Tropical Agricultural Research Vol. 29 (3): 230 – 240 (2018)

Preliminary Investigations on the Genetic Relationships and Origin of Domestication of the Tea Plant (*Camellia sinensis* (L.)) Using Genotyping by Sequencing

M.K. Meegahakumbura[1], M.C. Wambulwa, D.Z. Li and L.M. Gao

Key Laboratory for Plant Diversity and Biogeography of East Asia
Kunming Institute of Botany
Chinese Academy of Sciences
China

ABSTRACT: *A Tea is the most popular non-alcoholic beverage in the world. Asia accounts for 85% of the global tea production. Comprehensive studies on the genetic diversity and origin of domestication of tea plant are scarce, while the studies carried out to date also used fewer number of markers narrowing down its scope. Genotyping by Sequensing (GBS) is a novel next generation sequensing technique which generates large amounts of Single Nucleotide Polymorphysms (SNPs) that are vital for modern genetic analysis. Therefore, 114 wild, landraces and cultivated tea samples collected across 14 countries in Asia were subjected to GBS analysis to study the genetic relationships and origin of domestication of tea plant in Asia. A set of 247,760 high quality SNPs were generated and used for the genetic analysis of 112 samples. Multiple analysis with SNPs revealed three independent domestication events for cultivated tea confirming the results of simple sequence repeat analysis. Most of the wild species clustered together while few species/samples clustered differently showing their gene flow with the cultivated tea or possible hybrid origins. Structure and neighbour joining tree analysis clearly showed a differential clustering of Assam tea collected from India, Sri Lanka and other South Asian countries with the Assam tea collected from China and neighbouring countries in East Asia. Future studies with the recently published tea genome possibly identify differentially selected genes/biochemical pathways during tea domestication. Based on the findings of this most comprehensive study done on tea plant to date, incorporation of Chinese Assam tea germplasm into the breeding programmes in India, Sri Lanka and other South Asian countries is recommended.*

Keywords: *Origin of domestication, genetic relationships, genotyping by sequencing, next generation sequencing, tea plant*

INTRODUCTION

Tea is the world's most popular non-alcoholic beverage (Mondal *et al.*, 2004).South and South East Asia are the most important centers of tea cultivation and account for 85% of the global production. China and India are known to be the largest tea-producing countries and

[1]Corresponding author: muditha77@hotmail.com

are recognized as domestication centers. According to the classification system proposed by Ming (2000), the cultivated tea plant is currently treated as two varieties, *i.e. C. sinensis* var. *sinensis* (China type with small leaves) and *C. sinensis* var. *assamica* (Assam type with large leaves). The Cambod type with medium size leaves (*C. assamica* subsp. *lasiocalyx*) is treated as a distinct tea type under early classification systems, but identified as hybrids recently (Meegahakumbura *et al.,* 2016). Despite the number of studies that have been carried out in the past, comprehensive studies on the genetic diversity, relationships, and domestication history of the tea plant are limited. Recently, Meegahakumbura (2016) investigated the genetic diversity, structure, and relationships of 652 tea cultivars and wild relatives from 14 Asian countries using 23 SSRs markers and their results inferred three independent domestication centers for tea plant, and Chinese, Indian Assam tea were identified as distinct genetic entities based on genomic SSRs. Genetic relationships and breeding history of the African tea germplasm (Wambulwa *et al.,* 2016) and origins of African tea germplasm have also being investigated using SSRs and cpDNA sequencing (Wambulwa *et al.,* 2017). Yet, fewer number of molecular markers used was the noteworthy limitation in these studies.
Next generation sequencing (NGS) has enabled to discover, sequence and genotype hundreds to thousands of single nucleotide polymorphisms (SNPs) in tens to hundreds of individuals (Davey and Blaxter, 2010). The SNPs are the most abundant type of molecular markers extensively used in crop genetic studies and have transformed molecular biology research from genotyping to genome typing (Luikart *et al.,* 2003). Restriction site associated DNA (RAD) sequencing (Baird *et al.,* 2008) and genotyping by sequencing (GBS; Elshire *et al.,* 2011) are the two most popular next-generation sequencing techniques. Due to low-cost and ease of sample handling in GBS (Elshire *et al.,* 2011), it has been extensively used for SNPs discovery (Sonah *et al.,* 2013; Nimmakayala *et al.,* 2014), for understanding the genetic relationships (Donato *et al.,* 2013) and for domestication history studies (Nimmakayala *et al.,* 2014).

Limited attempts had been made during the past to develop and utilize SNPs in the tea plant. Zhang *et al.* (2014) developed 818 SNPs for tea plant using ESTs. Similarly, Fang *et al.*(2014) used 60 SNPs to identify tea cultivars in China, and the results revealed two genetic entities. Ma *et al.* (2015) developed SNPs and mapped 6042 markers to obtain the most informative genetic map available for tea plant. RAD Sequencing was recently used to develop genome wide SNPs and identified the relationships of 18 cultivated and wild relatives (Yang *et al.,* 2016). However, none of these previous studies used wide sampling methodologies to investigate the genetic relationships, origins of domestication of the tea plant. Therefore, in the present study, we collected tea samples including cultivated tea, landraces and close wild relatives from different Asian countries to develop SNPs using Genotyping by Sequencing (GBS) to understand the genetic relationships and origin of domestication of the tea plant.

MATERIALS AND METHODS

Plant materials

A set of 114 tea samples (Table 1), including 24 wild relatives, belonging to 12 species of cultivated tea [*Camellia sinensis* (L.) Kuntz.] landraces and cultivars were collected from 14 Asian countries (including Taiwan Province of China) and used for high quality DNA extractions based on a modified CTAB method. Samples were arranged in boxes with proper labeling, chilled using CO_2 crystals and carried to Novogene Company (Beijing, China) for GBS analysis.

Table 1. List of tea samples used for the GBS analysis

Species	Country	No. of samples
Camellia. taliensis	China	5
C. tachangensis	China	2
C. crassicolumna	China	2
C. kwangsiensis var. *kwangsiensis*	China	2
C. grandibracteata	China	2
C. gymnogyna	China	2
C. costata	China	2
C. leptophyla	China	2
C. ptilophyla	China	2
C. fengchengensis	China	1
C. pilosperma	China	1
C. costei (out group)	China	1
C.sinensis	China	41
C.sinensis	Republic of China (Taiwan)	3
C.sinensis	India	15
C.sinensis	Sri Lanka	5
C.sinensis	Nepal	4
C.sinensis	Bangladesh	3
C.sinensis	Laos	3
C.sinensis	Vietnam	3
C.sinensis	Thailand	2
C.sinensis	Indonesia	2
C.sinensis	Cambodia	2
C.sinensis	Myanmar	2
C.sinensis	Japan	2
C.sinensis	Pakistan	2
C.sinensis	Malaysia	1
Total		**114**

Analysis of GBS data

SNP calling

In the case of data analysis, quality control, assembly to a pseudo-reference sequence, mapping on the pseudo reference genome, and SNP calling were carried out as the major steps. In quality control, sequences of each sample were sorted out with the barcode. Row reads in the fasta format were processed through a series of quality control steps. Reads with ≥10% unidentified (N) nucleotides, >50% paired sequences, >10 nucleotides aligned to the adapter and reads with *Nla*III sequence were removed. Secondly, paired-end sequence reads of one sample were linked to forming artificial sequence tags and collapsed into clusters. The reads of this sample were clustered with stacks, allowing six base mismatches and more than three support reads. Thirdly, Barrow wheeler aligner ((BWA) was used to align the clean reads of each sample against the reference genome of *Arctinidia chinensis* belonging to the same Order Ericales, as the tea genome sequence was not available in 2016. Alignment files

were converted to BAM files using SAMtools (Li *et al.,* 2009) software. When multiple reads pairs with external coordinates were detected, read pairs with the highest mapping quality was retained. Finally, the SNP calling for each sample was carried out using SAMtools.

Variant filtering

Total data analysis was done using the Linux server (Super Computer) at the Kunming Institute of Botany, Chinese Academy of Science, China. The original cleaned SNP data set in the variant call format (vcf) was further filtered for minor allele frequency (MAF) ≥0.05, missing data per sample <5% and per locus <5 similar to previous studies. The filtering was done using VCFtools v0.1.14 (Auton and Marchetta, 2015).

Population structure

Population structure analysis was carried out with 112 tea samples using Frappe v1.1 (Tang *et al.,* 2005). The input file for Frappe is a plinkped file. The vcf file was converted into the plink format using VCFtools v0.1.14. The genotype data in the ped file was then recoded using Plink v1.07 (Purcel, 2015). Frappe was run using two input files: the parameter file (MyParamSim.txt) and the recoded genotype plinkped file (simdata.ped) with 10,000 maximum number of iterations and K values of 1 to 6. The output file of Frappe consisted of the ancestry proportion estimates for each. We used Distruct v1.1 (Rosenberg, 2004) to generate plots from the Q estimates.

Neighbor joining tree

The genetic distance matrix between individuals was generated and the neighbor joining tree was constructed using TASSEL 5(Bradbury *et al.,* 2007). *Camellia costei* of *C.* sect. *Theopsis* was treated as outgroup. The NJ tree was visualized and modified with FigTree 1.4 (Rambaut, 2012).

RESULTS AND DISCUSSION

GBS analysis

Of 114, samples (Table 1) sent for GBS analysis, one sample of Chinese Assam tea failed in the GBS library preparation. A total of 32.5 million GBS tags were generated for 113 tea samples with an average 287,624 GBS tags per sample. The mean sequencing depth of tags for each sample ranged from 12-29 with an average of 19. Row data generated ranged from 0.616-2.564 billion base pairs and cleaned data set ranged from 0.615-2.564 billion base pairs with an average of 99.99% very high effective rate. High-quality data set obtained with very low error rates (0.054%) and ambiguous bases (Q20=93.42%; Q30=85.11%) were removed during the cleaning process. Around 38.16% average GC content was reported for the tea samples and close wild relatives and 80.16% of the reads were mapped to the reference genome. One more sample from China reported 98.3% heterozygous SNPs, hence it was temporarily removed from the population structure and NJ tree analysis. The GBS analysis generated 1,286,662 SNPs, and upon variant filtering for minor allele frequency and missing data, 247,760 SNPs were selected for final analysis for 112 samples.

The GBS is an effective approach to develop large sets of SNP markers for crop genetic analysis (Elshire *et al.*, 2011). High quality reads were sequenced bypaired-endilluminasequencing for the tea plant and its close relatives, and the most comprehensiveSNPs dataset available for tea plant to-date was generated. A total of 247,760 high quality SNPs were developed for the 112 samples in this study, that was higher than 15,444 used for tea plant by Yang *et al.* (2016), yet lower than 8,174,678 SNPs of African wild rice used for domestication history study (Wang *et al.*, 2014). Mapping percentage reported here in the tea plant (80.16%) was lower than that of Sorghum (85%) reference genome (Morris *et al.*, 2013). This could be attributed to mapping with a reference genome of a related wild species in the same Order Ericales.

Population structure analysis

Population structure analysis was performed for 112 tea cultivars and wild relatives. Structure results for K=2 to 6 are illustrated in Figure 1. Based on the Q values obtained, three wild samples of *C. ptilophylla* (4455 and 4458) and *C. leptophylla* (4127) were grouped with China type tea, while two samples of *C. grandibracteata* (4600 and 4602) were included into Chinese Assam type tea groups. At K = 2, all 112 tea samples from Asia were assigned into two genetic groups. One group represented China type tea (in green), while the other with Assam type and wild tea relatives (in red). China tea cultivars collected from China, Japan, Republic of China (Taiwan), Southeast Asian countries (Thailand and Vietnam, Myanmar), India, Nepal and Pakistan, reported having China type gene pools and admixture genetic compositions. With K = 3, wild tea species formed a distinct cluster (in yellow) while samples of China type tea and Assam tea from China and India formed two distinct genetic groups consistent with K = 2. When K = 4, all other tea types and species showed a similar genetic grouping to K = 3, while Indian Assam type tea samples separated and formed a fourth distinct group (in blue). Both wild species and cultivated tea types reported a varying degree of genetic admixtures. At K=5 and K=6, increase in the heterogeneity within the groups was observed yet no new groups emerged. Therefore, K = 4 is probably the best clustering solution for the current data set.

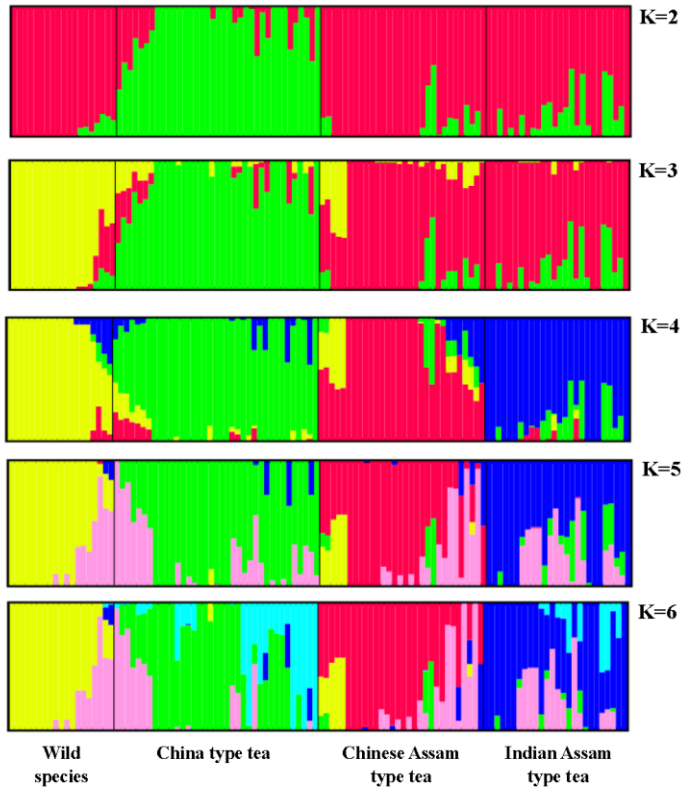

Figure 1. Results of the structure analysis for 112 tea cultivars, wild species with 247,760 SNPs

Results obtained at K = 4 offered further insights into the genetic composition of cultivated tea samples and related wild species in Asia (Figure 2). Among the wild species group, *C. taliensis, C. tachangensis, C. crassicolumna, C. kwangsiensis* and *C. gymnogyna* were the mostly pure gene pools (Figure 2: 1-15). In contrast, samples of four species namely, *C. gymnogyna*, *C. fengchengensis*, *C. costei*, and *C. pilosperma*, showed some degrees of genetic admixture of different tea types (16-19). China type tea group showed that the first seven samples (20-26) are genetic admixtures, with three samples of wild species *C. leptophylla* and *C. ptilophylla* showing similar admixture pattern with the China type landraces from Provinces of Guizhou and Hunan, China. In addition, four China type tea cultivars from India, Nepal, Pakistan and Myanmar exhibited genetic admixtures of China type and Indian Assam type tea. The first five tea samples (57-61) of Chinese Assam tea group had genetic admixtures of wild species and Chinese Assam tea. These samples were *C. grandibracteata* and hybrid samples of China tea. Chinese Assam type tea cultivars collected from Laos, Thailand, Vietnam, Taiwan and Myanmar were also found to be of admixture compositions (80-86). Among the Indian Assam type tea group, most of the samples showed genetic admixtures with China type tea.

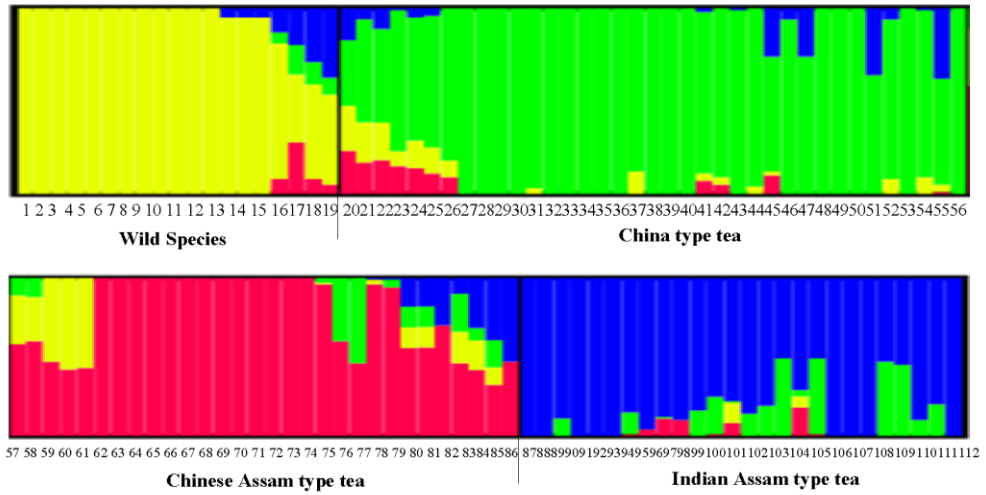

Figure 2. Magnifying result of Structure analysis at *K*=4

Interestingly, structure results at *K*=4 (Figure 1 and 2) with 247,760 SNPs was consistent with that of SSRs (Meegahakumbura, 2016). Structure results clearly defined three lineages of China type tea, Chinese Assam type tea and Indian Assam tea. Close genetic relationships between the wild species *C. leptophylla* and China type tea (*C. sinensis* var. *sinensis*) were also revealed by cpDNA sequencing. Genetic similarity of *C. grandibracteata* and hybrid landraces of Chinese Assam tea suggested a hybrid origin of *C. grandibracteata*, which might not be representing a separate "good" species but a hybrid landrace of Chinese Assam type tea. Few samples of China type tea from India, Nepal, Pakistan and Myanmar were found to be hybrids between China type tea and Indian Assam tea. Similarly, Indian Assam type tea was also found to have a higher number of cultivars with China type tea gene pool. Thus, China type tea may be a valuable tea germplasm for tea breeding in South Asian countries (Raina *et al.,* 2012).

Neighbor joining tree

The neighbor-joining tree (NJ tree) constructed for 112 tea cultivars, landraces and close wild relatives is illustrated in Figure3. Results indicated that all 112 tea samples could be broadly divided into four main clades namely, wild relatives, Chinese Assam type tea, Indian Assam type tea and China type tea. *Camellia fechengensis* was a sister to the whole cultivated tea clade. Interestingly, *C. leptophylla*was clustered with two landraces of Chinese Assam type tea hybrid from Laos and Thailand, suggesting that these samples have been misidentified. Hybrid landraces of China type tea from Guizhou, Hunan and two cultivars from Yunnan formed a subclade with *C. ptiliphyla*. *Camellia grandibracteata* and hybrid samples of Chinese Assam type tea fall into the clade of Chinese Assam tea, suggesting the genetic similarity of *C. grandibracteata* to Chinese Assam type teas.

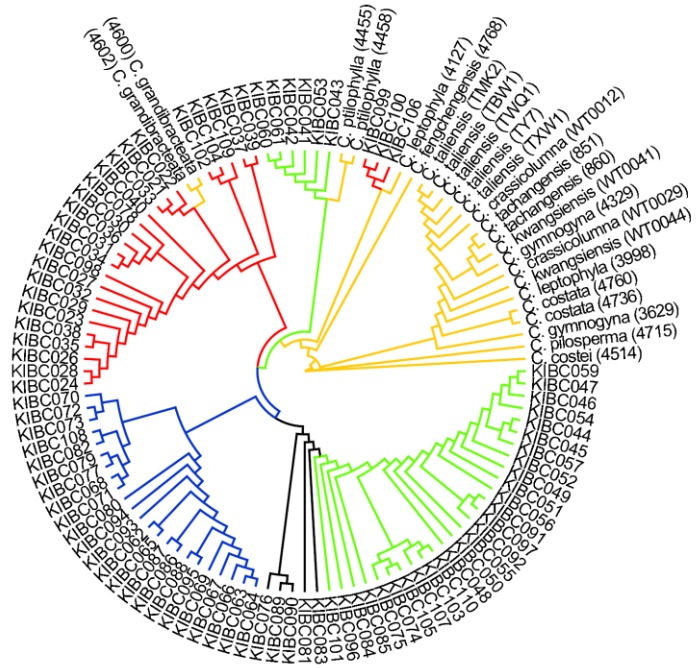

Figure 3. Neighbor joining tree constructed for 112 tea cultivars and related wild species collected from 15 countries in Asia

Dark Yellow - wild species; Red - Chinese Assam type tea; Blue - Indian Assam type tea; Black - Hybrids; Green - China type tea

Four main clades in the NJ tree generated with GBS data were similar to the results based on nSSRs data. Cultivated tea were clustered into three clades, corresponding with China type tea, Chinese Assam type tea and Indian Assam type tea, which was consistent with those of nSSR further confirming the three independent domestication centers of the tea plant (Meegahakumbura *et al.*, 2016). Yang *et al.* (2016) earlier reported differential clustering of China type tea (*C. sinensis* var. *sinensis*), Assam type tea (*C. sinensis* var. *assamica*) and three wild species using RADseqdata, which is consistent with our results of this study. Interestingly, landraces of China type tea collected from Guizhou, Hunan and two cultivars of Yunnan clustered with *C. ptilophylla* showing a genetic admixture, which differed from the China type tea cultivated in China and other Asian countries. Thus, *C. ptilophylla* may have contributed to the domestication of China type tea, and these landraces possibly selected from wild species *C. ptilophylla* or hybridization between *C. ptilophylla* and China type tea, subsequently isolated genetically from the common China type tea. Population level sampling of wild species and landraces in future needed to confirm these observations. Two samples of *C. grandibracteata* were grouped together with landraces of Chinese Assam type tea due to similarity in the structure results, which indicated *C. grandibracteata* was possibly of hybrid origin.

Structure and NJ tree analysis clearly showed the differential clustering of Assam type tea collected from India, Sri Lanka, Nepal, Bangladesh (South Asia) with the Assam type tea collected from China and neighboring countries (Laos, Myanmar, Thailand). Similar results

were revealed earlier with genomic SSRs (Meegahakumbura, 2016; Meegahakumbura *et al.,* 2016).

In future, detail analysis of SNPs of wild species and three cultivated tea types from China and India shall be used in the identification of domestication genes/genomic regions in the tea plant, similar to previous studies in other crops (Zhou *et al.,* 2015; Han *et al.,* 2016).The tea genome sequenced recently by Xia *et al.* (2017) will give a unique opportunity to study differentially selected genes/biochemical pathways in different tea types during their domestication histories. This study made a positive step towards the molecular breeding of the tea plant in future.

CONCLUSIONS

The NGS-based GBS technique was performed for the first time for tea and the most comprehensive SNP dataset was generated for 112 tea samples belonging to 63 cultivars, 25 landraces and 24 close wild relatives across 14 countries in Asia. The current analysis used 247,760 SNPs, the highest number employed for any genetic relationship study of tea to infer the genetic relationships, and origin of domestication. Three distinct genetic groups were defined for the cultivated tea confirming three independent domestications. With the results of the present study, we reconfirm the genetic differentiation between Indian and Chinese Assam type tea gene pools. We recommend the incorporationof the genetically distinct Chinese Assam type tea germplasm to the tea breeding programmes in India, Sri Lanka and other South Asian countries.

ACKWNOLEDGEMENTS

Authors wish to thank National Natural Science Foundation of China for providing the funding, and Tea Research Institute of Yunnan-China, Tea Board of India, Tea Research Institute of Sri Lanka and all other institutes and research personels for providing the tea leaf samples.Special thanks are due to Prof. D.K.N.G. Pushpakumara, Dean, Faculty of Agriculture, University of Peradeniya for his kind support through the “World Agroforestry Centre”to get this PhD scholarship.

REFERENCES

Auton, A. and Marchetta, A. (2015). VCFtools [on line]. [Accessed on 15.02.2016]. Available at https://vcftools.github.io/man_latest.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, *3*(10), p.e3376.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, *23*(19), 2633-2635.

Davey, J.W. and Blaxter, M.L. (2010). RADSeq: next-generation population genetics. Brief Funct. Genom. *9*(5), 416-423.

De Donato, M., Peters, S. O., Mitchell, S.E., Hussain, T. and Imumorin, I.G. (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS one, *8*(5), e62137.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, *6*(5), e19379.

Fang, W.P., Meinhardt, L.W., Tan, H.W., Zhou, L., Mischke, S. and Zhang, D. (2014). Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Hort. Res. *1*, 14035.

Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D.A., Yang, Z. and Zhang, Z. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytol. *209*(2), 871-884.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, *25*(16), 2078-2079.

Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. *4*(12), 981.

Ma, J.Q., Huang, L., Ma, C.L., Jin, J.Q., Li, C.F., Wang, R. K. and Chen, L. (2015). Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). PloS one, *10*(6), e0128798.

Meegahakumbura, M.K. (2016). Genetic assessment of Asian tea germplasm and the domestication history of the tea plant (*Camellia sinensis*). PhD thesis, University of Chinese Academy of Sciences, Beijing, China DOI: 10.13140/RG.2.2.21081.93282.

Meegahakumbura, M.K., Wambulwa, M.C., Thapa, K.K., Li, M.M., Möller, M., Xu, J.C. and Li, D.Z. (2016). Indications for three independent domestication events for the 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*(5), e0155369.

Ming, T.L. (2000). Monograph of the genus *Camellia*. Kunming: Yunnan Science and Technology Press. Kunming, China, 128-134.

Mondal, T.K., Bhattacharya A., Laxmikumaran, M. and Ahuja, P.S. (2004). Recent advances in tea (*Camellia sinensis*) biotechnology. Plant Cell Tiss. Organ. Cult. *76*, 195-254.

Morris, G.P., Ramu, P., Deshpande, S.P., Hash, C.T., Shah, T., Upadhyaya, H.D. and Harriman, J. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences *110*(2), 453-458.

Nimmakayala, P., Levi, A., Abburi, L., Abburi, V.L., Tomason, Y.R., Saminathan, T. and Mitchell, S.E. (2014). Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. BMC Genomics, *15*(1), 767.

Purcel, S. (2015). Plink.v1.90b3 [on line]. [Accessed on 20.02.2016]. Available at https://www.cog-genomics.org/plink2.

Raina, S.N., Ahuja, P.S. and Sharma, R.K. (2012). Genetic structure and diversity of India hybrid tea. Genet. Resou. Crop. Evol. *59*(7), 1527-1541.

Rambaut, A. (2012).FigTree version 1.4 [on line]. [Accessed on 22.02.2016]. Available at http://tree.bio.ed.ac.uk/software/figtree/

Rosenberg, N.A. (2004). Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes, *4*, 137–138.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B. and Belzile, F. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PloS one, *8*(1), e54603.

Tang, H., Peng, J., Wang, P. and Risch, N. (2005). Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. *28*, 289-301.

Wambulwa, M.C., Meegahakumbura, M.K., Kamunya, S., Muchugi, A., Möller, M., Liu, J. and Gao, L.M. (2016). Insights into the genetic relationships and breeding patterns of the African tea germplasm based on nSSR markers and cpDNA sequences. Front. Plant. Sci. *7*, 1244.

Wambulwa, M.C., Meegahakumbura, M.K., Kamunya, S., Muchugi, A., Möller, M., Liu, J. and Gao, L.M. (2017). Multiple origins and a narrow genepool characterise the African tea germplasm: concordant patterns revealed by nuclear and plastid DNA markers. Sci. Rep. *7*(1), 4053.

Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L. and Cossu, R.M. (2014). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. Nat. Genet. *46*(9), 982.

Xia, E.H., Zhang, H.B., Sheng, J., Li, K., Zhang, Q. J., Kim, C. and Huang, H. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol. Plant. *10*(6), 866-877.

Yang, H., Wei, C.L., Liu, H.W., Wu, J.L., Li, Z. G., Zhang, L. and Zhang, Z.Z. (2016). Genetic divergence between *Camellia sinensis* and its wild relatives revealed via genome-wide SNPs from RAD sequencing. PloS one, *11*(3), e0151424.

Zhang, C.C., Wang, L.Y., Wei, K. and Cheng, H. (2014).Development and characterization of single nucleotide polymorphism markers in *Camellia sinensis* (Theaceae). Genet. Mol. Biol. Res. *13*(3), 5822–5831.

Zhou, L., Wang, S.B., Jian, J., Geng, Q.C., Wen, J., Song, Q. and Zhang, J. (2015). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. Sci. Rep. *5*, 9350.