



# Domestication Origin and Breeding History of the Tea Plant (*Camellia sinensis*) in China and India Based on Nuclear Microsatellites and cpDNA Sequence Data

Muditha K. Meegahakumbura<sup>1,2,3,4</sup>, Moses C. Wambulwa<sup>1,2,3,5</sup>, Miao-Miao Li<sup>1,3</sup>, Kishore K. Thapa<sup>6</sup>, Yong-Shuai Sun<sup>7</sup>, Michael Möller<sup>8</sup>, Jian-Chu Xu<sup>9</sup>, Jun-Bo Yang<sup>2</sup>, Jie Liu<sup>1</sup>, Ben-Ying Liu<sup>10</sup>, De-Zhu Li<sup>1,2,3\*</sup> and Lian-Ming Gao<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Paul Gepts,  
University of California, Davis,  
United States

### Reviewed by:

Xun Xu,  
China National GeneBank (BGII), China  
Kevin S. Burgess,  
Columbus State University,  
United States

### \*Correspondence:

De-Zhu Li  
dzli@mail.kib.ac.cn  
Lian-Ming Gao  
gaolm@mail.kib.ac.cn

### Specialty section:

This article was submitted to  
Plant Genetics and Genomics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 21 February 2017

**Accepted:** 27 December 2017

**Published:** 25 January 2018

### Citation:

Meegahakumbura MK,  
Wambulwa MC, Li M-M, Thapa KK,  
Sun Y-S, Möller M, Xu J-C, Yang J-B,  
Liu J, Liu B-Y, Li D-Z and Gao L-M  
(2018) Domestication Origin and  
Breeding History of the Tea Plant  
(*Camellia sinensis*) in China and India  
Based on Nuclear Microsatellites and  
cpDNA Sequence Data.  
*Front. Plant Sci.* 8:2270.  
doi: 10.3389/fpls.2017.02270

<sup>1</sup> Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, <sup>2</sup> Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, <sup>3</sup> College of Life Science, University of Chinese Academy of Sciences, Kunming, China, <sup>4</sup> Genetics and Plant Breeding Division, Coconut Research Institute of Sri Lanka, Lunuwila, Sri Lanka, <sup>5</sup> Biochemistry Department, South Eastern Kenya University, Kitui, Kenya, <sup>6</sup> Department of Botany, Dinhat College, Dinhat, India, <sup>7</sup> Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China, <sup>8</sup> Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, <sup>9</sup> Centre for Mountain Ecosystem Studies, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, <sup>10</sup> Tea Research Institute of Yunnan Academy of Agricultural Sciences, Menghai, China

Although China and India are the two largest tea-producing countries, the domestication origin and breeding history of the tea plant in these two countries remain unclear. Our previous study suggested that the tea plant includes three distinct lineages (China type tea, Chinese Assam type tea and Indian Assam type tea), which were independently domesticated in China and India, respectively. To determine the origin and historical timeline of tea domestication in these two countries we used a combination of 23 nSSRs (402 samples) and three cpDNA regions (101 samples) to genotype domesticated tea plants and its wild relative. Based on a combination of demographic modeling, NewHybrids and Neighbour joining tree analyses, three independent domestication centers were found. In addition, two origins of Chinese Assam type tea were detected: Southern and Western Yunnan of China. Results from demographic modeling suggested that China type tea and Assam type tea first diverged 22,000 year ago during the last glacial maximum and subsequently split into the Chinese Assam type tea and Indian Assam type tea lineages 2770 year ago, corresponding well with the early record of tea usage in Yunnan, China. Furthermore, we found that the three tea types underwent different breeding histories where hybridization appears to have been the most important approach for tea cultivar breeding and improvements: a high proportion of the hybrid lineages were found to be F<sub>2</sub> and BCs. Collectively, our results underscore the necessity for the conservation of Chinese Assam type tea germplasm and landraces as a valuable resource for future tea breeding.

**Keywords:** *Camellia sinensis*, nSSR, cpDNA sequence, domestication center, breeding history

## INTRODUCTION

The tea plant, *Camellia sinensis* (L.) O. Kuntze, is an important economic tree crop (Mondal et al., 2004; Chen and Chen, 2012) that is currently grown in over 52 countries; China and India are the two largest global tea producers (FAOSTAT, 2015). Its taxonomic classification, however, is still a subject of great debate due, in part, to the plasticity of morphological traits that are used to discern tea taxa (Sealy, 1958; Wight, 1962; Banerjee, 1992; Ming, 2000). According to the classification system of Ming (2000), the cultivated tea plant is currently treated as two varieties, i.e., *C. sinensis* var. *sinensis* (China type tea) and *C. sinensis* var. *assamica* (Masters) Chang (Assam type tea). Although the Cambod type tea (*C. assamica* subsp. *lasiocalyx* Planch) has been treated as a synonym of *C. sinensis* var. *assamica*, recent genetic analyses have revealed that the Cambod type tea is in fact a hybrid between the China and Assam type teas (Wambulwa et al., 2016a). In addition, *C. sinensis* var. *assamica* in China and India has been shown to belong to two distinct genetic entities (Meegahakumbura et al., 2016). Apart from *C. sinensis*, there are eleven more species of *Camellia* sect. *Thea* that occurred in China and several species were used as beverage “tea” in Yunnan, Southwest China (Ming, 2000; Ming and Bartholomew, 2007). *Camellia taliensis* is a close wild relative of domesticated tea that is proposed to have been involved in the domestication and breeding of *C. sinensis* var. *assamica* in Yunnan, China (Li et al., 2015). A comprehensive genetic definition of tea lineages may help to resolve taxonomic delineations among tea taxa that are currently domesticated in China.

Crop domestication is a dynamic and ongoing process where wild progenitors give rise to landraces, which in turn, form the genetic basis of modern cultivars (Diamond, 1997; Zohary, 1999). Domesticated over 3,000 years ago, tea is one of the earliest tree crop species in China (Yamanishi, 1995). Recent excavations of the “Han Yangling Mausoleum” in Xi’an, Shaanxi province, China, revealed that tea drinking was popular among emperors in the Han dynasty more than 2,100 years ago (Lu et al., 2016) providing strong evidence that China type tea originated in China and has a long history of utilization and cultivation in the region. The origin of domestication of the tea plant in China, however, has long been controversial. Several regions, such as Sichuan and Yunnan provinces of China, Assam of India, and Indo-Burma, have long been suspected to be the origin of tea domestication (Stuart, 1919; Wight, 1959; Lu, 1974). In a recent study based on nuclear microsatellite (nSSR) data, two domestication events of tea in China and one in India have been detected (Meegahakumbura et al., 2016). However, the patterns of divergence among these three tea lineages as well as the historical timelines associated with their origins of domestication remain to be explored.

Tea breeding has a long history in China, and can be traced back over 1,000 years (Wu, 1987). Historically, cultivars were grown and selected from naturally occurring seed sources (Chen et al., 2007) and subsequently propagated vegetatively (Yao and Chen, 2012). More recently, artificial pollination and hybridization among selected tea types or with wild relatives have become an important breeding strategy for modern tea cultivar

development (Wachira et al., 1997; Sharma et al., 2010). To date, over 5,100 accessions of tea germplasm have been selected and conserved in China and India (Chen et al., 2007; Das et al., 2012; Yao and Chen, 2012) and all genetic stocks for most tea growing countries were directly or indirectly introduced from either China or India (Meegahakumbura et al., 2016), although some secondary exchange may have occurred between tea producing nations (Gunasekare, 2012; Sriyadi et al., 2012; Wambulwa et al., 2017). Unfortunately, the breeding history of tea germplasm and the extent of exchange of tea genetic resources between China and India remains poorly understood. Because the morphological characters that are traditionally used to define cultivars are highly plastic and easily influenced by environmental conditions (Yao et al., 2008), a molecular analysis that can not only discern patterns of lineage but also the origin of tea germplasm is required.

To further elucidate the origin of the domestication and breeding history of the tea plant in China, we used a combination of nSSR (402 samples) and cpDNA (101 samples) markers to genotype samples of tea landraces, cultivars and their wild relatives collected from China and India. nSSRs, which are co-dominant markers and bi-parentally inherited (Powell et al., 1996) were incorporated into demographic models and NewHybrids analyses were carried out to investigate the domestication history (Carnille et al., 2012; Diez et al., 2015). cpDNA, which is maternally inherited in *Camellia* (Kaundun and Matsumoto, 2011), was used to generate network and neighbor joining (NJ) trees to study the origins of maternal lineages associated with tea domestication. Specifically, we aimed to (1) investigate the demographic history and domestication origin of the tea plant, (2) explore patterns of tea breeding history in the domestication centers of China and India, and (3) clarify the genetic relationships, and origins of tea cultivars in both countries. Our findings will contribute directly to the development of modern tea breeding programmes and future tea germplasm conservation strategies.

## MATERIALS AND METHODS

### Sampling

A total of 402 samples from China and India were included in the present study, 392 samples were from a *Camellia sinensis* dataset of a previous study (Meegahakumbura et al., 2016) and 10 samples came from the closely related wild relative, *C. taliensis* (Table S1). Based on our previous work involving the 392 tea samples, 276 samples were grouped as “pure” tea types and the remaining 116 were classified as a Mosaic group representing hybrids (Meegahakumbura et al., 2016; Table S2). The tea samples were divided into landraces (samples collected from ancient tea trees improved by traditional selection methods) and cultivars (samples generated from modern tea breeding programmes). Hereafter, *C. sinensis* var. *sinensis* will be referred to as “China type tea”; *C. sinensis* var. *assamica* occurring in China as “Chinese Assam type tea”; and *C. sinensis* var. *assamica* that occurs in India as “Indian Assam type tea.” All Cambod tea accessions were treated as mosaic cultivars based on our previous study (Meegahakumbura et al., 2016). Finally, a total of

12 subgroups were defined for the convenience of presentation and discussion. These include: CT: *C. taliensis*; CCL: Landrace of China type tea in China; CCC: Cultivars of China type tea in China; CCCM: Mosaic cultivars of China type tea in China; CIC: Cultivars of China type tea in India; CICM: Mosaics cultivars of China type tea in India; ACL: Landraces of Chinese Assam type tea; ACLM: Mosaic landraces of Chinese Assam type tea; ACC: Cultivars of Chinese Assam type tea; ACCM: Mosaic cultivars of Chinese Assam type tea; AIC: Cultivars of Indian Assam type tea; AICM: Mosaic cultivars of Indian Assam type tea. Vouchers for most of the samples were collected and deposited in the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN).

## DNA Extraction, nSSR Genotyping, and cpDNA Sequencing

Extraction of total genomic DNA was carried out using a CTAB method with slight modifications (Doyle and Doyle, 1990). For the 10 samples of *C. taliensis*, nSSR genotyping was conducted using a set of 23 nSSR primers as described in Meegahakumbura et al. (2016). A total of 101 samples were selected to represent all nSSR NJ tree clades, tea types and origins that were found in our previous study (Meegahakumbura et al., 2016). These samples were then sequenced for three cpDNA intergenic spacer regions: *ndhF-rpl32*, *trnSGG-trnSr*, and *trnSf1-trnGGG* (Yang et al., 2013). Detailed information of the primer sequences and the methodology for cpDNA sequencing are given in our recent study (Wambulwa et al., 2016b). The sequences of the 31 haplotypes were deposited in GenBank (Table S3).

## nSSR Data Analysis

A total of 23 nSSR loci were used to genotype all 402 samples (Table S1), and the nSSR data were checked manually as described in Meegahakumbura et al. (2016). Using the admixture model in STRUCTURE (Prichard et al., 2000), a population structure analysis was performed for samples grouped as “pure” and Mosaic. These were evaluated for 1–10 genetic clusters ( $K$ ) with 20 permutations for each  $K$  value. STRUCTURE was run with 100,000 generations of burn-in followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations. To estimate the optimal number of genetic clusters,  $\Delta K$  was obtained using the method of Evanno et al. (2005) as implemented in STRUCTURE HARVESTER web v.0.6.94 (Earl and von Holdt, 2012). The optimal  $K$  value based on the Log-likelihood method was calculated with STRUCTURE HARVESTER. Genetic diversity indices were then calculated for the grouped dataset (pure tea types) excluding the Mosaic samples. The observed heterozygosity ( $H_o$ ) and percentage of rare alleles ( $R_A\%$ ) (with frequencies  $\leq 0.05$  according to White et al., 1999) for each group were calculated in GenA1Ex v.6.5b4 (Peakall and Smouse, 2006). The total allele number ( $A$ ), gene diversity ( $H_s$ ), allele richness ( $Ar$ ) and inbreeding coefficient ( $F_{is}$ ) for each tea group were estimated with FSTAT v.2.9.3.2 (Goudet, 1995).

We computed three hypothetical models (Figure 1) to infer the possible demographic history and genetic divergence of the tea plant in China and India. “Model A” assumed that Indian Assam type tea originated via hybridization between China

type tea and Chinese Assam type tea. In “Model B,” Indian Assam type tea was assumed to have diverged from China type tea, while “Model C” assumed Indian Assam type tea derived from Chinese Assam type tea. A stepwise nSSR allele evaluation method was used in this analysis. The grouped nSSR dataset of 276 samples, excluding samples categorized as Mosaics, was used to generate 11 summary statistical parameters in ARLEQUIN v.3.5 (Excoffier and Lischer, 2010), following the method of Sun et al. (2014). A standard algorithm of the Approximate Bayesian Computing (ABC) toolbox (Wegmann et al., 2010) and the computer program Fastsimcoal (Excoffier and Foll, 2011) was employed. A total of  $3 \times 10^6$  simulated samples ( $1 \times 10^6$  for each model) were generated to compare the three models. Based on these simulations, a “Bayes-factor” (BF) was calculated to identify the best-fit model for the dataset. The ABC toolbox was also used to calculate the mutation rate for each of the 23 nSSR loci where the mean mutation rate was used to estimate the divergence time. To decrease redundancies associated with the summary statistics, partial least square (PLS) components for a total of 11 statistics were recalculated using the R script of the ABC toolbox (Wegmann et al., 2010). Following the root mean square error plot and conversion equations inferred for 10,000 samples simulated by the standard algorithm, a total of 6 PLS components were used to analyse each of the three models. A regression-adjusted general linear model (GLM) was employed to generate posterior probability distributions for each parameter and to calculate the divergence time between each genetic lineage assumed in the models.

A NewHybrids analysis (Anderson and Thompson, 2002), was carried out following Wambulwa et al. (2016a), with slight modifications. Based on the results of a previous study (Meegahakumbura et al., 2016), three distinct genetic groups were used as parents in the NewHybrids analysis for different datasets: China type tea (*C. sinensis* var. *sinensis*), Chinese Assam type tea (*C. sinensis* var. *assamica* in China), and Indian Assam type tea (*C. sinensis* var. *assamica* in India). For the dataset composed of tea samples collected from India, China type tea and Indian Assam type tea were assigned as parents given that Chinese Assam type tea from India has not been reported (Meegahakumbura et al., 2016). For the dataset of China type tea samples from China (CCCM), China type tea and Chinese Assam type tea were used as parents. In addition, for the dataset of Chinese Assam samples from Yunnan Province of China (ACLM+ACCM), Chinese Assam type tea and *C. taliensis* were used as parents, because *C. taliensis* naturally occurs in sympatric populations with Chinese Assam type tea and has been recognized as contributing to the domestication of Assam type tea in China (Li et al., 2015).

## cpDNA Sequence Data Analysis

The original trace files of the sequences for each region were assembled and edited using Sequencher v.5.0 (Genecodes Co., Michigan, USA). The final sequences of the three cpDNA were concatenated using SequenceMatrix v.1.7.8 (Vaidya et al., 2011) and then aligned in Geneious Pro v.4.8.5 (Biomatters Ltd., USA), and subsequently, manually adjusted where necessary. The combined sequences were assigned to different haplotypes using

DNASP v.5.10 (Librado and Rozas, 2009), with each indel treated as a single mutation event. A neighbor joining (NJ) tree was generated in MEGA v.6 (Tamura et al., 2013) with the pairwise deletion and K2P distance model. A median joining haplotype network was constructed using the program NETWORK v.4.6.1.3 (available at <http://www.fluxus-engineering.com>) with the maximum parsimony (MP) criterion.

## RESULTS

### Genetic Diversity

The summary statistics for the different genetic groups is given in **Table 1**. The highest genetic diversity ( $H_s = 0.731$ ,  $H_o = 0.690$ ) was observed for AIC, followed by CIC ( $H_s = 0.706$ ,  $H_o = 0.652$ ), while *C. taliensis* (CT) had the lowest genetic diversity ( $H_s = 0.642$ ,  $H_o = 0.517$ ). As expected, landraces of both China type tea in China (CCL) and Chinese Assam type tea (ACL) exhibited relatively higher gene diversity (0.702 and 0.642) than those of cultivars (0.684 and 0.625). The highest allelic richness ( $A_r$ ) was recorded for CCL and AIC (4.9), while the lowest was for ACC (4.2). The highest number of alleles ( $A$ ) was observed in CCC (225), followed by AIC (200), and the lowest was recorded for CIC (102). ACL had the highest percentage of rare alleles ( $R_A$ : 48.62%), followed by ACC (47.16%), while no rare alleles were found in CIC.

### Structure Analysis

The optimal  $K$  value was found to be two and four based on the  $\Delta K$  (**Figure S1**) and the Log-likelihood method (**Figure S2**), respectively. The STRUCTURE results with the “admixture model” are given with  $K = 3$  and 4, but not for  $K = 2$  because three distinct tea types were defined in our previous study (Meegahakumbura et al., 2016; **Figure 2**). When  $K$  was 3, three genetic groups were found, representing the China type tea group (green), Chinese Assam type tea (red), and Indian Assam type tea + *C. taliensis* (blue) (**Figure 2**). Interestingly, Indian Assam type tea was not separated from *C. taliensis* at  $K = 3$ . With  $K = 4$ , *C. taliensis* was separated from Indian Assam type tea and formed a distinct genetic group (yellow) (**Figure 2**). Samples of the Mosaic group showed genetic admixtures between tea types or between tea types and *C. taliensis*. For example,

ACCM samples showed admixture between China type tea, Chinese Assam type tea and *C. taliensis*; CICM samples showed admixture between China type tea and Indian Assam type tea; and AICM samples showed admixture between Indian Assam type tea and China type tea. In most accessions, ACLM showed a mixed genetic make-up between Chinese Assam type tea and *C. taliensis*. Compared to a previous study by Meegahakumbura et al. (2016), three samples (LMM159, LMM337, and LMM649) of ACL were re-assigned to the Mosaic group (ACLM) with a genetic admixture between Chinese Assam type tea and *C. taliensis*. Overall, a total of 119 samples were assigned to the Mosaic (hybrids) group in the present study.

### Demographic Modeling Analysis

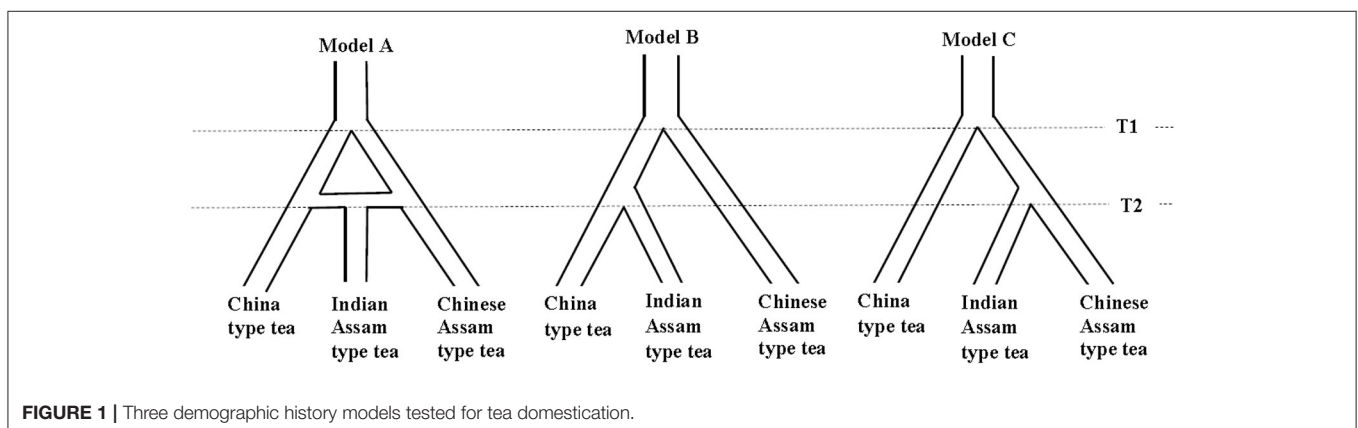
The mean mutation rate ( $\mu$ ) across the 23 nSSR loci was estimated to be  $0.515 \times 10^{-3}$  which was used to evaluate the three models described above. The highest Bayes factor support (BF = 75.3) was found for Model C which assumed that the Indian Assam type tea and Chinese Assam type tea lineages diverged at the time T2 (**Figure 1**), suggesting this to be the

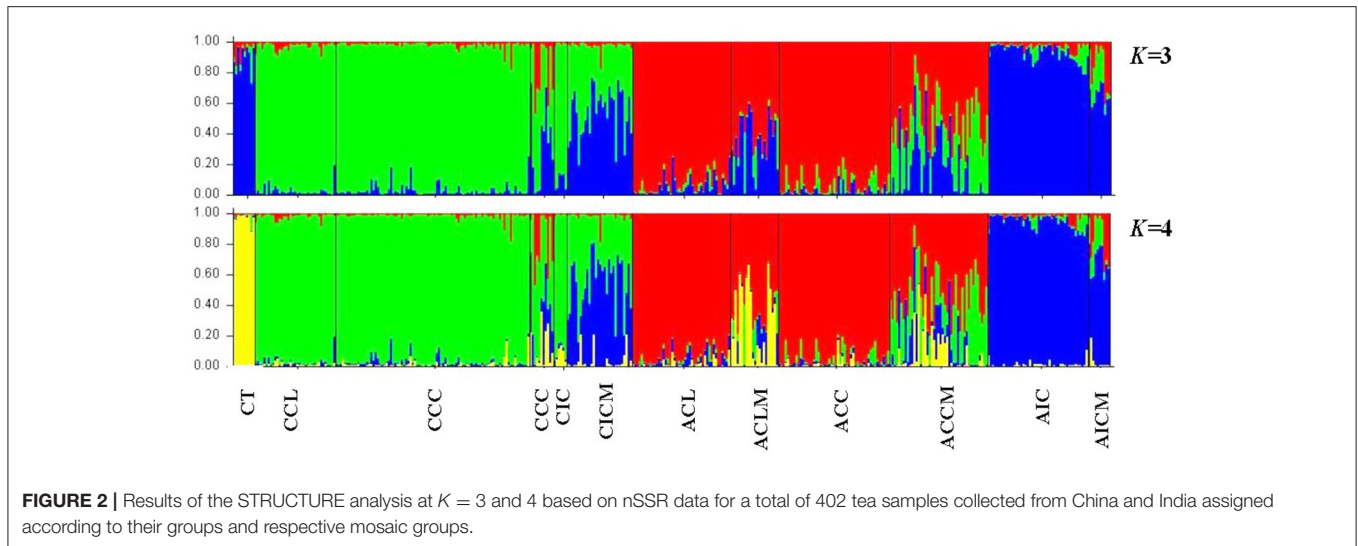
**TABLE 1** | Genetic diversity parameters of the grouped dataset of landraces and modern cultivars excluding the mosaic group.

Group	$N$	$A$	$R_A\%$	$A_r$	$H_o$	$H_s$	$F_{is}$
CT	10	120	31.67	4.3	0.517	0.624	0.171
CCL	36	189	37.57	4.9	0.679	0.702	0.034
CCC	89	225	45.78	4.8	0.620	0.684	0.094
CIC	6	102	0	4.4	0.652	0.706	0.076
ACL	45	181	48.62	4.3	0.547	0.642	0.147
ACC	51	176	47.16	4.2	0.548	0.625	0.122
AIC	46	200	38.50	4.9	0.690	0.731	0.055
Total	283	301		5.6	0.608	0.673	0.1

$N$ , number of samples;  $A$ , number of alleles;  $R_A\%$ , Percentage of rare alleles;  $H_o$ , observed heterozygosity;  $H_s$ , gene diversity;  $F_{is}$ , inbreeding coefficient.

CT, *Camellia taliensis*; CCL, Landrace of China type tea in China; CCC, Cultivars of China type tea in China; CIC, Cultivars of China type tea in India; ACL, Landraces of Chinese Assam type tea; ACC, Cultivars of Chinese Assam type tea; AIC, Cultivars of Indian Assam type tea.





**FIGURE 2** | Results of the STRUCTURE analysis at  $K = 3$  and  $4$  based on nSSR data for a total of 402 tea samples collected from China and India assigned according to their groups and respective mosaic groups.

best-fit model for the demographic history of tea domestication. Model A (the hybrid model) had the lowest Bayes factor support ( $BF = 0.0477$ ), and Model B with the assumption of Indian Assam type tea and China type tea diverging at time T2 also exhibited a low Bayes factor support ( $BF = 1$ ). Therefore, Model C was used to estimate population divergent times T1 and T2. The divergence time between China type tea and Assam tea (T1) was estimated with 1,875 generations, and the divergence time between Indian Assam type tea and Chinese Assam type tea (T2) was estimated with 231 generations. With an estimated generation time for the tea plant of around 12 years (Duke, 1983), we calculated that China type tea and Assam tea diverged approximately 22,552 years ago (95% Highest Probability Density Interval [HPDI]: 2,598–1,043,593), while Chinese and Indian Assam type tea diverged more recently, approximately 2,770 years ago (95% HPDI: 1,200–235,932).

### NewHybrids Analysis

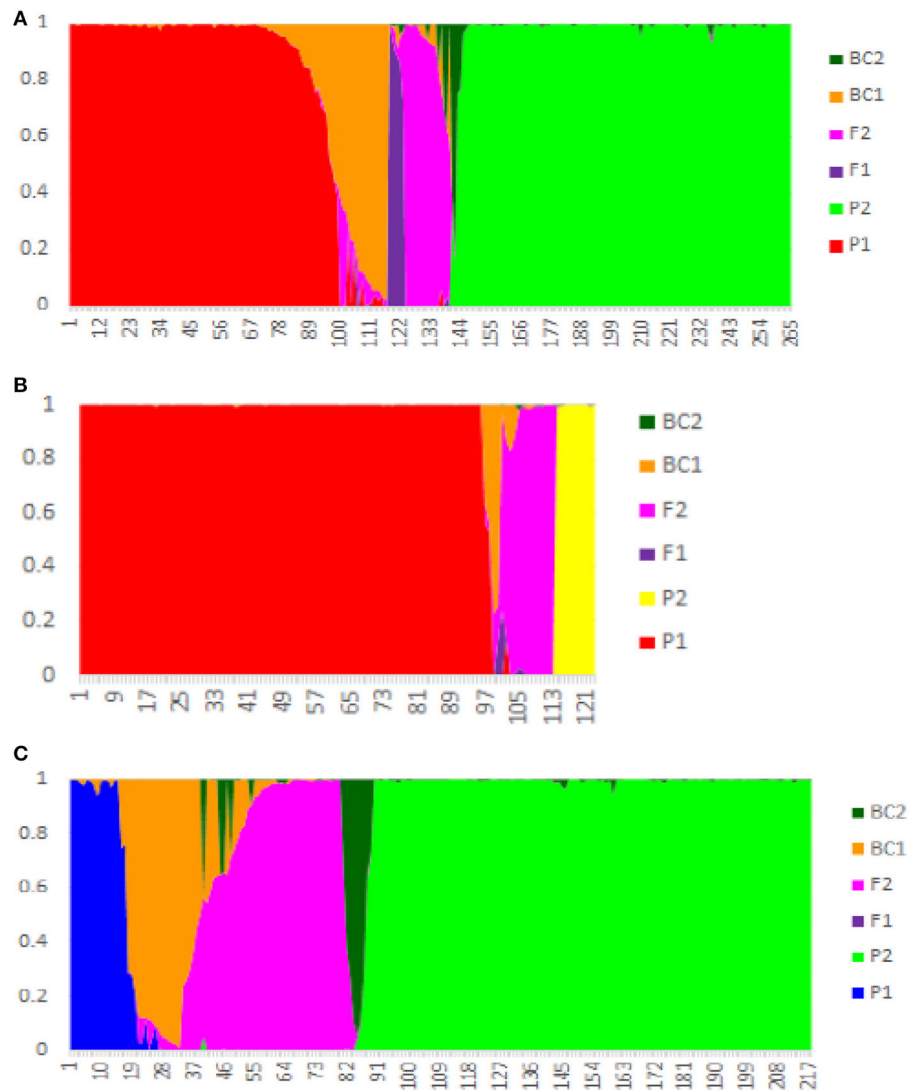
Among the 119 hybrid samples included here, only six (5.0%) were identified as F<sub>1</sub> hybrids, while 64 (53.8%) were assigned as F<sub>2</sub> hybrids, and 30 (25.21%) as backcross (BC) hybrids. The remaining 16 samples showed a complex genetic admixture from at least three tea group gene pools, suggesting that they underwent a complex breeding process and were excluded from the NewHybrids analysis (Figure 3 and Table 2). Among the 30 CICM samples (mainly from Darjeeling, India), 22 were F<sub>2</sub> hybrids between Indian Assam type tea and China type tea and five were backcrosses to Indian Assam type tea. The remaining three samples (144, 190, and 192) were assigned as China type tea, although they were defined as hybrids in the STRUCTURE analysis. Of the 22 samples of ACLM, 11 were defined as hybrids (10 F<sub>2</sub> and one BC<sub>2</sub>) between Chinese Assam type tea and *C. taliensis* and seven represented complex hybrids involving more than two parents. The remaining four samples were identified as representing BC<sub>1</sub> (3 hybrids) and one F<sub>1</sub> hybrid between Chinese Assam type tea and China type tea. Among the 45 ACCM samples, 33 were assigned as hybrids between Chinese

Assam type tea and China type tea, six were hybrids between Chinese Assam type tea and *C. taliensis* (three F<sub>2</sub> and three BC<sub>1</sub>) and six showed multiple hybridization events with more than two possible parents. All 10 samples of AICM were F<sub>2</sub> hybrids between Indian Assam type tea and China type tea (Table 2). In addition, a total of 42 samples that were identified as “pure” in STRUCTURE were assigned as hybrids according to NewHybrids analysis (Table S2). Out of these, 31 were Indian Assam type tea, which included 22 BC<sub>1</sub> and nine F<sub>2</sub> hybrids, the remaining 11 samples (CCC: 4; ACL: 2 and ACC: 5) were BC hybrids.

### cpDNA Sequence Data and Genetic Relationships

The length of the three cpDNA regions, *ndhF-rpl32*, *trnSGG-trnSr*, *trnSf1-trnGGG* was 662, 692, and 581 bp, respectively. A total of 31 haplotypes were defined for the 101 samples that were sequenced (Table S3). Four haplotypes were defined for five *C. taliensis* samples, of which H6 (TMK2) and H7 (TBW1) were private, while H15 (TY7, TWQ1) and H17 (TXW1) were shared with five landrace samples of Chinese Assam type tea (Table S3). H1 was the dominant haplotype of China type tea detected in the 22 samples that were from China and India, as well as in the hybrids of China type tea (MW153, 144) and a Chinese Assam type tea sample (704). H22 and H24 were the two dominant haplotypes of Indian Assam type tea, but were also present in a few samples of China type tea hybrids from India. H11 and H23 were two haplotypes that were private to Indian Assam type tea. Chinese Assam type tea had a dominant haplotype H26 and eight private haplotypes (H10, H12, H14, H16, H21, H29, H30, and H31). For the 101 sequenced samples, 43 were found to be hybrids according to the NewHybrids and STRUCTURE analyses (Table S3).

The network tree for the 31 haplotypes was separated into three major clades (Figure 4). Clade A included thirteen haplotypes: three private haplotypes of *C. taliensis*, eight haplotypes of Chinese Assam type tea from Southern Yunnan (Xishuangbanna and Pu'er) of China, and one private haplotype



**FIGURE 3** | NewHybrid analysis based on nSSR data of 402 tea samples from China and India. **(A)** Chinese Assam type tea as P1 and China type tea in China as P2; **(B)** Chinese Assam type tea as P1 and *Camellia taliensis* as P2; **(C)** Indian Assam type tea as P1 and China type tea from China and India as P2.

of Indian Assam type tea (H11). This clade also included one private haplotype (H18) of China type tea which showed a distinct divergence from the other China type tea samples in Clade B. Clade B comprised four haplotypes of China type tea (H1, H3, H4, and H5) plus one haplotype (H2) private to Indian Assam type tea. Clade C contained eight haplotypes of Chinese Assam type tea that were mainly from Western Yunnan (Lincang and Baoshan) in China as well as four haplotypes of Indian Assam type tea. No haplotype was shared between Chinese Assam type tea and Indian Assam type tea.

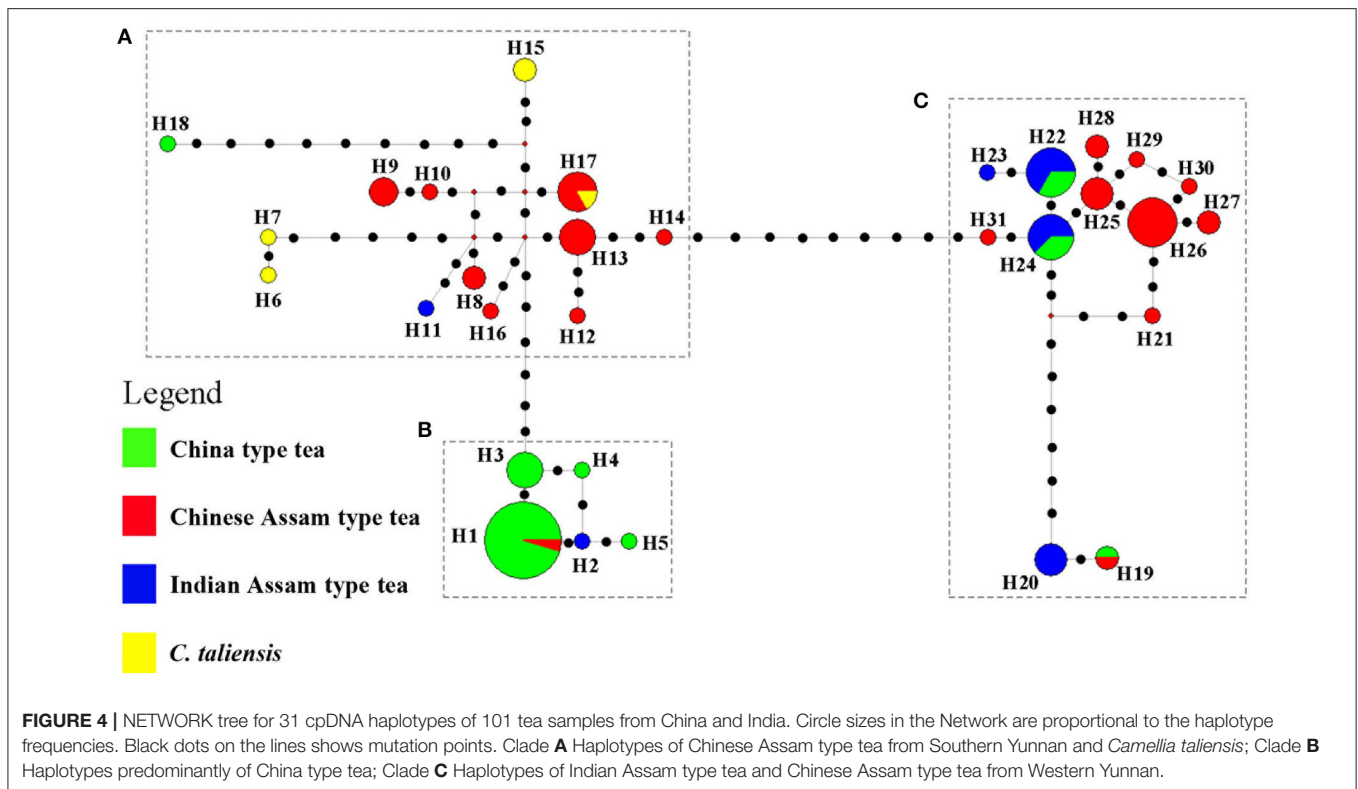
The NJ tree of the 101 sequenced samples showed similar relationships to those in the haplotype network tree (Figure 5). Clade A consisted of all haplotypes of China type tea and also included several hybrid samples of China type tea (832, MW 153, 144) (Table S2). The China type tea sample 885 collected from Jiangxi was sister to the remaining samples of this clade. Clade

B included haplotypes of Chinese Assam type tea from South Yunnan, *C. taliensis*, and the haplotype of one Indian Assam type tea sample (120). Some hybrid Chinese Assam type tea samples (YX 1, MQ 7, MH 2, MZ 9, JM 13, NL 2, NL 11, 701, and 760) were also in this clade. Cluster C comprised haplotypes of Chinese Assam type tea from Western Yunnan, Indian Assam type tea and hybrids of China type tea from India.

## DISCUSSION

### New Insights into the Origin and Domestication of the Tea Plant

In a previous study, we defined three distinct genetic lineages among tea plant accessions; Southern China, the Southwestern Yunnan Province of China, and Assam of India, representing



the three potential areas of domestication of China type tea, Chinese Assam type tea and Indian Assam type tea, respectively (Meegahakumbura et al., 2016). In the present study, the existence of three independent gene pools and separate origins for the tea plant was confirmed by the cpDNA data. Chinese Assam type tea, however, which formed a distinct genetic group based on nSSR data (Figure 2), had cpDNA haplotypes that grouped into two distinct clades in the network (Figure 4) and NJ trees (Figure 5). The haplotypes of Chinese Assam type tea from Southern Yunnan (Pu'er and Xishuangbanna) of China formed a distinct clade with haplotypes of *C. taliensis*, whereas the haplotypes of Chinese Assam type tea from Western Yunnan (Lincang and Baoshan) grouped together with haplotypes of Indian Assam type tea. We speculate that the Southern Yunnan Chinese Assam type tea possibly had a different maternal parent than the Western Yunnan Chinese Assam type tea. This pattern may be indicative of two possible independent origins for Assam type tea in Yunnan, China. The inconsistency between nSSR and cpDNA results for the Chinese Assam type tea may be explained by historical exchanges of tea germplasm (Freeman and Ahmed, 2010) and frequent gene flow (especially in the modern tea cultivar breeding programs) between the two neighboring regions in Yunnan which, in turn, might have resulted in the homogenization of the nuclear genetic composition and has obscured pre-existing genetic boundaries. Although the cpDNA haplotypes of Chinese Assam type tea in Western Yunnan were genetically close to Indian Assam type tea, with an exception of the hybrid TF 5 with haplotype H19 (Figure 4), no haplotype was shared between each other. This

result agreed with our previous results and confirms the existence of independent domestication events of China type and Indian Assam type teas (Meegahakumbura et al., 2016).

Our results indicated that Chinese Assam type tea in Western Yunnan and Indian Assam type tea in Assam of India may have arisen from a single ancestral population from an area where Southwest China, Indo-Burma, and Tibet meet (Kingdon-Ward, 1950). From this origin of domestication, consecutive independent domestication events likely occurred in Western Yunnan of China and Assam of India. In addition, both nSSR and cpDNA results indicated that China type tea is a distinct genetic lineage, which suggests a separate domestication event for this tea type (Meegahakumbura et al., 2016). Although we propose Southern China as the area of domestication of China type tea, the specific origin of domestication remains unknown due to the lack of wild populations. Thus, future studies that include a broader sampling of wild relatives of the tea plant as well as more samples of ancient trees from or near wild populations would likely enhance our resolution of the exact origin of domestication of tea in China.

## Demographic Modeling and the Beginning of Tea Domestication

The tea plant was initially domesticated in China over 4,000 year ago (Yamanishi, 1995). Demographic modeling with Approximate Bayesian Computing (ABC) has been used to estimate divergence times of crop plants based on the evolutionary rates of nSSR data (Carnille et al., 2012; Diez et al., 2015). Furthermore, Duke (1983) proposed a generation





**TABLE 2** | Parentage and types of hybrids in the Mosaic group obtained using NewHybrids.

Mosaic category	P1 (maternal)	P2 (paternal)	F <sub>1</sub>	F <sub>2</sub>	BC1	BC2	Total
CCLM	Chinese Assam type tea	China type tea				1	1*
CCCM	Chinese Assam type tea	China type tea	1	3			4
	Indian Assam type tea	China type tea		1		3	4
	Complex***						3
CICM	Indian Assam type tea	China type tea		22	3	2	27
		China type tea					3**
ACLM	Chinese Assam type tea	<i>C. taliensis</i>		10		1	11
	Chinese Assam type tea	China type tea	1		3		4
	Complex***						7
ACCM	Chinese Assam type tea	<i>C. taliensis</i>		3	3		6
	Chinese Assam type tea	China type tea	4	15	14		33
	Complex***						6
AICM	Indian Assam type tea	China type tea		10			10
Total			6	64	23	7	119

\*Single sample of China type tea landrace was belonging to Mosaic group, hence a separate CCLM category was not formed; \*\*Three samples of CICM were assigned to Parental China type tea in NewHybrid analysis; \*\*\*Complex: 16 samples in total showing a complex mixing of more than two parental gene pools according to STRUCTURE results were defined as "Complex" hybrids and excluded from analysis.

samples) were found to be F<sub>2</sub> and BC hybrids. It also seems likely that closely related wild relatives of the tea plant have contributed to tea cultivar development in China (Li et al., 2015). For example, the Indian Assam type tea samples 115, 122, 124, 201, possess haplotype H20, which is genetically distinct from other Indian Assam type tea cultivars (Figure 5). Furthermore, results of BLAST searches in GenBank for the three cpDNA regions used in this study matched the wild species *C. pubicosta*. These samples are likely F<sub>2</sub> hybrids or backcross generations between Indian Assam type tea and *C. pubicosta* as the maternal parent suggesting the introduction of wild tea plant germplasm into the Indian tea breeding program since at least before 1890, the planting date of one of our samples (201). Similarly, sample 885, which is a backcross hybrid (Table S2) possessing the H18 haplotype, possibly hybridized with a wild *Camellia* species of *Camellia* sect. *Thea*. Collectively, our results indicate that germplasm of wild species of *Camellia* was involved in the domestication and breeding of the tea plant.

Comparison of the breeding histories in different tea domestication regions could provide valuable information for the direction of future breeding programs. Unfortunately, different countries have adopted different breeding objectives and have applied different genetic systems (markers) to identify germplasm and confirm lineages, which makes it difficult to make comparisons among studies (Chen et al., 2012). This study is the first to use the same genetic markers to compare the origin and breeding history of the three main tea types currently in cultivation in China and India. Our results indicated that China type tea, Chinese Assam type tea and Indian Assam type tea may have undergone different breeding processes in their respective domestication regions. More specifically, China type tea was possibly derived from a single domestication event in

Southern China (Meegahakumbura et al., 2016). This suggests that the largely "pure" genetic composition of China type tea might be attributable to the limited gene flow between China type tea and its wild relatives, the wild species is not in the close vicinity of major growing areas of China type tea in China. Interestingly, the landraces of China type tea showed a lower proportion of hybrids compared to modern China type tea cultivars (2.7 vs. 15%), but exhibited relatively higher gene diversity than the cultivars ( $H_e = 0.692$  vs.  $H_s = 0.680$ ). This result might indicate that the seeds of China type landraces were collected from a relatively wide gene pool, while modern cultivars may have been bred by selection and/or artificial hybridization involving a few tea individuals representing a limited gene pool (Tan et al., 2015). Two modern cultivars of China type tea collected from Guangdong, China are shown to be hybrids with Indian Assam type tea (MW157, MW163), indicating that Indian Assam tea has been used recently in tea breeding programs in China. They both possess a maternal lineage of Indian Assam type tea (Figure 5), thus China type tea was used as paternal parent. Overall, our results indicated that hybridization between the tea types and their closely related wild species, and/or among different tea types might have played a major role in tea breeding.

The cpDNA results revealed that Chinese Assam type tea possibly has two origins in China (Southern and Western Yunnan) and possesses high haplotype diversity. According to the nSSR analysis, landraces of Chinese Assam type tea exhibited relatively higher gene diversity than cultivars ( $H_e: 0.633$  vs.  $0.618$ ). Natural introgression from other sympatric wild relatives in this region could have possibly contributed to the higher gene diversities found in this study. For example, approximately 50% of landrace samples of Chinese Assam type tea (ACLM) exhibited a genetic admixture between Chinese Assam type tea

and *C. taliensis* (Figure 2 and Table 2). Furthermore, *C. taliensis* haplotype H17 was shared with several samples of Chinese Assam type tea (Figure 4; Table S3). *C. taliensis* is a related wild species of the tea plant that has a long history of cultivation in ancient tea gardens alongside Chinese Assam type tea plants in Yunnan (Zhao et al., 2014). Thus, gene flow between Chinese Assam type tea and *C. taliensis* has likely resulted in Chinese Assam type tea having relatively higher genetic diversity in hybrid individuals. In turn, it seems likely hybrid landraces were selected by farmers for their unique taste, due to their low caffeine and high theobromine content (Ogino et al., 2009). Interestingly, more than 73% (33 out of 45) of the samples of ACCM (Figure 2; Table 2) showed genetic admixture between China type tea and Chinese Assam type tea, with the latter usually serving as the maternal parent. In addition, 13 hybrid landraces / cultivars of Chinese Assam type tea showed genetic admixture from more than two species/tea types. This suggested that multiple hybridization events between landraces/cultivars and wild species/different tea types may have occurred during the breeding of tea cultivars. Collectively, these results indicated that modern breeding programs should implement mating between distinct tea types or related wild species via artificial hybridization to breed new cultivars and bolster genetic diversity of germplasm stocks.

Although the history of tea plantations in India is relatively short, less than 200 years (Das et al., 2012), the Indian Assam type tea has been proposed to have been domesticated independently in Assam of India (Meegahakumbura et al., 2016). It is known that China type tea germplasm was introduced to India in 1836 (Sharma et al., 2010). However, it did not grow well and was hybridized with indigenous tea accessions and wild relatives from a wide range of geographic areas in India. Subsequently, the selection of tea cultivars was performed from a highly heterogeneous gene pool (Sharma et al., 2010). The high genetic diversity and higher proportion of F<sub>2</sub> and BC hybrids among the accessions of Indian Assam type tea detected in our study (Table 2; Table S2) possibly reflects early hybridization events between the introduced China type tea and local Assam tea accessions, and maybe even with wild species of *Camellia* in India (Raina et al., 2011). This seems likely given that all the commercial cultivars in our study that were from Darjeeling, India were found to be hybrids between China type tea and Indian Assam type tea, and most of them with Indian Assam type tea serving as the maternal parent. Furthermore, China type tea, which naturally comes from cold climates, might have contributed cold tolerance genes to hybrids typically found in high elevations of the Darjeeling area. Thus, hybrids between different tea types exhibiting higher environmental adaptability or possessing more favored agricultural traits than their parents, may have played a substantial role in tea cultivation in India and China.

## Implications for Tea Germplasm Conservation

The three distinct types of the tea plant represent a rich gene pool, which is valuable for future tea breeding and cultivar

development. For example, Assam tea from India and China type tea are the only two gene pools that are currently used for the development of hybrid tea cultivars in approximately 30 countries worldwide (Ellis, 1995). In particular, Chinese Assam type tea from Southwestern Yunnan, China was recently found to represent a distinct gene pool (Meegahakumbura et al., 2016) and both landraces and cultivars of Chinese Assam type tea possessed a high proportion of rare alleles and private haplotypes. Given these results, Chinese Assam type tea is a highly valuable, yet underutilized, germplasm resource for future tea breeding. Furthermore, the existence of ancient trees that are presumed to be over 1,000 years old may present a unique genetic window into tea domestication in China. Therefore, the preservation of these tea plant germplasm resources through *in-situ* conservation should be considered a conservation priority in Yunnan, China.

## DATA ACCESSIBILITY

The Microsatellite Data sets supporting the results in this study have been deposited in the Dryad Digital Repository under the doi: 10.5061/dryad.6gq1g. The cpDNA sequence data in this study have been deposited in NCBI GenBank under accession numbers MG676240-MG676332 as shown in the Table S3.

## AUTHOR CONTRIBUTIONS

L-MG, D-ZL, and J-CX conceived and designed the study. MKM, M-ML, KT, L-MG, J-BY, B-YL, and J-CX collected plant material. MKM, M-ML and MW performed the experiments. MKM, MW, Y-SS, JL, MM, and L-MG analyzed and interpreted the data. MKM, MW, Y-SS, JL, MM, J-CX, D-ZL, and L-MG wrote and revised the manuscript. All authors reviewed and approved the final manuscript.

## ACKNOWLEDGMENTS

We thank J. Yang, H. T. Li and Z. Y. Zhang from the Molecular Biology Centre of the Germplasm Bank of Wild Species, CAS Kunming Institute of Botany for their help of lab work. We are grateful to C. N. Fu, S. X. Yang, Q. Wei, X. J. Ge, X. Hong, Y. X. Qiu, Y. X. Zhang, L. N. Zhang, J. Y. Zhou, D. Wu, L. Fang, F. Luo, and P. H. Fu for providing tea germplasm in China, and the Tea Board India and Tea Research Association are thanked for providing their tea germplasm for this study. We are also grateful to R. Barhakur and J. S. Kandal of Amalgamated Plantations, India for providing samples from their tea germplasm collections. Dr. L. Perera and Dr. S. A. C. N. Perera kindly acknowledge for their comments on the manuscript. This work was funded by the National Natural Science Foundation of China (31161140350 to L-MG; 31160175 to B-YL), and the Program of Science and Technology Talents Training of Yunnan Province, China (2017HA014). The Royal Botanic Garden Edinburgh is supported by the Rural and Environment Science and Analytical Services division (RESAS) in the Scottish Government.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02270/full#supplementary-material>

## REFERENCES

- Anderson, E. C., and Thompson, E. A. (2002). A model-based method for identifying species hybrids using multi-locus genetic data. *Genetics* 160, 1217–1229.
- Banerjee, B. (1992). “Botanical Classification of Tea,” in *Tea: Cultivation and Consumption*, eds K. Wilson, and N. Clifford (London: Chapman and Hall), 25–51.
- Carnille, A., Gladieux, P., Smulder, M. J. M., Roldén-Ruiz, I., Laurens, F., and Cam, B. L. (2012). New insight into history of domesticated apple: secondary contribution of European wild apple to genome of cultivated varieties. *PLoS Genet.* 8:e1002703. doi: 10.1371/journal.pgen.1002703
- Chen, L., Apostolides, Z., and Chen, Z. M. (2012). *Global Tea Breeding*. Hangzhou; New York, NY: Zhejiang University Press; Springer Heidelberg.
- Chen, L., Zhou, Z. X., and Yang, Y. J. (2007). Genetic improvement and breeding of the tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. *Euphytica* 154, 239–248. doi: 10.1007/s10681-006-9292-3
- Chen, Z. M., and Chen, L. (2012). “Delicious and healthy tea: an overview,” in *Global Tea Breeding: Achievements Challengers and Prospective*, eds L. Chen, Z. Apostolides, and Z. M. Chen (Hangzhou; Berlin; Heidelberg: Zhejiang University Press; Springer-Verlag), 1–11.
- Cook, C. G., Jones, R. T., Langdon, P. G., Leng, M. J., and Zhang, E. (2011). New insight into late quaternary Asian palaeomonsoon variability and the timing of the last glacial maximum in Southwest China. *Q. Sci. Rev.* 30, 808–820. doi: 10.1016/j.quascirev.2011.01.003
- Cook, C. G., Leng, M. J., Jones, R. T., Langdon, P. G., and Zhang, E. (2012). Lake ecosystem dynamics and linked to climate change inferred from a stable isotope and organic palaeorecords from Mountain Lake in Southwest China. *Q. Res.* 77, 132–137. doi: 10.1016/j.yqres.2011.09.011
- Das, S. C., Das, S., and Hazarika, M. (2012). “Breeding of Tea Plant (*Camellia sinensis*) in India,” in *Global Tea Breeding: Achievements Challengers and Prospective*, eds L. Chen, Z. Apostolides, and Z. M. Chen (Hangzhou; Berlin; Heidelberg: Zhejiang University Press; Springer-Verlag), 69–124.
- Diamond, J. (1997). *Guns Germs and Steel: The Fates of Human Societies*. New York, NY: NW Norton.
- Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206, 436–447. doi: 10.1111/nph.13181
- Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.
- Duke, J. (1983). *Handbook of Energy Crops*. West Lafayette: Purdue University.
- Earl, D. A., and von Holdt, B. M. (2012). Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Res.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Ellis, R. T. (1995). “Tea,” in *Evolution of Crop Plants, 2nd Edn.* eds J. Smart and N. W. Simmons (Harlow: Longman Scientific and Technical), 22–27.
- Endicott, P., Ho, S. Y. W., Metspalu, M., and Stringer, C. B. (2009). Evaluating the mitochondrial timescale of human evolution. *Trends Ecol. Evol.* 24, 515–521. doi: 10.1016/j.tree.2009.04.006
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., and Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27, 1332–1334. doi: 10.1093/bioinformatics/btr124
- Excoffier, L., and Lischer, H. E. L. (2010). ARlquinsuite ver 3.5: a new series of programmes to perform population genetics analysis under Linux and Windows. *Mol. Ecol. Res.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- FAOSTAT (2015). *FAO Database*. Food and Agriculture Organization; United Nations. Available online at: <http://faostat3.fao.org/download/Q/QC/E> (Accessed May 10, 2015).
- Freeman, M., and Ahmed, S. (2010). *Tea Horse Road: China's Ancient Trade Road to Tibet*. Bangkok: River Book Press.
- Goudet, J. (1995). FSTAT(version 1.2): a computer program to calculate F statistics. *J. Hered.* 85, 485–486. doi: 10.1093/oxfordjournals.jhered.a111627
- Gunasekare, M. T. K. (2012). “Tea Plant (*Camellia sinensis*) breeding in Sri Lanka,” in *Global Tea Breeding: Achievements Challengers and Prospective*, eds L. Chen, Z. Apostolides, and Z. M. Chen (Hangzhou; Berlin; Heidelberg: Zhejiang University Press; Springer-Verlag), 125–176.
- Kaundun, S. S., and Matsumoto, S. (2011). Molecular evidence for maternal inheritance of the chloroplast genome in tea, *Camellia sinensis* (L.) O. Kuntze. *J. Sci. Food Agric.* 91, 2660–2663. doi: 10.1002/jsfa.4508
- Kingdon-Ward, F. (1950). Does wild tea exist? *Nature* 165, 297–299. doi: 10.1038/165297a0
- Li, M. M., Meegahakumbura, M. K., Yan, L. J., Liu, J., and Gao, L. M. (2015). Genetic involvement of *Camellia taliensis* in the domestication of *Camellia sinensis* var. assamica (Assamica Tea) revealed by nuclear microsatellite markers. *Plant Divers. Res.* 37, 29–37. doi: 10.7677/ynzwjy201514048
- Librado, P., and Rozas, J. (2009). DnaSP V5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Lu, H. Y., Zhang, J. P., Yang, Y. M., Yang, X. Y., Xu, B. Q., Yang, W. Z., et al. (2016). Earliest tea as evidence for one branch of the silk road across the Tibetan Plateau. *Sci. Rep.* 6:18955. doi: 10.1038/srep18955
- Lu, Y. (1974). *The Classic of Tea*. Introduced and Transl. by F. Ross Carpenter and H. Demi (Boston, MA: Little Brown & Co.).
- Meegahakumbura, M. K., Wambulwa, M. C., Thapa, K. K., Li, M. M., Möller, M., Xu, J. C., et al. (2016). Indications for three independent domestication events for tea plant (*Camellia sinensis* (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PLoS ONE* 11:e0155369. doi: 10.1371/journal.pone.0155369
- Ming, T. L. (2000). *Monograph of the genus Camellia*. Kunming: Yunnan Science and Technology Press.
- Ming, T. L., and Bartholomew, B. (2007). “Theaceae,” in *Flora of China* Vol 12. eds Z. Y. Wu, P. H. Raven, and D. Y. Hong (Beijing; St. Louis, MI: Science Press; Missouri Botanical Garden Press), 366–478.
- Mondal, T. K., Bhattacharya, A., Laxmikumar, M., and Ahuja, P. S. (2004). Recent advances in tea (*Camellia sinensis*) biotechnology. *Plant Cell Tiss. Org.* 76, 195–254. doi: 10.1023/B:TICU.0000009254.87882.71
- Ogino, A., Tanaka, J., Taniguchi, F., Yamamoto, M. P., and Yamada, K. (2009). Detection and characterization of caffeine less tea plant originated from inter-specific hybridization. *Breed. Sci.* 59, 277–283. doi: 10.1270/jsbbs.59.277
- Peakall, R., and Smouse, P. (2006). GenALEX 6. genetic analysis in excel. population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x
- Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2, 225–238. doi: 10.1007/BF00564200
- Prichard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Raina, S. N., Ahuja, P. S., Sharma, R. K., Das, S. C., Bhardwaj, P., Negi, R., et al. (2011). Genetic structure and diversity of India hybrid tea. *Genet. Resour. Crop. Evol.* 59, 1527–1541. doi: 10.1007/s10722-011-9782-6
- Sealy, J. R. (1958). *A Revision of the Genus Camellia*. London: Royal Horticultural Society.

**Figure S1** | Best *K* values for STRUCTURE analysis using two methods of (A)  $\Delta K$  and (B) Log Likelihood (*K*).

**Table S1** | SSR dataset for *Camellia taliensis*.

**Table S2** | Collection Information of the 402 tea samples used in the current study.

**Table S3** | Details of the haplotypes obtained with cpDNA sequencing.

- Sharma, R. K., Negi, M. S., Sharma, S., Bhardwaj, P., Kumar, R., Bhattacharya, E., et al. (2010). AFLP-based genetic diversity assessment of commercially important tea germplasm in India. *Biochem. Genet.* 48, 549–564. doi: 10.1007/s10528-010-9338-z
- Sriyadi, B., Suprihatini, R., and Khomaeni, H. S. (2012). “The development of high yielding tea clones to increase Indonesian tea production,” in *Global Tea Breeding: Achievements Challenges and Prospective*, eds L. Chen, Z. Apostolides, and Z. M. Chen (Hangzhou; Berlin; Heidelberg: Zhejiang University Press; Springer-Verlag), 299–308.
- Stuart, C. P. (1919). A basis for tea selection. *Bull. Jard. Bot. Buitenzorg* 1, 193–320.
- Sun, Y. S., Abbott, R. J., Li, L., Zou, J. B., and Liu, J. Q. (2014). Evolutionary history of purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet plateau: homoploid hybrid origin and Pleistocene expansion. *Mol. Ecol.* 23, 243–259. doi: 10.1111/mec.12599
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetic analysis version 6.0. *Mol. Ecol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tan, L. Q., Peng, M., Xu, L. Y., Wang, L. Y., Chen, S. X., Zou, Y., et al. (2015). Fingerprinting 128 Chinese clonal tea cultivars using SSR markers provide new insights into their pedigree relationships. *Tree Genet. Genomes* 11:90. doi: 10.1007/s11295-015-0914-6
- Vaidya, G., Lohman, D. J., and Meier, R. (2011). Sequence Matrix: concatenation software for the fast assembly of multigene datasets with character set and codon information. *Cladistics* 27, 171–180. doi: 10.1111/j.1096-0031.2010.00329.x
- Wachira, F. N., Powell, W., and Waugh, R. (1997). An assessment of genetic diversity among *Camellia sinensis* L (cultivated tea) and its wild relatives based on randomly amplified polymorphic DNA and organelle-specific STS. *Heredity* 78, 603–611. doi: 10.1038/hdy.1997.99
- Wambulwa, M. C., Meegahakumbura, M. K., Chalo, R., Kamunya, S., Muchugi, A., Xu, J. C., et al. (2016a). Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree Genet. Genomes* 12:11. doi: 10.1007/s11295-015-0963-x
- Wambulwa, M. C., Meegahakumbura, M. K., Kamunya, S., Muchugi, A., Möller, M., Liu, J., et al. (2016b). Insights into genetic relationships and breeding patterns of African tea germplasm based on nSSR and markers and cpDNA sequences. *Front. Plant Sci.* 7:1244. doi: 10.3389/fpls.2016.01244
- Wambulwa, M. C., Meegahakumbura, M. K., Kamunya, S., Muchugi, A., Möller, M., Liu, J., et al. (2017). Multiple origins and narrow genepool characterize the African tea germplasm: concordant patterns revealed by nuclear and plastid DNA markers. *Sci. Rep.* 7:4053. doi: 10.1038/s41598-017-04228-0
- Wegmann, D., Leuenberger, C., Neuenschwander, S., and Excoffier, L. (2010). ABCtoolbox: a versatile toolkit for Approximate Bayesian computations. *BMC Bioinformatics* 11:116. doi: 10.1186/1471-2105-11-116
- Wight, W. (1959). Nomenclature and classification of tea plant. *Nature* 138, 1726–1728. doi: 10.1038/1831726a0
- Wight, W. (1962). Tea Classification revised. *Curr. Sci.* 31, 298–299.
- White, G. M., Boshier, D. H., and Powell, W. (1999). Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Mol. Ecol.* 8, 1899–1909. doi: 10.1046/j.1365-294x.1999.00790.x
- Wu, J. N. (1987). *Review on 'Cha Ching'*. Beijing: Agriculture Press.
- Yamanishi, T. (1995). Special issue on tea. *Food Rev. Int.* 11, 371–546.
- Yang, J. B., Yang, S. X., Li, H. T., Yang, J., and Li, D. Z. (2013). Comparative chloroplast genomes of *Camellia* species. *PLoS ONE* 8:e73053. doi: 10.1371/journal.pone.0073053
- Yao, M. Z., and Chen, L. (2012). “Tea germplasm and breeding in China,” in *Global Tea Breeding: Achievements Challenges and Prospective*, eds L. Chen, Z. Apostolides, and Z. M. Chen (Hangzhou; Berlin; Heidelberg: Zhejiang University Press; Springer-Verlag), 13–68. doi: 10.1007/978-3-642-31878-8\_2
- Yao, M. Z., Chen, L., and Liang, Y. R. (2008). Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in the breeding programs. *Plant Breed.* 127, 166–172. doi: 10.1111/j.1439-0523.2007.01448.x
- Yao, Y. F., Song, X. Y., Wortley, A. H., Blackmore, S., and Li, C. S. (2015). A 22570 year record old vegetational and climate change from Wenhai Lake in Hengduan Mountain biodiversity hotspot, Yunnan, Southwest China. *Biogeoscience* 12, 1525–1535. doi: 10.5194/bg-12-1525-2015
- Zhao, D. W., Yang, J. B., Yang, S. X., Kato, K., and Luo, J. P. (2014). Genetic diversity and domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by microsatellite markers. *BMC Plant Biol.* 14:14. doi: 10.1186/1471-2229-14-14
- Zhao, F. R., and Yin, Q. Y. (2008). The Khmer Meng nationalities in China earliest domesticated cultivated tea. *J. Simao Teacher's College* 24, 28–34.
- Zohary, D. (1999). Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genet. Res. Crop Evol.* 46, 133–142. doi: 10.1023/A:1008692912820

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Meegahakumbura, Wambulwa, Li, Thapa, Sun, Möller, Xu, Yang, Liu, Liu, Li and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.