

## MOLECULAR PHYLOGENY OF *Camellia sinensis* var. *assamica* FROM VIETNAM

Le Chi Toan

*Faculty of Biology, Hanoi Pedagogical University 2*

**Abstract.** The tea plant *Camellia sinensis* is an economically important species worldwide, especially in Asia. *C. sinensis* var. *assamica* is one of the four varieties of *C. sinensis*, and the tea from *C. sinensis* var. *assamica* has a high economic value. Several phylogenetic studies on *Camellia* were conducted, however, the phylogenetic studies to investigate the relationship of *C. sinensis* var. *assamica* from Vietnam are limited. Identifying the phylogenetic position and genetic relationship of *C. sinensis* var. *assamica* from Vietnam is significant in providing essential information for taxonomic study and economic resources. The present study, based on the molecular data of four DNA regions supported the monophyly of *C. sinensis*. The *Camellia sinensis* var. *assamica* from Vietnam formed a clade together, but it is far from *Camellia sinensis* var. *assamica* from India and China. The results support the evolution in the genetics of *C. sinensis* var. *assamica* between regions, however, the current genetic variations do not influence classification yet. *C. sinensis* has been grown in several regions in Asia, however, the conservation of native populations of *Camellia sinensis* var. *assamica* is necessary and should be carried out for sustainable development.

**Keywords:** *Camellia*, tea, phylogeny, molecular, morphology.

### 1. Introduction

*Camellia* L. (Theaceae) including about 200 species is evergreen shrubs or trees that are native to East, South, and Southeast Asia. The species of *Camellia* are economically important ornamental flowers, edible oil sources, and tea plants [1]. Today, several *Camellia* species are cultivated and produced all over the world. According to the Food and Agriculture Organization of the United Nations statistics ([www.fao.org/faostat/](http://www.fao.org/faostat/)), tea production has expanded by 66% in terms of area, reaching 5.3 million tons on 3.5 million hectares across 50 tea-growing nations in the last decade.

*Camellia sinensis*, tea, was first recorded and used by the Chinese more than two millennia ago, and it has become one of the most popular beverage sources in the world now [2].

---

Received April 11, 2023. Revised June 24, 2023. Accepted June 30, 2023.

Contact Le Chi Toan, e-mail address: [lechitoan@hpu2.edu.vn](mailto:lechitoan@hpu2.edu.vn)

*Camellia sinensis* is also cultivated in several countries in Asia including Vietnam. The species consists of four varieties *C. sinensis* var. *sinensis*, *C. sinensis* var. *pubilimba*, *C. sinensis* var. *assamica*, *C. sinensis* var. *dehungensis*. Among the four varieties, the tea from *C. sinensis* var. *assamica* has the highest economic value. *C. sinensis* var. *assamica* is an evergreen shrub or tree that adapted to altitudes from 1500 m (Figure 1).



**Figure 1. Morphology of *Camellia sinensis* var. *assamica* from Vietnam. A: Branch with flower, B: Flower, C: Branch with bud and young leaf in dry**

Several phylogenetic studies on *Camellia* were conducted [3-6] based on molecular data. Vijayan et al. [3] investigated paralogous ITS (internal transcribed spacer region of 18S–26S nuclear ribosomal DNA) fragments of 112 species of *Camellia*. Eight major clades were determined within *Camellia*. The study only supported the monophyly of sects. *Thea* and *Furfuracea*. However, the sample of *Camellia sinensis* is limited to only two individuals. Furthermore, the study used the synonym name *Camellia assamica* instead of the accepted name *C. sinensis* var. *assamica*. Thus, the question of the relationship between varieties of *C. sinensis* remains.

The phylogenetics and biogeography of Theaceae were performed by Li et al. [4], the study used multiple plastid genes from 27 species of Theaceae representing all genera except *Laplacea*. The results clarified the relationship of Theaceae with three

tribes in the family. The evolution of Theaceae also was discussed. Unfortunately, the limitation of the *Camellia* sample was not a support to clarify the relationship within *Camellia*. Similarly, the phylogenomic study by Cheng et al. [5] did not focus on the relationship of *C. sinensis*.

Zhao et al. [6] reconstructed a robust phylogenetic tree of *Camellia*. The results of the study supporting the monophyly of *Camellia* with ten clades were recognized with the genus, and some taxonomic treatments provided were based on the phylogenetic results. Additionally, the historical biogeography of *Camellia* was investigated, and the crown age of *Camellia* was estimated at 39.5 Ma. However, the study did not discuss about *Camellia sinensis* and its varieties.

*Camellia sinensis* var. *assamica* recorded is distributed in SE Asia and some areas of Vietnam such as Ha Giang, Yen Bai, Nghe An, and Son La. Identifying the phylogenetic position and genetic relationship of *Camellia sinensis* var. *assamica* from Vietnam is significant, as it can provide essential information for taxonomic study and economic resources. Thus, the present study aims to (1) reconstruct the phylogeny of the genus *Camellia sinensis* and its allies based on molecular data; and (2) clarify the phylogenetic relationship of *Camellia sinensis* var. *assamica* from Vietnam.

## 2. Content

### 2.1. Materials and methods

#### 2.1.1. Taxon sampling, DNA extraction, amplification, sequencing

We collected the five *Camellia sinensis* var. *assamica* individuals from Ha Giang and Son La provinces, Vietnam. Three of the four varieties of *Camellia sinensis* were sampled. Additionally, we assembled molecular data from NCBI of 16 individuals of *Camellia sinensis* and its allies to reconstruct the phylogenetic trees of *Camellia sinensis* var. *assamica* based on the data including a total of 28 samples (Table 1). Four molecular makers were used in this study including ITS, *rbcL*, *matK*, and *trnL-F* (Table 1). Three species of *Polyspora* (*P. hainanensis*, *P. longicarpa*, *P. specioca*) were selected as outgroups.

**Table 1. Source information and GenBank accession numbers for DNA sequences generated or used in this study. “–” indicates missing data, and “XXX” represents sequences newly generated in this study**

Species	Source	<i>rbcL</i>	<i>trnL-trnF</i>	<i>matK</i>	ITS
<i>Camellia fangchengensis</i>	China	MW119914	MW120242	MW119750	–
<i>Camellia gymnogyne</i>	China	MW119925	MW120253	MW119761	MH270492
<i>Camellia kwangsiensis</i>	China	MW119937	MW120265	MW119773	FJ432106
<i>Camellia</i>	China	AF380037	AY943277	AF380077	AF315492

*Molecular phylogeny of Camellia sinensis var. assamica from Vietnam*

<i>sinensis</i>					
<i>Camellia sinensis</i>	India	MW248322	MK890365	MW259057	MW251737
<i>Camellia sinensis</i>	India	MW248317	MK890364	MW259052	–
<i>Camellia sinensis</i>	China	MW120015	MW120343	MW119851	–
<i>Camellia sinensis</i>	India	MW248321	MK890363	MW259056	MW251736
<i>Camellia sinensis</i> var. <i>assamica</i>	Vietnam	OR198078	OR148368	OR148363	OR139847
<i>Camellia sinensis</i> var. <i>assamica</i>	Vietnam	OR198079	OR148369	OR148364	OR139848
<i>Camellia sinensis</i> var. <i>assamica</i>	Vietnam	OR198080	OR148370	OR148365	OR139849
<i>Camellia sinensis</i> var. <i>assamica</i>	Vietnam	OR198081	OR148371	OR148366	OR139850
<i>Camellia sinensis</i> var. <i>assamica</i>	Vietnam	OR198082	OR148372	OR148367	OR139851
<i>Camellia sinensis</i> var. <i>assamica</i>	India	MK262770	–	MN138452	MW252037
<i>Camellia sinensis</i> var. <i>assamica</i>	China	MW120014	MW120342	MW119850	–
<i>Camellia sinensis</i> var. <i>assamica</i>	India	MK262744	–	MW248378	MW252033
<i>Camellia sinensis</i> var. <i>assamica</i>	India	MK262745	–	MW248379	MW252034
<i>Camellia sinensis</i> var. <i>assamica</i>	India	MK262748	–	MN138434	OL588265
<i>Camellia</i>	India	MK262746	–	MN138435	OL588266

<i>sinensis</i> var. <i>assamica</i>					
<i>Camellia</i> <i>sinensis</i> var. <i>dehungensis</i>	China	KJ806279	–	KJ806279	–
<i>Camellia</i> <i>sinensis</i> var. <i>sinensis</i>	India	MK262761	–	MN473193	MW252044
<i>Camellia</i> <i>sinensis</i> var. <i>sinensis</i>	India	MK262757	–	MN138443	MW252043
<i>Camellia</i> <i>sinensis</i> var. <i>sinensis</i>	India	MN494893	–	MN473198	MW252046
<i>Camellia</i> <i>sinensis</i> var. <i>sinensis</i>	India	MN494892	–	MN473197	MW252045
<i>Camellia</i> <i>tachangensis</i>	China	MW120022	MW120350	MW119858	–
<i>Polyspora</i> <i>hainanensis</i>	China	AF380049	AY216566	AF380092	AY214932
<i>Polyspora</i> <i>longicarpa</i>	China	MW120041	MW120369	MW119877	AF456264
<i>Polyspora</i> <i>speciosa</i>	China	MW120040	MW120368	MW119876	–

The genomic DNA of *Camellia sinensis* var. *assamica* was extracted from silica gel-dried tissues using the CTAB procedure [7]. Polymerase chain reactions and sequencing were performed using the primers designed by Vijayan et al. [3], Le et al. [8], and Taberlet et al. [9]. The primers used for conducting PCR and sequencing are presented in Table 2.

The PCR amplification reactions used MasterMix of the BioMed company. The PCR program consisted of 5 min at 95 °C, 37 cycles of 30 s at 95 °C, 50 s at 52 °C, and 1 min 30 s at 72 °C, with a final extension of 10 min at 72 °C.

PCR products were purified on 1.0 % agarose gels. The PCR products were purified using BioMed multifunctional DNA fragment purification recovery kits and sequenced using the amplification primers. The bidirectional sequencing was completed using the ABI 3730 DNA Sequencer (Applied Biosystems, Carlsbad, California, USA). The sequences were aligned in Geneious v.8.0.5 [10].

Table 2. Primers used for PCR and sequencing in this study

Locus	Primer	Sequence 5'–3'
<i>matK</i>	F56	CCTATCCATCTGGAAATCTTAG
	R1326	GTTCTAGCACAAGAAAGTCG
<i>rbcL</i>	1F	ATGTCACCACAAACAGARAC
	889R	CTATCAATAACTGCATGCAT
<i>trnL-F</i>	C	CGAAATCGGTAGACGCTACG
	F	ATTTGAACCTGGTGACACGAG
ITS	IT1–F	AGTCGTAACAAGGTTTCC
	IT2–R	GTAAGTTTCTTCTCCTCC

### 2.1.2. Phylogenetic analyses

The phylogenetic analyses of *Camellia sinensis* were carried out by using both the maximum likelihood (ML) and Bayesian inference (BI). The ML analysis was performed using the program RAxML 8.2.10 [11, 12] with the best-fitting model for each molecular marker and the combined dataset at the Cyber Infrastructure for Phylogenetic Research (CIPRES; www.phylo.org). ML bootstrap analysis was implemented with 1000 replicates. Bayesian inference was conducted in MrBayes 3.1.2 [13]. The best-fitting models for each marker and the combined data set were determined by the Akaike Information Criterion (AIC) as implemented in jModelTest 2.1.6 [14]. Bayesian analysis of the combined data set used the GTR + I + G model as determined in jModelTest. The MCMC algorithm was run for 5,000,000 generations with four Markov chain Monte Carlo (MCMC) and trees were sampled every 1000 generations. The effective sample size (ESS) for all pertinent parameters was checked using the tool Tracer 1.6 [15], which indicated that stationarity had likely been attained (>200). With the first 25% of sampled generations (2500 trees) discarded as burn-in, a 50% majority-rule consensus tree and posterior probabilities (PP) were obtained using the remaining trees.

### 2.2. Results and discussion

The study generated 20 new sequences of *Camellia sinensis* var. *assamica*. The lengths of individual data sets of ITS, *rbcL*, *matK*, and *trnL-F* are 805, 1254, 740, and 1003 bps, respectively. The combined dataset included 3802 aligned ingroup and outgroup positions (Figure 2).

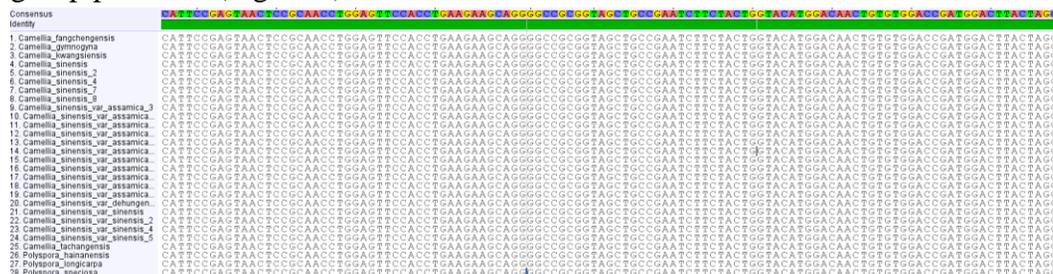
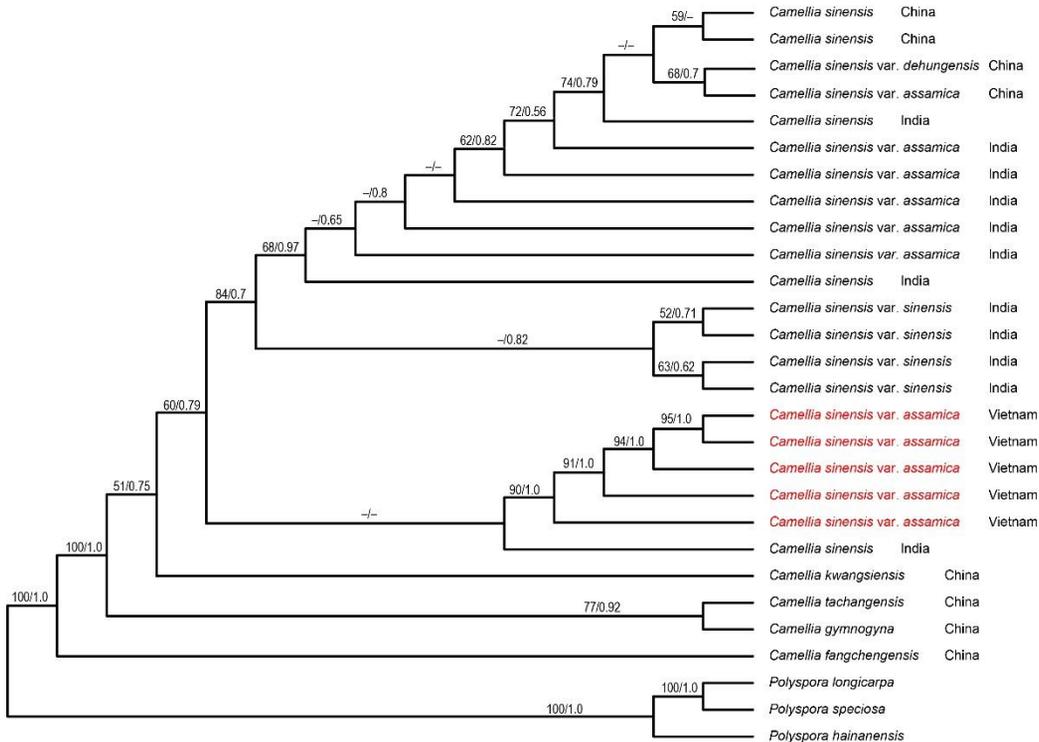


Figure 2. A section of the combined molecular dataset from four DNA regions (ITS, *rbcL*, *matK*, and *trnL-F*) of *Camellia*

The results from ML and BI trees were highly congruent, the few differences had low solutions. Thus, we combined the results in the ML tree with BS and PP values. The phylogenetic relationship within *Camellia* is presented in Figure 3.

Our molecular results well supported the monophyly of *Camellia*, the major clades in the phylogenetic tree are well solution. *C. fangchengensis* was supported as a sister to the remaining members. While *C. gymnogyna* and *C. tachangensis* formed a clade. The three varieties of *Camellia sinensis* were formed together in a clade (Figure 3). The *Camellia sinensis* var. *assamica* from Vietnam is placed together and closely relative to *C. sinensis* from India (Figure 3). Whereas, *Camellia sinensis* var. *assamica* from India and China were placed together but with low support. The position of *Camellia sinensis* var. *dehungensis* likely unstable, but it seems closely related to taxa from China. In addition, *Camellia sinensis* var. *sinensis* weakly formed a clade.

The individuals of *Camellia sinensis* var. *assamica* from Ha Giang and Son La provinces, Vietnam placed together suggesting the genetic conservatism of the *Camellia sinensis* var. *assamica* event they have a large distribution distance (Figures 2, 3). However, *Camellia sinensis* var. *assamica* from Vietnam placed far from individuals from India and China. These results support the evolution in the genetics of *C. sinensis* var. *assamica* between regions, however, the current genetic variations have not influenced classification yet. *Camellia sinensis* var. *dehungensis* is an endemic taxon in Yunnan, China, the molecular data did not resolve the position of this lineage. However, it seems closely related to its allies from China.



**Figure 3. Maximum likelihood tree showing the phylogenetic relationship of *Camellia*. Nodal support is given above the branches as ML bootstrap values/Bayesian posterior probabilities**

The phylogenetic relationship within *C. sinensis* is quite clear with the placement of the three varieties. In morphology, *Camellia sinensis* var. *assamica* was characterized by the leaf blade abaxially densely spreading villous, especially in the bud and young leaf (Figure 1) [16, 17]. This variety is distributed in evergreen broad-leaved forests at altitudes of 1500-1900 m where the remaining three varieties of *C. sinensis* rarely occur [16]. In general, the varieties of *Camellia sinensis* are highly conserved in both morphology and genetics.

All four varieties of *C. sinensis* were used as a tea for thousands of years in Asia. The tea from *C. sinensis* var. *sinensis* is likely more popular, while *C. sinensis* var. *assamica* is exploited and processed into tea products with higher economic value. Although *C. sinensis* has been grown in several regions in Asia, however, the conservation of native populations of *Camellia sinensis* var. *assamica* is necessary and should be carried out for sustainable development.

### 3. Conclusions

The study based on the molecular data of comprehensive taxon sampling supported the relationship of *Camellia sinensis*. The genetic relationship within *C. sinensis* is quite clear with the placement of the three varieties. The individuals of *C. sinensis* var. *assamica* from Vietnam are placed together which suggests the genetic conservatism of the *Camellia sinensis* var. *assamica* even though they have a large distribution distance. However, *Camellia sinensis* var. *assamica* from Vietnam placed far from individuals from India and China. These results support the evolution in the genetics of *C. sinensis* var. *assamica* between regions, however, the current genetic variations have not influenced classification yet. The conservation of native populations of *Camellia sinensis* var. *assamica* is necessary even though it has been grown in several regions in Asia.

### REFERENCES

- [1] T. L. Ming, 2000. *Monograph of the genus Camellia*. Kunming: Yunnan Science and Technology Press.
- [2] D. W. Zhao, J. A. N. Parnell, T. R. Hodgkinson, 2017. Names of Assam tea: Their priority, typification, and nomenclatural notes. *Taxon*, Vol. 66, pp. 1447-1455.
- [3] K. Vijayan, W. J. Zhang, C. H. Tsou, 2009. Molecular taxonomy of *Camellia* (Theaceae) inferred from nrITS sequences. *Am. J. Bot.*, Vol. 96, pp. 1348-1360.
- [4] L. Li, Y. Hu, M. He, B. Zhang, W. Wu, P. Cai, D. Hou, Y. Hong, 2021. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics*, Vol. 22, pp. 138.
- [5] L. Cheng, M. Li, Q. Han, Z. Qiao, Y. Hao, T. S. Balbuena, Y. Zhao, 2022. Phylogenomics resolves the phylogeny of Theaceae by using low-copy and multi-copy nuclear gene makers and uncovers a fast radiation event contributing to tea plants' diversity. *Biology*, Vol. 11, pp. 1007.

- [6] D. Zhao, T. R. Hodkinson, J. A. N. Parnell, 2023. Phylogenetics based on three nuclear regions and its implications for systematics and evolutionary history of global *Camellia* (Theaceae). *J. Syst. Evol.*, Vol. 61, pp. 253-439.
- [7] J. J. Doyle, J. L. Doyle, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*, Vol. 19, pp. 11-15.
- [8] C. T. Le, B. Liu, R. L. Barrett, L. M. Lu, J. Wen, Z. D. Chen, 2018. Phylogeny and a new tribal classification of Opiliaceae (Santalales) based on molecular and morphological evidence. *J. Syst. Evol.*, Vol. 56, pp. 56-66.
- [9] P. Taberlet, L. Gielly, G. Pautou, 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.*, Vol. 17, pp. 1105-1109.
- [10] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, A. Drummond, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, Vol. 28, pp. 1647-1649.
- [11] A. Stamatakis, 2006. RAxML-VI-HPC, maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, Vol. 22, pp. 2688-2690.
- [12] Stamatakis, P. Hoover, J. Rougemont, 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Syst. Biol.*, Vol. 57, pp. 758-771.
- [13] F. Ronquist, J. P. Huelsenbeck, 2003. MrBayes 3, Bayesian phylogenetic inference under mixed models. *Bioinformatics*, Vol. 19, pp. 1572-1574.
- [14] D. Darriba, G. L. Taboada, R. Doallo, D. Posada, 2012. jModel Test 2: more models, new heuristics, and parallel computing. *Nature Methods*, Vol. 9, pp. 772.
- [15] A. Rambaut, A. J. Drummond, 2007. Tracer, Version 1.4. <http://beast.bio.ed.ac.uk/Tracer>.
- [16] T. L. Ming, B. Bruce, 2007. Theaceae. In: Z. Y. Wu, P.H. Raven, D.Y. Hong (eds.). *Flora of China*. Beijing. *Science Press & St. Louis: Missouri Botanical Garden Press*, Vol. 12 pp. 366-478.
- [17] H. H. Pham, 2003. *An illustrated flora Vietnam*. Ho Chi Minh Young Publishing House, Part 1, p. 430-431.