *C. cuspidata*, *C. grandibracteata, C. albogigas.* Using the molecular markers together with morphological characteristics will increase with precision and rapidly identify the

difference among organisms. Furthermore, many research papers have reported the identification of many specific DNA regions that can be used as DNA barcoding, which is individually specific and able to recognize plants at different levels. Selecting the specific DNA region as DNA barcoding depends on a group of specific plants and research proposals (Kress *et al*., 2008). Interestingly, the *mat*K gene expressed in chloroplast has been reported as a molecular marker to identify species and under-species in many plant species (Yakawa *et al.*, 2006; Storchova *et al.*, 2007; Ford *et al.*, 2009).

In this paper, we report a nucleotide comparison of three DNA regions (*matK*, *ITS2*, *rbcL*) for two species of *Camellia*, *C. euphlebia* và *C. chrysantha*. These results can be used as a method to classify and identify species of Camellia that enhances the conservation efficiency and the development of valuable gene sources in Vietnam.

2. RESEARCH METHODOLOGY

2.1. Materials

DNA samples: 6 fresh leaves from each species were collected from natural forests Son Dong - Bac Giang and Ba Che - Quang Ninh.

Chemicals for total DNA extraction: CTAB, SDS, EDTA, Tris-HCl, NaCl, PVP, Ascorbic acid, Mercaptoethanol, Potassium acetate, Sodium acetate, and Ethanol from Wako (Japan) and Merck (Germany); PCR and electrophoresis from Fermentas (Germany), Bioneer (Korea), Research Organics (America).

Primer sets for PCR amplification of *mat*K, *rbc*L *và ITS2* (Table 1).

Table 1. Nucleotide sequence of primers

Genes	Sequence (5'-3')	Fragment size	Reference
*mat*K	Forward: ACCCAGTCCATCTGGAAATCTTGGTTC Reverse: CGTACAGTACTTTTGTGTTTACGAG	920 bp	Cuenoud *et al*., 2002
*rbc*L	Forward: GTAAAATCAAGTCCACCACG Reverse: GTAAAATCAAGTCCACCGCG	600 bp	Kress *et al*., 2007
ITS2	Forward: ATGCGATACTTGGTGTGAAT Reverse: TCCTCCGCTTATTGATATGC	550 bp	Chen *et al*., 2010

2.2. Methods

Plant samples: Samples were collected and labelled in order as following CeBG1, CeBG2, CeBG3, CeQN1, CeQN2 và CeQN3. Fresh-leaves samples immediately put into a freezed container and stored at -80^{0}C until DNA extraction.

Total DNA extraction: Total DNA was extracted by CTAB method (Saghai Maroof *et al.,* 1984). Briefly, about 100 mg sample was ground in 600 mL CTAB solution (2% CTAB, 20 mM EDTA, 1.4 M NaCl, 1% beta-mercaptoethanol, 100 mM Tris-HCl pH 8.0). Then, the total solution was transferred into 1.5 mL centrifuge tubes and incubated at 65^{0}C in 30 minutes. Next step, the same volume of chloroform was added and mixed by inverting tubes 10 times. All samples were centrifuged at 10000 rpm in 10 minutes. The supernatant was transferred into new 1.5 mL tubes and precipitated DNA by adding 500 µl cold isopropanol. The mixture was incubated at -20^{0}C for 2 hours before centrifuging at 10000 rpm in 10 minites. The pellet was then washed twice with 70% ethanol solution and dried at room temperature for 30 minutes. The pellet was dissolved in 50 µl TE buffer.

PCR amplification and DNA sequencing: The sequences of three genes (*matK*, *rbcL* và *ITS2*) were amplified by PCR using PCR machine (model: 9700 Thermal Cycler Applied Biosystems, American). The PCR master mix for one reaction (25 µl) consists of 2.5 µl 10X Taq buffer, 2.0 µl dNTPs (2.0 mM), 1.0 µl for each of forward and reverse primer (10 nM), 0.5 µl Taq DNA polymerase (5U/µl), 1.0 µl DNA templates (50 ng/µl), and water was added to get the final volume of 25 µl. The PCR

program as followed: 95^0C: 3 minutes, 40 cycles of {95^0C: 30s, 57^0C - 62^0C: 30s, 72^0C: 1 minute}, 72^0C: 7 minutes. PCR samples were kept at 4^0C and separated by electrophoresis on 1.2% agarose gel. The PCR product was then gel purified using a Norgen biotek kit (Canada) and sequenced by using ABI PRISM®3730xl DNA Analyzer (ABI, Foster City, CA, USA).

Analysis of DNA sequence (DNA barcode): The sequences were processed using BioEdit software (version 7.2.5) and blasted on NCBI website Using BLAST NUCLEOTIDE tool (*https://blast.ncbi.nlm.nih.gov/Blast.cgi)*.

3. RESULTS AND DICUSSION

3.1. DNA extraction and PCR amplification of three genes (*matK*, *rbcL* and *ITS2*)

Total DNA was extracted by CTAB methods as described by Saghai Maroof *et al.* (1984). The result was shown in figure 1. As can be seen, the specific bands were clean and have no contamination of protein and RNA. This suggests the total DNA can be used as template for PCR amplification.

The PCR results presented all three genes which were successfully amplified from six samples (Figure 2 and Table 2). These indicated that the good quality of extracted DNA and PCR products are clean enough for sequencing. Moreover, the results also showed the PCR procedure was optimized for amplification of the target DNA fragments.

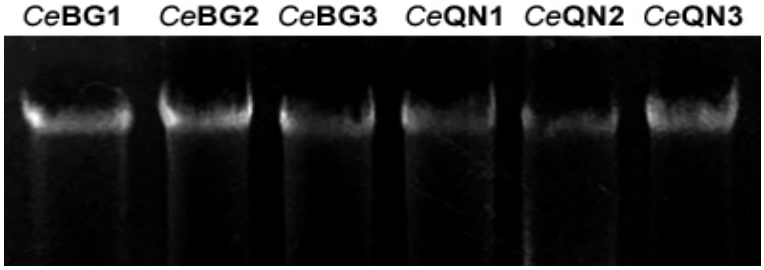

Figure 1. Genomic DNA extraction from six leaf samples

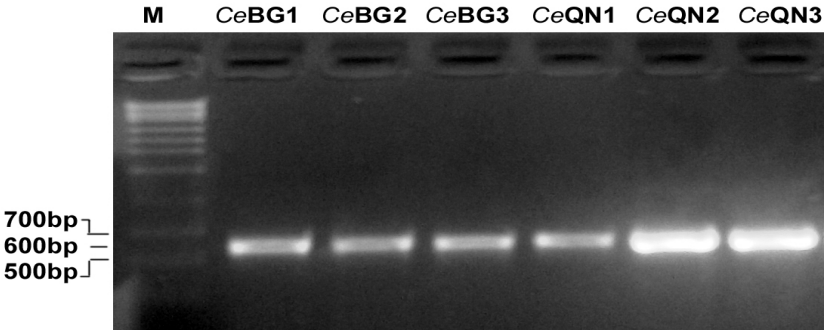

Figure 2. PCR amplification of *rbc*L gene

(DNA template from six leaf samples (CeBG1, CeBG2, CeBG3, CeQN1, CeQN2, CeQN3) were used. The bands as expected with 589 bp. M: DNA marker 1kb)

Table 2. Analysis of DNA sequence of three proposed genes

No.	Aspects	DNA region		
		matK	***rbcL***	***ITS2***
1	PCR amplification (%)	100	100	100
2	Correct Sequence (%)	100	100	100
3	Expected size (bp)	921	589	367
4	Different position	0	2	3
5	Number of different nucleotide	0	2	3
6	Distention (% different nucleotide)	0	0.34	0.82

3.2. Analysis of sequence of three DNA regions

3.2.1. Sequence analysis of matK gene

The sequence result of *mat*K gene from all Camellia samples showed that the PCR products were clean and not be affected by background. These sequences were then edited and aligned by using Bioedit software (ver. 7.2.5). The results indicated that the sequence matched with DNA regions that have 921 bp, 589 bp and 367 bp for *mat*K, *rbc*L và *ITS2*, respectively (Table 2).

In the next step, we compared the DNA region of three genes between two Camellia samples, one from Son Dong - Bac Giang and the other from Ba Che - Quang Ninh, to find the number and position of different nucleotides. These results can be useful for classification of species, in particular, confirming that *Camellia species* from Quang Ninh and Bac Giang are different species or belonging to the same species. Basically, it is known that the difference of nucleotide sequence of DNA barcoding regions among species is higher than that in one species, then this difference can be used as DNA barcoding for that species (Fabrizio *et al.,* 2011).

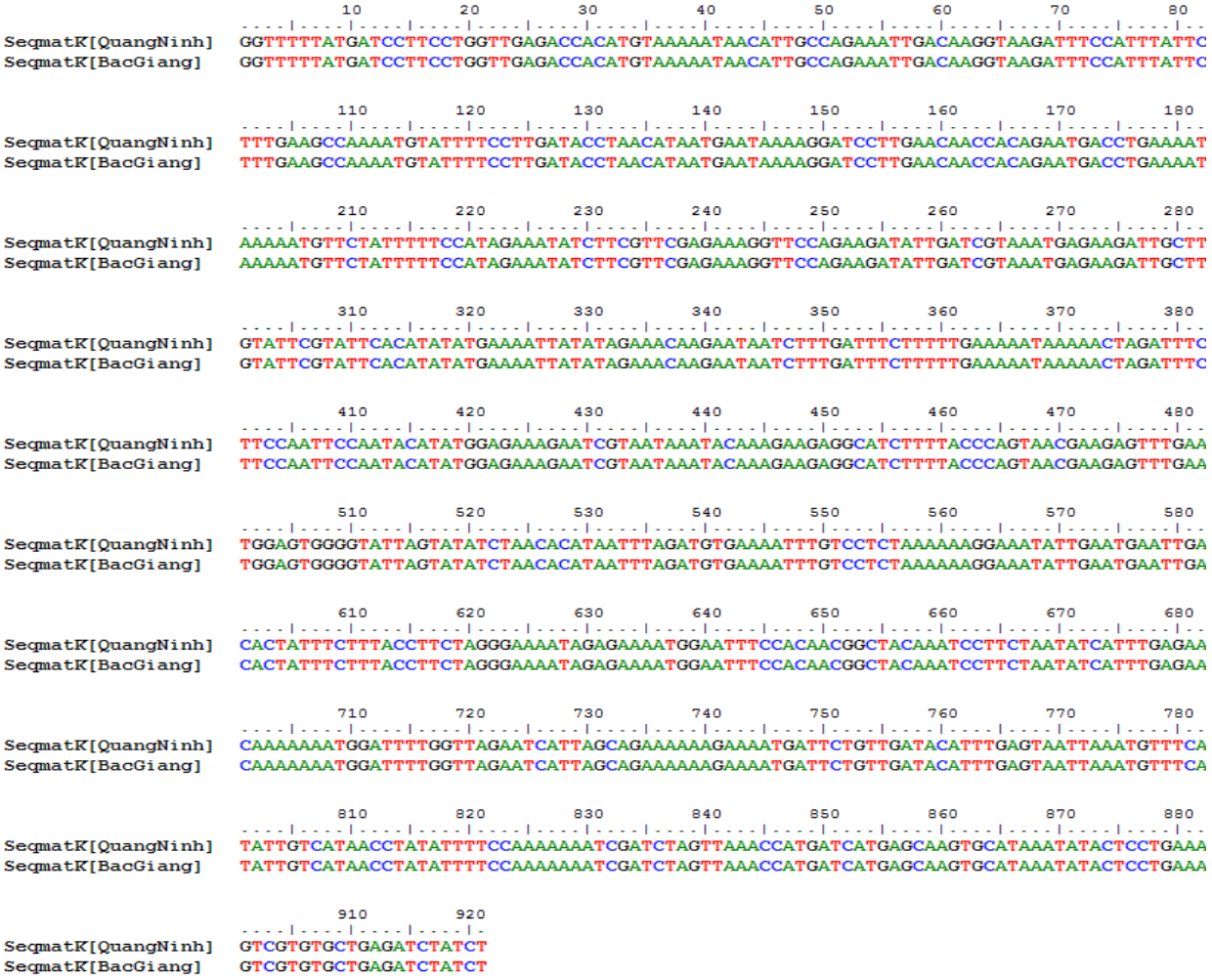

Figure 3. The alignment of *mat*K genes from Camellia samples of Quang Ninh and Bac Giang

As we expected, the alignment analysis of *mat*K DNA region from our studies with the *matK* sequence of *Camellia sp.* from GeneBank showed the difference is from 1 - 2%. This is consistent with previous studies. Interestingly, the alignment results of samples from Quang Ninh and Bac Giang presented 0% for the differences (Figure 3), which is normal for individuals in one species.

3.2.2. Sequence analysis of rbcL gene

We applied the same method to analyze sequence of *rbc*L gene. The results showed the

difference is from 1 - 3% for our *rbc*L sequence compared with *rbc*L sequence from the Genebank. This is consistent with previous studies. However, the nucleotide difference between two Camellia samples from Quang Ninh and Bac Giang is 0.34% suggesting two samples are from one *Camellia species*.

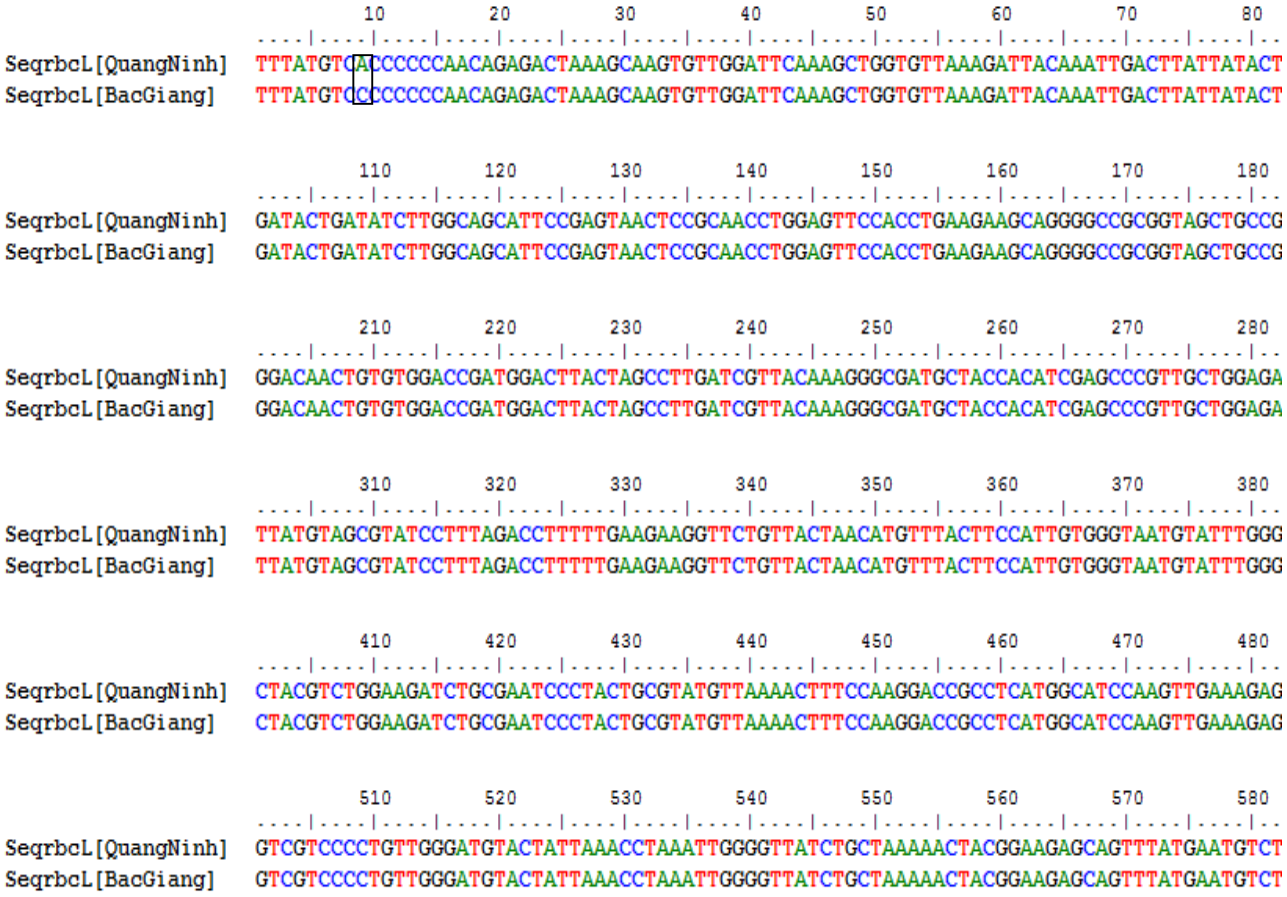

Figure 4. The alignment of *rbc*L genes from Camellia samples of Quang Ninh and Bac Giang

3.2.3. Sequence analysis of ITS2 gene

In next step, the same method was applied to analyze *ITS2* sequences. The alignment of our *ITS2* sequencewith the one from Genebank showed the difference is from 1 - 3%, while the nucleotide difference between two Camellia samples from Quang Ninh and Bac Giang is 0.82% suggesting two samples are from one *Camellia species*.

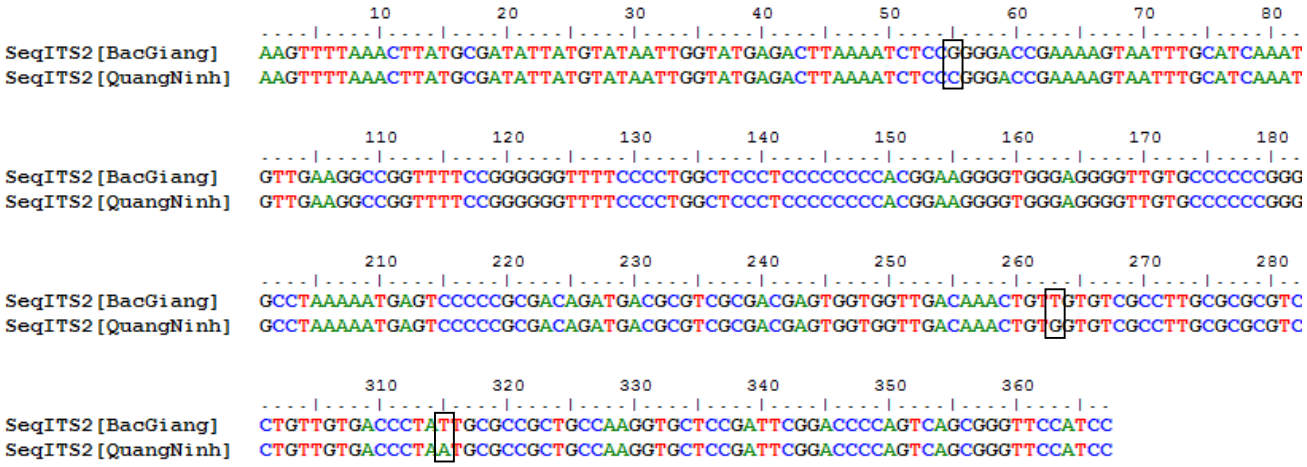

Figure 5. The alignment of *ITS2* genes from Camellia samples of Quang Ninh and Bac Giang

All in all, our results suggest that among proposed DNA regions, *mat*K and *ITS2* are specific for an evolution in which the nucleotide difference is enough for species identification, meanwhile, *rbc*L is specific for the highly conserved evolution that is normally used for identification of individuals within a species (CBOL, 2009). All three DNA regions presented the nucleotide sequence for samples from Quang Ninh and Bac Giang that belong to a range for individuals of one species. That result is consistent with the study of Ngo Thi Minh Duyen *et al.* (2011).

4. CONCLUSION

All three proposed genes (*matK, rbcL* và *ITS2*) were successfully amplified and sequenced. The nucleotide difference between *Camellia species* from Quang Ninh and Bac Giang ranged from 0 - 0.82% The result initially confirmed that *Camellia* samples from Quang Ninh and Bac Giang are the same species.

REFERENCES

1. CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc Natl Acad Sci USA,* 106: 12794-12797.

2. Chen S (1), Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5(1): e8613.

3. Cuenoud P, Savolainen V, Chatrou L.W, Powell M, Grayer R.J, Chase MW (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *American Journal of Botany*, 89: 132-144.

4. Fabrizio De Mattia, Ilaria Bruni, Andrea Galimberti, Francesca Cattaneo, Maurizio Casiraghi, Massimo Labra (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Research International*, 44: 693-702.

5. Ford C.S, Ayres K.L, Toomey N, Haider N, Stahl J.V.A, Laura J, Kelly, Wikstrom N, Peter M, Hollingsworth R, Dupp R.J, Sarah B, Hoot R.S, Cowan, Mark W, Chase, Mike J, Wilkinson. (2009). Selection of candidate coding DNA barcoding regions for use on DNA plants. *Botanical Journal of the Linnean Society*, 159 (1): 1-11

6. Hakoda N, Kirino S.H, Ninh T (2007). New species of genus camellia in Vietnam. *International Camellia Journal*, 39:54-57.

7. Kress J.W, Erickson D.L (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proc NatlAcad Sci U S A*, 105(8):2761-2762.

8. Kress WJ, Erickson DL (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *PLoS ONE* 2(6): e508.

9. Luong Thinh Nghiep (2000). Chinese name of *Camella sp. Kim Thuan publisher, Bac Kinh - Trung Quoc*.

10. Ngo Thi Minh Duyen, Ngo Quang Hung, Le Sy Doanh, Ngo Quy Cong, Nguyen Van Khuong (2011). Evaluation of growth and generation ability of *Camella sp. f*rom Northern provines. *Forestry Science and technology journal*, 4: 1954 - 1965.

11. Ninh T. (2007). Diversity of wild *Camellia species* of Tam Dao Nationnal Park. *Journal of Science - Natural Science and Technology*, 23:152-154.

12. Saghai - Maroof M.A, Soliman K.M, Jorgensen R.A, Allard R.W (1984). Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci*, 81: 8014-8019.

13. Storchova H, Olson M.S (2007). The architecture of the chloroplast *psbA-trnH* non coding region in angio sperms. Plant systematic and evolution. *Biomedical and life sciences*, 268: 235 – 256.

14. Tran Ninh, Hakoda Naotoshi (2010). *Camellia* species in Tam Dao national park. *Project for management of Tam Dao national park and buffer zone*, pp. 30.

15. Yakawa M, Tsudzuki T, Sugiura M (2006). The chloroplast genome of Nicotianas sylvestris and Nicotiantomentosisformis: complete sequencing confirms that the Nicotiana sylvestris progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol genet genomics*, 275: 367 - 373.

16. *https://blast.ncbi.nlm.nih.gov/Blast.cgi.*

THỬ NGHIỆM BA VÙNG DNA TIỀM NĂNG (*matK, rbcL* và *ITS2*) CHO NHẬN DẠNG LOÀI TRÀ HOA VÀNG TẠI BẮC GIANG *(Camellia euphlebia)* VÀ QUẢNG NINH *(Camellia chrysantha)*

Nguyễn Văn Việt[1], Phạm Quang Chung[1], Đỗ Quang Trung[1], Trần Việt Hà[1], Sounthone Douangmala[2]

[1]Trường Đại học Lâm nghiệp

[2]Trường Cao đẳng Nông - Lâm Bolykhămxay, Viêng Chăn, Lào

TÓM TẮT

Trà hoa vàng (*Camellia sp.*) là loài cây đa tác dụng có giá trị kinh tế cao, dùng làm đồ uống bổ dưỡng, dược liệu và trang trí cảnh quan. Bài báo này công bố kết quả so sánh trình tự nucleotide của ba vùng DNA (*matK*, *ITS2*, *rbcL*) giữa loài Trà hoa vàng ở Ba Chẽ, Quảng Ninh (*Camellia chrysantha*) và loài Trà hoa vàng ở Sơn Động, Bắc Giang (*Camellia euphlebia*). Tỷ lệ thành công cho khuyếch đại PCR cho ba đoạn mã vạch là 100%. Tỷ lệ đọc thành công trình tự hai chiều đạt được từ sản phẩm PCR là 100% cho ba đoạn mã vạch DNA. Độ dài trình tự nucleotide phân tích thuộc vùng DNA lần lượt là 921bp, 589 bp và 367 bp cho *mat*K, *rbc*L và *ITS2*. Kết quả phân tích ba vùng DNA lựa chọn đều cho thấy sự sai khác nucleotide giữa hai loài Trà hoa vàng trong nghiên cứu dao động trong khoảng từ 0% đến 0,82% thuộc sự khác biệt giữa các thể trong loài, có nghĩa là Trà hoa vàng ở Ba Chẽ, Quảng Ninh (*Camellia chrysantha*) là một dẫn xuất của loài Trà hoa vàng ở Sơn Động, Bắc Giang (*Camellia euphlebia*).

Từ khóa: *Camellia chrysantha, Camellia euphlebia*, DNA mã vạch, giám định loài, Trà hoa vàng.

Received : 01/11/2018
Revised : 08/5/2019
Accepted : 15/5/2019