Towards the marker development applicable for the reconsideration of genetic relationships in the genus *Camellia*

Kei Fukuyama[1], Hironori Katayama[2], Thi Lieu Nguyen[1], Van Dung Luong[3], Nguyet Hai Ninh Le[4], Tana Wuyun[5], Huimin Liu[5], Naotoshi Hakoda[6], Chiyomi Uematsu[1]
1. Botanical Gardens, Graduate School of Science, Osaka City University
2. Graduate School of Agriculture, Kobe University
3. Da Lat University
4. Hoa Lu University
5. Paulownia Research and Development Center of China
6. Emeritus Professor of Keisen University
Contact e-mail address for CU: uematsu@sci.osaka-cu.ac.jp

Keywords: morphology, *mat*K sequence, chloroplast DNA, structural alteration

Abstract
In the genus *Camellia*, it is known to exist a discrepancy between traditional taxonomic system and recently reported molecular phylogeny. For the reconsideration of genetic relationships in the genus *Camellia*, three different approaches were examined in this study. In total 265 individuals of 161 species from 19 sections of the genus *Camellia* were used. These materials were obtained from *Camellia* gardens in Japan and Vietnam, also collected from wild habitat in Vietnam. Accessions of China were prepared and analyzed in China. *Schima liukiuensis* was used as an out-group. Morphological measurement was done for 37 characters of flower. Data obtained were subjected to the principal component analysis. Accessions were divided into two types based on the flower characters, one was perule type and the other was bract type. This agreed with traditional classification system. Some traits have been considered to be important, e.g., number of locule/ovary, did not show high contribution rate. Phylogenetic tree obtained from *mat*K gene sequence didn't form a clade corresponding to each subgenus. By using 6 structural alterations in cpDNA, 161 species were divided into 10 haplotypes in 3 groups by two large and 4 small in/dels. Even each group didn't support traditional classification system, structural alterations of cpDNA may contribute to reconsider relationships in the genus *Camellia*. Further investigation will be required.

Introduction
In the genus *Camellia*, about 250 species are distributing from East to Southeast Asia (Kirino, 2013). Chang and Bartholomew (1984) classified into 4 subgenera, 20 sections mainly based on their flower characters. However recent phylogenetic analysis using molecular markers resulted in a disagreement with conventional classification system (Xiao and Parks, 2003; Vijayan et al., 2009). Recently about 50 species were reported from Vietnam, even after that new species were found from Vinh Phuc province and Lam Dong province in Vietnam (Dung

et al., 2016). These newly reported species were classified based on their morphological characteristics, but not subjected to molecular phylogenetic analysis.

In this study we try to find out suitable methods to reconsider relationships in the genus *Camellia*. In order to clarify which morphological trait is a key character to distinguish species, detailed morphological observation and numerical evaluation has been carried out for flower. For molecular phylogenetic analysis, higher ploidy level made it difficult to use nuclear DNA as markers. Therefore in this study, chloroplast genome was examined whether it is applicable for phylogenetic analysis in the genus *Camellia* or not. One is *mat*K gene sequence, which is widely used for plant taxonomic study because of its high evolutionary rate. Another is structural alteration found in the chloroplast (cp) DNA among different species. Insertion or deletion bigger than some dozens or hundreds bp were considered useful to trace the evolutionary lineage (Ingvarsson et al., 2003). In order to examine applicability of *mat*K gene sequence and cpDNA structural alterations, 265 individuals of 161 species inclusive 59 species from Vietnam were analyzed in this study. These could cover 19 sections out of 20 sections presented by Chang and Bartholomew (1984).

Materials and Methods

Plant materials

In total 265 individuals of 161 species, which cover 19 sections out of 20 sections by Chang and Bartholomew (1984) were used in this study. Table 1 showed number of species and individuals in each section, according to four subgenera. Number in parentheses indicates the number of species distributing in Vietnam.

In Japan, following materials were collected, i.e., 47 individuals of 44 species from Kurume Camellia Garden, 37 individuals of 34 species from World Camellia Hall in Kurume both in Fukuoka prefecture, 35 individuals of 34 species from Inokuchi Tsubaki kan in Toyama prefecture, and *Camellia japonica* individuals from Oshima Island in Tokyo and Ofunato in Iwate Prefecture. 125 individuals of 59 species were collected from Vietnam, 19 individuals of 19 species were collected and analyzed in China.

Flower morphology

Three flowers were collected and observed from one tree. If it was difficult to collect three flowers, one or two flowers were collected depend on situation. Finally flowers could be collected from 96 individuals of 84 species belong to 15 sections were subjected to the morphological analysis. On site, flower position was recorded whether terminal or axillary, longest and shortest diameter was measured by digital caliper. Following characteristics such as, size, number, colour and condition of perule, bracteole, sepal, petal, stamen, pistil and ovary were observed and measured by using dissection microscope. These characteristics were described in Hakoda and Akihama (1988), in this study flower position and number of stigma were newly added. In total 37 characters were subjected to the Principal Component Analysis after standardizing the measured value (Table 2).

DNA extraction for molecular analysis

Leaves collected from each individual were dried by silica gel and stored unde 4^{o}C until use. Twenty mg of dried leaf was grinded using Mixer mill MM300 (Retsch, Germany) under the chilled condition with LN_2. Well grinded fine powder was used for DNA extraction.

Table 1. Number of accessions used in this study.

Subgenus	*Section*	Number of species		Number of individuals
Protocamellia	*Archecamellia*	6	(4)	11
	Stereocarpus	4	(3)	10
	Piquetia	5	(5)	14
Camellia	*Oleifera*	4	(1)	6
	Furfuracea	5	(2)	8
	Paracamellia	10	(2)	12
	Pseudocamellia	1	(0)	1
	Tuberculata	3	(0)	3
	Luteoflora	1	(1)	1
	Camellia	28	(1)	33
Thea	*Corallina*	3	(3)	5
	Brachyandra	1	(1)	3
	Longipedicellata	2	(2)	6
	Chrysantha	45	(27)	96
	Calpandria	2	(1)	2
	Thea	2	(2)	4
	Longissima	1	(1)	1
	Glaberrima	—		—
Metacamellia	*Theopsis*	15	(3)	16
	Eriandria	6	(1)	8
Unclassified		17	(16)	25
Total		161	76	265

Number in parentheses indicates the number of species distributed in Vietnam.

DNeasey Plant Mini Kit (Qiagen, Germany) was used following the manufacturer's protocol with a slight modification of final elution solution changed to 50 µl of TE10:0.1. Concentration of obtained DNA solution was measured using Qubit dsDNA HS Assay Kits (Thermo Fisher Scientific, U.S.A.) and the purity was confirmed by 1.5 % agarose gel electrophoresis using 1x TAE buffer.

Phylogenetic analysis using *mat*K sequence

For the phylogenetic study, chloroplast *mat*K gene sequence was widely used. In this study, 711 bp region of *mat*K gene known to be variable among 13 *Camellia* species (Huang *et al.*, 2014) was selected for sequencing. Two sets of primers, i.e., *mat*K4-F: CTGTTGATA AGGTTTACCGAGG and *mat*K4-R:GGTTCCAGAAGATATTGATCG, *mat*K5-F: ATACCCCACTCCA TCCATCTG and *mat*K5-R:ACCCCTGCGAAGTAGAAGAAG, were constructed by using Primer3 web version 4.1.0 (Whitehead Institute for Biomedical Research, U.S.A.) based on the chloroplast genome sequence of 6 *Camellia* species reported by Yang et al. (2013).

PCR amplification was performed for these primer pairs respectively using KOD polymerase (TOYOBO, Japan). Amplification was confirmed by 1.5 % agarose gel electrophoresis and fragment size was estimated.

Table 2. Flower characteristics observed and used for Principal Component Analysis.

	Characteristic	PC1	PC2
1	Petal color	-0.19	-0.03
2	Flower type	0.02	-0.21
3	Flower position	0.13	0.05
4	Flower diameter	-0.20	-0.12
5	Pedicel length	0.06	-0.23
6	Pedicel width	-0.10	-0.30
7	Pedicel hair	-0.14	-0.02
8	Number of Perule	-0.26	0.01
9	Max Perule length	-0.26	-0.03
10	Max perule width	-0.26	-0.04
11	Perule hair	-0.26	0.07
12	Number of bracteoles	0.24	-0.10
13	Max bracteole length	0.22	-0.16
14	Max bracteole width	0.24	-0.16
15	Bracteole hair	0.23	-0.05
16	Number of Sepal	0.25	-0.10
17	Max sepal length	0.22	-0.18
18	Max sepal width	0.22	-0.19
19	Sepal hair	0.23	-0.05
20	Number of Petal	0.07	-0.11
21	Max petal length	-0.20	-0.21
22	Max petal width	-0.19	-0.21
23	Petal hair	-0.12	0.00
24	Filament length	-0.13	-0.28
25	Number of filaments	0.01	-0.28
26	Filament hair	-0.02	-0.11
27	Number of stigmata	-0.07	-0.11
28	Style length	-0.08	-0.30
29	Style hair	-0.08	-0.11
30	Ovary diameter	-0.11	-0.27
31	Number of locules / ovary	-0.06	-0.12
32	Number of ovules / locules	-0.04	-0.17
33	Ovary hair	-0.13	0.08
34	A rate of connation between petals	-0.01	-0.21
35	A rate of connation between petal and filament	0.01	-0.26
36	A rate of connotation between filaments	-0.03	-0.15
37	A rate of connation between style	-0.09	0.08

Principal component loading for each characteristic is given in the column of PC1 and PC2.

When amplification was confirmed, that PCR product was treated with ExoSAP-IT (Thermo Fisher Scientific, U.S.A.) in order to remove remaining primers and nucleotides. Resulted solution was used as a template for sequencing. Big Dye Terminater v3.1 and x5 sequencing buffer (Thermo Fisher Scientific, U.S.A.) were used following sequencing reaction. After repeating 25 cycles of 96^{o}C for 10sec, 50^{o}C for 5 sec, and 60^{o}C for 4 min, resulted solution

was precipitated with Ethanol, 3.0M NaOAc (pH 9.0) and 0.5 M EDTA (pH 8.0). Precipitate was washed with 70% Ethanol and dried, then Hi-Di Formamid (Thermo Fisher Scientific, U.S.A.) was added. DNA sequence was analyzed using 3500/3500xL Genetic Analyzer (Thermo Fisher Scientific, U.S.A.). Based on the DNA sequence data set obtained from 94 species for 711 bp region of *mat*K gene, phylogenetic tree was constructed by NJ method by using ClustalX2.1 (Thompson *et al*., 1997).

Detect the structural alterations among *Camellia* cpDNAs

Entire chloroplast DNA sequence of seven individuals of *Camellia* species revealed by Yang *et al.* (2013) were obtained from Genbank (National Center for Biotechnology Information, U.S.A.). Multiple alignment was done for seven sequence data by using GeneDoc Version 2.7.000 (Pittsburgh Supercomputing Center, U.S.A.). Then bigger than 30 bp of insertion or deletion found at least one species was searched for entire sequence. Six regions were found and corresponding 6 primer pairs were designed by using Primer3 web version 4.1.0 (Whitehead Institute for Biomedical Research, U.S.A.) (Table 3). Each primer pair could amplify the region including insertion or deletion. DNAs of 265 accessions were subjected to PCR amplification by using 6 primer pairs. PCR products were applied for agarose gel electrophoresis in order to confirm the amplification and estimate the fragment size. 1.5 % or 3 % agarose gel was used depend on the expected fragment size, bigger or smaller than 400 bp respectively.

Results and Discussion

Flower morphology

Principal Component Analysis (PCA) was performed using 37 flower characteristics. The principal component loading of the first principal component (PC1) and PC2 for 37 characteristics were shown in Table 2. The value was bigger in PC1 for the characteristics related to flower size, perule, bracteole and sepal, and bigger in PC2 for the characteristics such as pedicel size, petal size, No. of filament and length, style length, ovary diameter and rate of connation between petal.

Eigen value and contribution of PC1 were 3.58 and 34.5 % respectively, and those of PC2 were 2.68 and 19.4 % respectively. Cumulative contribution of PC1 and PC2 was 54.0 %. Scatter diagram based on the principal component scores of PC1 and PC2 was shown in Figure 1. Accessions were well grouped according to the subgenera except for accessions belong to *Protocamellia*. This result reflected the traditional taxonomic system mainly based on the flower morphology.

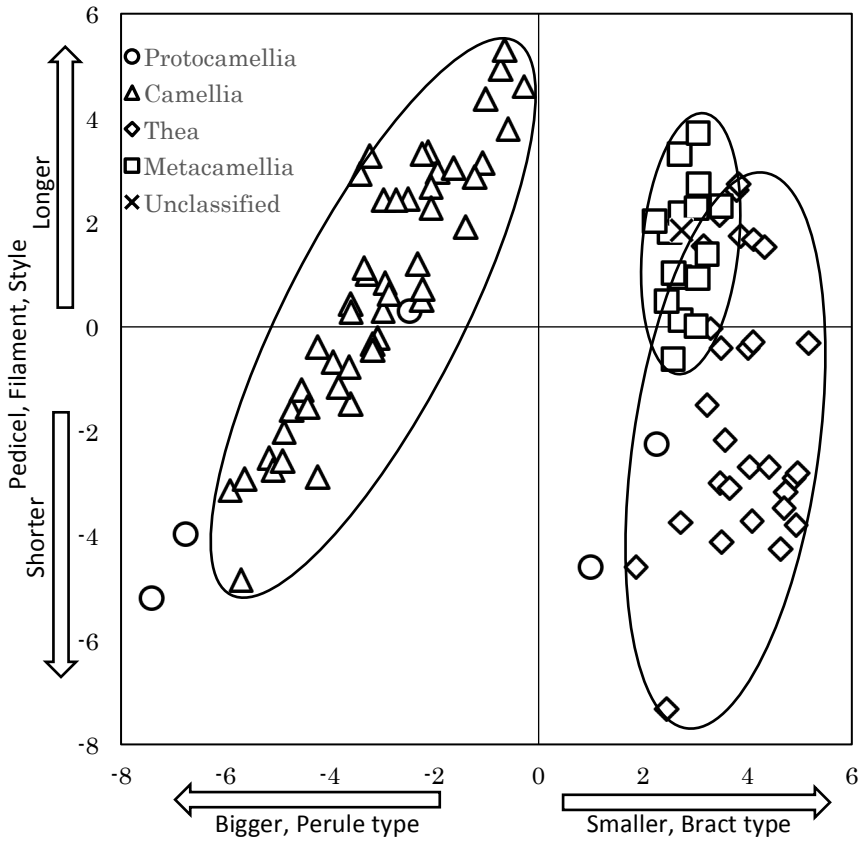

Fig. 1. Scatter diagram based on the principal component scores of PC1 and PC2.
Horizontal axis represents the score of PC1, and vertical-axis represents the score ofPC2. Symbols indicate each subgenus. In horizontal-axis, positive value indicates bract type and smaller flower, negative value indicates perule type and bigger flower. In vertical-axis, positive value indicates longer pedicel, filament and style, negative value indicates shorter pedicel, filament and style.

Phylogenetic relationships revealed by *mat*K sequence

Phylogenetic tree for 94 species of the genus *Camellia* constructed by NJ method was shown in Figure 2. Species were highlighted by 4 different colour according to 4 subgenera. Basically the resolution was very low and species belong to four different subgenera mixed together without forming a clade corresponding to each subgenus. In some occasion, small clade was formed based on a single nucleotide substitution. Also 6 bp deletion resulted in a small clade composed of 9 species belong to 3 subgenera. These results indicated the close relationships among species in the genus *Camellia* irrespective of 4 subgenera.

Chloroplast haplotypes characterized by the structural alterations in cpDNA

Ten chloroplast haplotypes were identified among 266 individuals (inclusive one out group) based on the in/del at 6 regions of cpDNA (Table 3, 4). At the region 1 between *atp*H and *atp* I genes, 400bp of fragment size difference was detected. By this difference, accessions were divided into two groups, Group I harbouring shorter fragment (deletion type) and others harbouring longer one (insertion type).

25. K. Fukuyama, *et al* : Toward marker development for reconsideration of genetic relationships in the genus *Camellia*

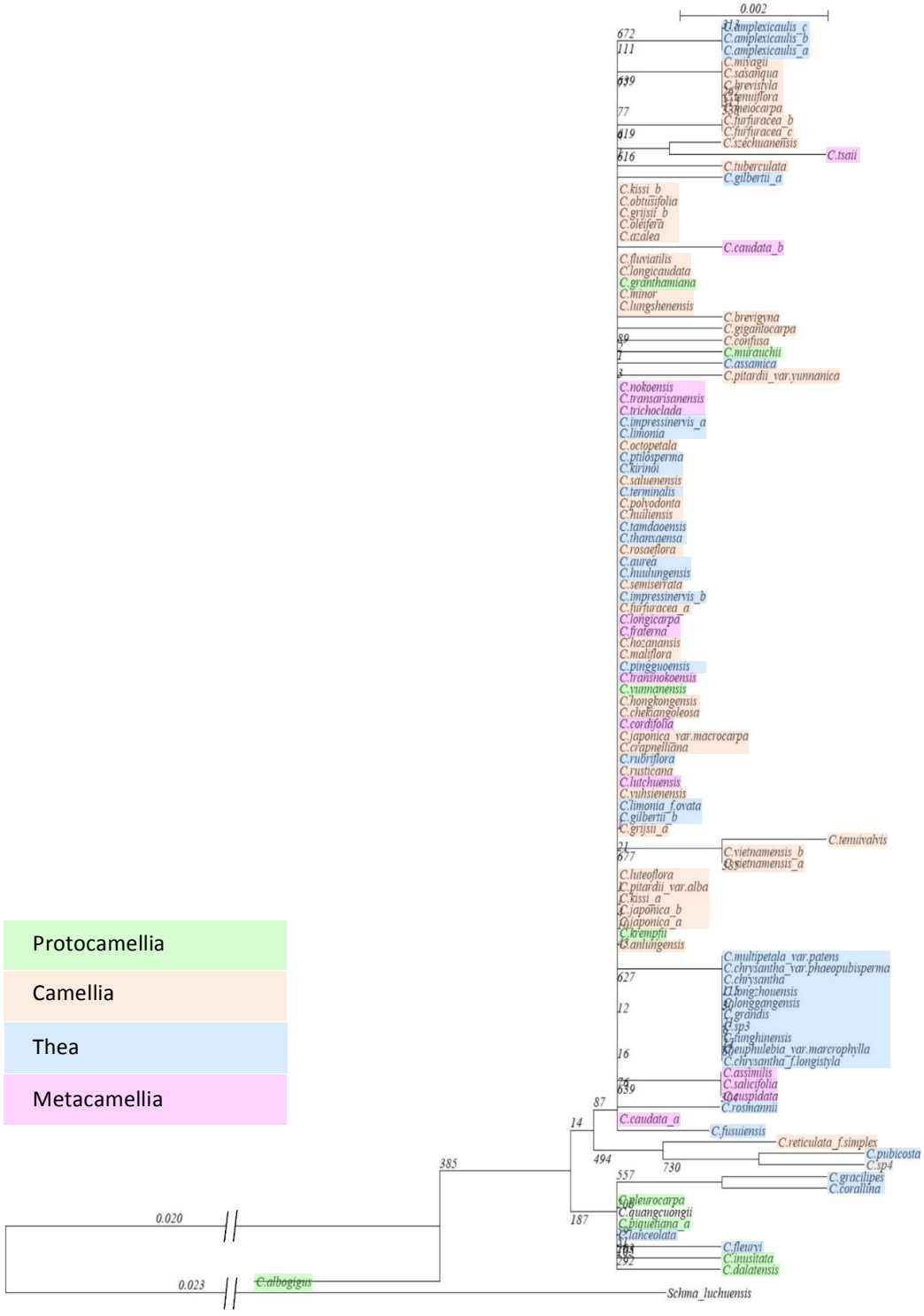

Fig. 2 Phylogenetic tree composed of *mat* K sequence of 94 *Camellia* species by NJ methods. Numbers on the branch represents bootstrap value. Color represents each subgenus.

Then latter accessions were divided into two groups again based on the 433 bp difference in the region 2 between *psb* D and *psb* M genes, Group II harbouring shorter fragment (deletion type) and Group III harbouring longer one (insertion type). Mutation occurred at the region 1 and 2 could be considered to be a bigger event. Four more mutations of 30 to

40 bp of fragment size difference were detected and named region 3, 4, 5, and 6. Finally three major groups I, II and III were divided into 10 haplotypes, A to J.

There are two major haplotypes, one is haplotype C in the Group I, 86 accessions carrying this type, another is haplotype I in the Group III, 146 accessions carrying this type (Table 4). These two types were found in many sections, i.e., haplotype C was seen in 8 sections of 3 subgenera and haplotype I was seen in all sections analyzed (Table 5). Other haplotypes were minor because less than 10 accessions carried those haplotypes except for haplotype J possessed by 12 accessions.

Table 3. Sequence and features of primer pairs used to detect the structural alterations.

Region	Primer name		Sequence(5'-3')	Expected product length	Expected In/Del length
1	*atp* H-*atp* I	F1	TTCGGCTCTCCCTCTCCTAAC	588 or 188 bp	400 bp
		R2	GGAGGGCCATCATTGACTAG		
2	*psb* D-*psb* M	F2	AGAGAGATGTCCTGAACCACTAG	840 or 407 bp	433 bp
		R2	TGGCGTTACTCTACCACTGA		
3	*ycf* 1-*rps* 7_3	F1	GGCATCCTAACAGACCGGTA	225 or 193 bp	32 bp
		R1	TTCTTGACTGGAGGGGACAC		
4	*ycf* 1-*rps* 7_5	F1	TCATGATCGGGATAGCGGAC	176 or 135 bp	31 bp
		R1	TGGCTGAGGAAGAGTTACGG		
5	*ycf* 1-*rps* 7_6	F1	AGCTACTGAGGAACAACGGG	192 or 232 bp	40 bp
		R1	TGGAATTGGGATATGGATGGAAT		
6	*ycf* 2_1	F1	GGAGCAAGAATTTCCAGGAACA	264 or 234 bp	30 bp
		R1	AGATCTCGGACCTATGAATGGG		

The primer name indicates the region where amplified, between the genes or within the gene.

Table 4. Distribution of structural alterations found in each haplotype.

	Haplotype	No. of accession	Region of structural alteration occurred given in Table1					
			1	2	3	4	5	6
Group I	A	1	−	+	+	−	+	+
	B	1	−	+	+	+	+	−
	C	86	−	+	+	+	+	+
	D	7	−	+	−	+	+	+
Group II	E	1	+	−	+	−	+	+
	F	5	+	−	+	+	+	+
	G	3	+	−	+	+	−	+
Group III	H	3	+	+	+	−	+	+
	I	146	+	+	+	+	+	+
	J	12	+	+	−	+	+	+
	OUT	1	+	−	−	+	+	+

+: longer fragment amplified, -: shorter fragment amplified.

In the section *Camellia, Cuspidata and Theopsis,* 5 haplotypes were observed, and in *Chrysantha,* 6 haplotypes were found. Also 6 other sections were including 2 or 3 haplotypes.

Remaining 7 sections had only one haplotype in this study.
Based on these results, chloroplast genome of the genus *Camellia* seemed to be divergent including 2 big in/dels and smaller ones. But most of taxa such as subgenus or section could not be defined by cpDNA haplotype. This might suggest the disagreement between traditional taxonomic system and chloroplast haplotypes.

Table 5. Distribution of haplotype in each section of four subgenera.

Subgenus	*Section*	Group I				Group III			Group II		
		A	B	C	D	E	F	G	H	I	J
Archecamellia	*Archecamellia*			5					1	5	
	Streocarpus								1	9	
	Piquetia			8						6	
Camellia	*Oleifera*			2						4	
	Furfuracea			2						6	
	Paracamellia			4						8	
	Puseudocamellia									1	
	Tuberculata									3	
	Luteoflora									1	
	Camellia	1		10	1					19	2
Thea	*Corallina*			1						4	
	Brachyandra									3	
	Longipedicellata										6
	Chrysantha		1	47	6		2			39	1
	Calpandria									2	
	Thea									4	
	Longissima									1	
	Glaberrima										
Metacamellia	*Cuspidata*					1	2	2		9	2
	Theopsis						1	1	1	4	1
	Unclassified			7						18	
	Total	1	1	86	7	1	5	3	3	146	12

Conclusion

Based on the flower morphology, PCA analysis revealed 3 subgenera, *Camellia*, *Thea* and *Metacamellia* were grouped respectively. This result reflected the traditional classification system, however *Protocamellia* could not form a group. Phylogenetic tree constructed from *mat*K sequence could not solve the relationships in the genus *Camellia*. This might indicate

the very close relationships among species in the genus *Camellia* caused by recent speciation. On the other hand, chloroplast haplotype analysis based on the structural alterations gave rise 10 haplotypes in spite of conservativeness, it suggest the possibility to use hypervariable regions such as cpDNA SSR in chloroplast genome for the further research.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16K07239 from 2016 to 2018 and MEXT Cooperative Joint Research Grant from 2017 to 2019.

Literature cited

Chang, H. T., and Bartholomew, B. 1984. *Camellias*. B. T. Batsford ltd., London, U.K.

Dung, L. V., Son, H. T., Ninh, T., and Nhan, P. H. 2016. *Camellia quangcuongii* (*Theaceae*), a new species from Vietnam. *J. Jpn. Bot.* 91:226–230.

Kirino, S. 2013. Inokuchi Tubaki Kan Illustrated Guide to wild Camellia. (in Japanese) Inokuchi Tubaki-Kan, Toyama pref., Japan

Hakoda, N., and Akihama, T. 1988. Morphological classification of cultivars in *Camellia sasanqua* THUNB. using principal component analysis and cluster analysis. Journal of the Japanese Society for Hort. Sci. 57:233-242

Huang, H., Shi, C., Liu, Y., Mao, S.-Y., and Gao, L.-Z. 2014. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology 14:151.

Ingvarsson, K.P., Ribstein, S., and Douglas, R.T. (2003) Molecular Evolution of Insertions and Deletion in the Chloroplast Genome of *Silene*. Molecular Biology and Evolution 20(11) :1737–1740.

Ming, T. L. 2000. Monograph of the genus '*Camellia*'. Kunming Institute of Botany, Chinese Academy of Sciences, Yunnan Science and Technology Press, Kunming, China.

Sealy, J. R. 1958. A revision of the genus *Camellia*. Royal Horticultural Society, London, U.K.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25:4876-4882.

Vijayan, K., Zhang, W.-J., and Tsou, C.-H. 2009. Molecular taxonomy of Camellia (Theaceae) inferred from nrITS sequences. American Journal of Botany 96:1348-1360.

Xiao, T. J., and Parks, C. R. 2003. Molecular analysis of the genus *Camellia*. International Camellia Journal 35:57-65.

Yang, J.-B., Yang, S.-X., Li, H.-T., Yang, J., Li, D.-Z. 2013. Comparative Chloroplast Genomes of Camellia Species. PLoS ONE 8(8): e73053.