A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Construction of a Genetic Linkage Map Based on RAPD, AFLP, and SSR Markers in China Type Tea Plants

소엽종 차나무에 있어서 RAPD, AFLP 및 SSR 표지인자 기반의 유전적 연관지도 작성

Yali Chang

August, 2017

DEPARTMENT OF HORTICULTURE SCIENCES

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

Construction of a Genetic Linkage Map Based on RAPD, AFLP, and SSR Markers in China Type Tea Plants

Yali Chang

(Supervised by Professor **Kwan Jeong Song**, Ph.D)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Agriculture

June, 2017

Yali Chang, a candidate for the degree of Doctor of Philosophy in Agriculture is approved by the committee members

Chairman _____
Hoon Kang, Ph.D.

Vice-Chairman _____
Kwan Jeong Song, Ph.D.

Member _____
Ho Bang Kim, Ph.D.

Member _____
Seong Cheol Kim, Ph.D

Member _____
Yong-Hwan Jung, Ph.D.

Construction of a Genetic Linkage Map Based on RAPD, AFLP, and SSR Markers in China Type Tea Plants

Yali Chang

(Supervised by Professor Kwan Jeong Song, Ph.D)

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Agriculture

2017.8

DEPARTMENT OF HORTICULTURE SCIENCES

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

ABSTRACT

According to the double pseudo-testcross theory, a set of 79 F1 seedlings crossed from two China type tea plants (*Camellia sinensis* var. *sinensis*), with a Japanese tea plant cultivar 'Fushun' as female parent and a Korean tea plant cultivar 'Kemsull' as male parent, were used to construct the preliminary genetic linkage maps with RAPD and SSR markers for two tea plant cultivars. Firstly, a total of 660 Operon decamer random primers and 60 public SSR primer pairs were screened against two parents and six F1 offspring, and 41 decamer random primers and 11 SSR loci were selected for next genotyping analysis, respectively. Two preliminary genetic linkage maps for tea cultivars 'Fushun' and 'Kemsull' were constructed based on these markers. The map of 'Fushun' included 37 loci evenly spread into eight linkage groups (LGs), covered a total length of 255.7 cM and with an average distance of 6.9 cM between two adjacent markers. The map of 'Kemsull' located 30 loci that evenly distributed into five LGs, spanned a length of 196.1 cM and with a map density of 6.5 cM. The collinearities of LGs of two maps were unambiguously identified by shared markers between two parents, such as f1 to k2, f3 to k3, f4 to k1, f5 to k4, and f8 to k5.

Using same mapping population, two genetic linkage maps for 'Fushun' and 'Kemsull' were separately developed with AFLP markers on the basis of double pseudo-testcross theory. For the AFLP study, 2,439 bands were obtained from 27 primer combinations and with an average number of 90.3 per each. For 495 AFLP markers, 400 (80.8%) with Mendelian segregation ratios ($p < 0.01$) were found, of which 136 (34.0%) were 3:1 segregation ratio and 264 (66.0%) were 1:1 segregation ratio. The map of 'Fushun' consisted of 92 loci that distributed into eight LGs, covered a total length of 470.8 cM and with an average distance of 5.2 cM between two adjacent markers; the map for 'Kemsull' included 17 LGs and totally located 152 loci, owned a map length of 1,053.1 cM with a mean interval of 6.9 cM among two neighbouring markers. Clear collinearities of six LGs of two maps were clearly identified by shared markers which were heterozygous at both parents, such as f1 to k1, f2 to k3, f3 to k6, f4 to k5, f6 to k2, and f7 to k14. After combining 143 RAPD markers and 11 SSR loci with 495 AFLP markers as one data set, an integrated genetic map was constructed using this strategy of 'One-step method' for tea plant by JoinMap 4.0. This developed genetic map contained 295 loci (76 RAPD, 5 public SSR, and 214 AFLP markers) which evenly distributed into 15 LGs, covered a total length of 1,376.9 cM with an average interval between two adjacent markers of 4.7 cM, in which eight candidate segregation

distorted regions were identified which may contain couple genes causing segregation distortion.

Using high throughput RNA sequence technique, more informative markers were also developed and then attempted to incorporate into the moderately saturated map previously developed. We analyzed and compared the transcriptome sequences of flowers and leaves between ‘Fushun’ and ‘Kemsull’. Afterwards, 1,800 potential polymorphic SSR markers were successfully mined and 296 of them were selected and experimentally validated with a subset of tea plants (including two parents and six F1 offspring), in which 75 (25.3%) could repeatably produce anticipated bands and also be polymorphic between two parents. Finally, 29 (38.7%) newly mined SSR markers that were heterozygous in ‘Fushun’ and /or ‘Kemsull’ and showed segregated genotypes in F1 seedlings and adoptable by JoinMap, were attempted to incorporate into the existing combined genetic linkage map. Finally, 11 of them were successfully merged into this map. The new genetic map included 79 RAPDs, 5 public SSRs and 11 newly developed SSRs, and 214 AFLPs, covered 1,441.6 cM and with the average distance of 4.7 cM between two adjacent markers, which will lay a foundation for qualitative or quantitative trait loci (QTLs) analysis of important agronomic traits for tea plant in the future study.

CONTENTS

ABSTRACT.....i
CONTENTS..... iii
LIST OF TABLES.....v
LIST OF FIGURES..... vi

INTRODUCTION..... 1
Literature Cited..... 3

LITERATURE REVIEW..... 5
Plant Genome Mapping..... 5
Common Molecular Markers for Genetic Map Construction.....12
Development of Genetic Linkage Map in Tea Plant..... 20
Other Applications of Molecular Marker in Tea Plant..... 26
Literature Cited..... 29

CHAPTER Ⅰ. CONSTRUCTION OF PRELIMINARY GENETIC LINKAGE MAP WITH RAPD AND SSR MARKERS FOR CHINA TYPE TEA PLANTS.....37
Introduction.....38
Materials and Methods.....40
Results and Discussion..... 45
Literature Cited..... 55

CHAPTER Ⅱ. CONSTRUCTION OF GENETIC LINKAGE MAPS WITH RAPD, SSR, AND AFLP MARKERS FOR CHINA TYPE TEA PLANTS..... 58
Introduction.....59
Materials and Methods.....61
Results and Discussion..... 65
Literature Cited..... 89

CHAPTER Ⅲ. DEVELOPMENT OF RNA-SEQ BASED SSR MARKERS AND INCORPORATION INTO A GENETIC LINKAGE MAP WITH RAPD, SSR, AND AFLP MARKERS FOR CHINA TYPE TEA PLANTS..... 93

Introduction.....94
Materials and Methods.....96
Results and Discussion..... 100
Literature Cited..... 117

ABSTRACT IN KOREAN.....119
ACKNOWLEDGEMENTS.....122

LIST OF TABLES

Table 1. A list of some reported linkage maps in tea plant worldwide.....24

Table 2. PCR amplification of 41 primers selected from 660 random decamer primers.....47

Table 3. Genotypes of polymorphic SSR loci between 'Fushun' and 'Kemsull'.....50

Table 4. Sequences and codes of primers used for PCR amplification in AFLP analysis.....64

Table 5. Primer combinations and their codes for selective amplification of PCR.....66

Table 6. Distribution of AFLP polymorphic markers in F1 progenies.....68

Table 7. Distribution of AFLP Mendelian segregation markers in F1 progenies.....69

Table 8. Polymorphism AFLP markers obtained from PCR amplifications of 27 PCs.....71

Table 9. Summarized data of each linkage group of genetic map for 'Fushun'.....74

Table 10. Summarized data of each linkage group of genetic map for 'Kemsull'.....77

Table 11. Summarized data of each linkage group of three linkage maps.....87

Table 12. RNA-seq based SSR markers being polymorphic between 'Fushun' and 'Kemsull'.102

Table 13. Genic SSR markers developed by RNA-seq and BLASTX hit results in The Arabidosis Information Resource (TAIR) database.....106

Table 14. Summarized data of each linkage group of three linkage maps.....115

LIST OF FIGURES

Fig. 1. AFLP flowchart. Adaptor DNA= short double strand DNA molecules, 18-20 in length, representing a mixture of two types of molecules. Each type is comparable with one restriction enzyme generated DNA end. Pre-amplifications uses selective primers, which contain an adaptor DNA sequence plus one or two random bases at the 3' end for reading into the genomic fragments. Primers for re-amplification (selective amplification) have the pre-amplification primer sequence plus one or two more bases at the 3'ends. A tag (*) is attached at the 5' end of one of the re-amplification primers for detecting amplified molecules (Xu, 2009)..... 18

Fig. 2. Characteristics of tea cultivars of 'Fushun' (*Camellia sinensis*) and 'Kemsull' (*C. sinensis*). a and b, 'Kemsull' planted in the field; c and d, 'Fushun' planted in the filed; e, the mature leaf of new shoot; f, two leaves and a bud..... 41

Fig. 3. Electrophoresis results of DNA samples of 'Fushun', 'Kemsull', and all F1 offspring. f, 'Fushun' (*Camellia sinensis*), female parent; k, 'Kemsull' (*C. sinensis*), male parent..... 46

Fig. 4. Agarose profiles of amplicons of PCR with five random decamer oligonucleotide primers from two parents 'Fushun' (lane 2) and 'Kemsull' (lane 3) and 21 F1 hybrids (lane 4-24) that were seedlings from No. 1 to No. 24. DNA samples were successively amplified with OPB-01, OPC-19, OPC-20, OPG-15, and OPV-17 from up to down. All amplified products were resolved in a 1% agarose gel and electrophoresed with 0.5 × TBE buffer at 100 V for 1.5 h. Lane 1 was DNA ladder marker 1000..... 48

Fig. 5. Agarose profiles of amplicons of PCR with five SSR primer pairs for two parents 'Fushun' (lane 2) and 'Kemsull' (lane 3) and 21 F1 hybrids (lane 4-24) that were seedlings from No. 1 to No. 24. DNA samples were successively amplified from five pairs primer of MSG0232, MSG0240, MSE0316, MSG0327, and MSG0811 from up to down. All amplified products were resolved in 3% agarose gel and electrophoresed with 0.5 × TBE buffer at 100 V for 1.5 h. Lane 1 was DNA ladder marker 1000..... 51

Fig. 6. A preliminary genetic linkage map based on RAPD and SSR markers for Japanese tea cultivar 'Fushun'. Loci names were given to the right of each linkage group; genetic distance (cM) were to the left. Markers shared between two parents were postfixed with "s". Markers with distorted segregation ratios were marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001)..... 53

Fig 7. A preliminary genetic linkage map based on RAPD and SSR markers for Korean tea cultivar ‘Kemsull’. Notes were same as Fig. 6.....54

Fig. 8. PCR amplification results of two parents and six offspring with eleven primer combinations (E-AT/M-NNN). Six PCs including E2/M2, E2/M3, E2/M5, E2/M6, E2/M7, and E2/M11 were selected to conduct the following genotyping study for all F1 seedlings..67

Fig. 9. AFLP fingerprinting patterns of parents and F1 offspring amplified by the primer combination of E3/M9. Notes: lane 79, ‘Fushun’ (F), female parent; lane 78, ‘Kemsull’ (K), male parent; Lane 80, DNA marker; lanes 1-76, 76 F1 individuals, from No. 1 to No. 79. Three seedlings including No.13, No.22, and No.23 were dead.....72

Fig. 10. The genetic linkage map of female parent ‘Fushun’ with AFLP markers. Loci names were given to the right of each linkage group; genetic distances (cM) were to the left. Markers with suffix “s” were heterozygous at both parents. Markers with distorted segregation ratios were marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001).....75

Fig. 11. The genetic linkage map of male parent ‘Kemsull’ with AFLP markers. Labels and notes were same as Fig. 9.....78

Fig. 12. Genetic linkage map based on RAPD, SSR and AFLP markers for tea plant. LGs were numbered from k1-c1-f1 to k15-c15-f15. The map in the middle (c1-c15) is the consensus map constructed from the combined dataset. The genetic linkage maps on the left (k1-k15) and on the right (f1-f15) are the paternal (Kemsull) and maternal (Fushun) maps, respectively. Loci names were given to the right of each linkage group; genetic distances (cM) were to the left. Regions of linkage groups considered to be candidate segregation distortion region (SDR) were denoted by black fill. Markers with suffix ‘f’ or ‘m’ were just heterozygous at female parent or male parent, respectively. Marker with suffix ‘s’ belonged to shared markers of two parents, which was used for denoting collinearity and homology among three maps. Markers with distorted segregation ratios are marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001).....83

Fig. 13. Electrophoresis results of RNA samples of leaves and flowers of ‘Fushun’ (*Camellia sinensis*) and ‘Kemsull’ (*C. sinensis*). Notes: a, lane 2 to lane 7 were RNA samples of leaves of ‘Fushun’, lane 8 to lane 15 were RNA samples of flowers of ‘Fushun’; b, lane 2 to lane 7 were RNA samples from leaves of ‘Kemsull’, lane 8 to lane 15 were RNA samples of flowers of ‘Kemsull’; lane 1 was DNA ladder marker 100..... 97

Fig. 14. 8% denatured poly-acrylamide gel profiles of PCR amplification of five SSR primer pairs for the whole mapping population.....108

Fig. 15. Three-lined genetic linkage maps based on RAPD, AFLP, and SSR markers for tea plant. The arrangement of three maps and significance of all symbols and marks were same as Fig. 12..... 111

INTRODUCTION

Tea plant [*Camellia sinensis* (L.) O. Kuntze] is one of the most widely consumed non-alcoholic beverage crops throughout the world. It is a very important commercial crop and contributes a mass of job opportunities and revenue for developing countries including China, India, Kenya, Sri Lanka, and Vietnam (Chen et al., 2012). The whole harvest area of tea plant is 4.2 million ha. with the production of approximately 4.9 million metric tons in the world (ITC, 2013). Tea plant originated in the southwestern region of China and belongs to the family Theaceace, genus *Camellia*, section Thea, which could be mainly divided into China type (*C. sinensis* var. *sinensis*) and Assam type (*C. sinensis* var. *assamica*) according to its morphological, biochemical, and physiological characteristics (Chen et al., 2012).

During recent decades, tea has attracted attention from scientists and consumers due to its health-promotion activity and protective effect against chronic diseases such as cancer, cardiovascular, hepatic cirrhosis, and neurological disease (Mainaak et al., 2016). Therefore, there has been an imperious demand for new tea cultivars improved not only in the content of important secondary metabolites such as catechins and theanine but also in the resistance to abiotic and biotic stress or other attractive agronomic characteristics. The genetic study and breeding of this species have been impeded by highly heterogeneous genome resulted from self-incompatible reproductive system and long generation cycle. The release of a new tea cultivar by conventional breeding system usually takes more than 20 years (Jiang et al., 2009). However, if breeding materials could be screened by an employment of molecular markers closely linked to desirable agronomic traits at an early developmental stage before or after planted in the field, it will significantly shorten breeding period and save the investment of time and cost. Therefore, the marker-assisted selection (MAS) has been developed as a very efficient strategy for genetic improvement of crops, which combined molecular markers with conventional breeding technology (Collard et al., 2005). A saturated linkage map is the foundation for MAS for crop improvement and will be a powerful tool for the mapping analysis of qualitative or quantitative trait loci (QTLs) and other genome studies. A number of QTLs for important agronomic traits have been detected and characterized for perennial woody plants not only in fruit trees such as apple, citrus, and pear (Weber et al., 2003; Kunihisa et al., 2014; Yamamoto et al., 2014), but also in tea plant (Ma et al., 2014).

To date, more than ten genetic linkage maps have been developed for diverse tea plants with different molecular marker types in the F1 or BC1 segregation populations. They have been mostly constructed using the F1 progenies according to the double pseudo-testcross theory (Tanaka, 1996; Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2005; Huang

et al., 2006; Mewan et al., 2007; Taniguchi et al., 2007, 2012; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015). Among those applied molecular markers, randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) belong to the dominant molecular markers. These two types of molecular markers have been preferentially appropriate approaches for producing a large scale of reliable markers and increasing the density of genetic map quickly to discover specific genes/QTLs. Especially the AFLP marker technique has been considered as the most efficient approach since it can generate a mass of reproducible markers with a relatively low input of labor and time (Vos et al., 1995; Huang et al., 2005). Simple sequence repeat (SSR), also known as microsatellite marker, which is co-dominant and highly reproducible and locus-specific, could be used as anchors for mapping analysis and shared between different laboratories and institutes. Another marker of single nucleotide polymorphism (SNP) has been regarded as an preferably alternative for the construction of high-density linkage map due to its extreme abundance throughout the genome. The first genetic map for tea plant based on a great number of SNPs was constructed by Ma et al. (2015), in which the average interval between two neighboring markers reached to 1.6 cM.

In this study, 79 clones obtained from a cross of Japanese tea cultivar ‘Fushun’ (*C. sinensis*) and Korean tea cultivar ‘Kemsull’ (*C. sinensis*) were used to construct an integrated genetic map based on RAPD, AFLP, and SSR markers. This generated linkage map here will be greatly useful for QTL mapping analysis for tea plant in the next study.

Literature Cited

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169-196.

Chen L, Zeno A, Chen ZM (2012) Global tea breeding: achievements, challenges and perspectives. Hangzhou: Zhejiang University Press.

Hackett CA, Wachira FN, Paul S (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85(4):346-355.

Huang JA, Li JX, Huang YH, Luo JW, Gong ZH, Liu ZH (2005) Construction of AFLP molecular markers linkage map in tea plant. J Tea Sci 25 (1):7-15.

Huang FP, Liang YR, Lu JL, Chen RB (2006) Genetic mapping of first generation of backcross in tea by RAPD and ISSR markers. J Tea Sci 26(3):171-176.

Hu CY, Lee TC, Tsai HT (2013) Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). Euphytica 191:141-152.

International Tea Committee (ITC) (2013) Annual bulletin of statistics. London: ITC, P.48-53.

Jiang CJ, Liang YR, Luo JW (2009) Science of tea breeding. Beijing: China agriculture press.

Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, Kim H, Nishitani C, Terakami S et al. (2014) Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. Breed Sci 64: 240–251.

Ma JQ, Yao MZ, Ma CL, Wang XC, Jin JQ, Wang XM, Chen L (2014) Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (*Camellia sinensis*). PLoS ONE 9(3): e93131.

Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, Zheng HK, Yao MZ, 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.

Mainaak M, Tapan KM, Pradeep KC (2016) Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): a review. Plant Cell Rep 35: 255 – 287.

Mewan KM, Saha MC, Konstantin C (2007) Construction of a genomic and EST-SSR based saturated genetic linkage map of tea (*Camellia sinensis* L.). The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha

(tea) Culture and Science (ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Ota S, Tanaka J (1999) RAPD-based linkage mapping using F1 segregating population derived from crossings between tea cultivar 'Sayamakaori' and strain 'Kana-Ck17'. Breeding Res (1):16.

Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM et al. (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). PloS ONE 8(11): e81611.

Tanaka J (1996) The possibility of using RAPD analysis to create genetic map of tea. Tea Res J 84:44-45.

Taniguchi F, Tanaka J, Kono I (2007) Construction of genetic linkage map of tea using SSR markers. The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science (ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, Fukuoka H, Tanaka J (2012) Construction of a high-density reference linkage map of tea (*Camellia sinensis*). Breeding Sci 62:263-273.

Vos P, Hogers R, Bleeker M, Reijans M, de Lee TV, Homes M, Frijters A, Pot J, Peleman J et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21): 4407-4414.

Weber CA, Moore GA, Deng ZA, Gmitter Jr FG (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* × *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. J Amer Soc Hort Sci 128: 508–514.

Yamamoto T, Terakami S, Takada N, Nishio S, Onoue N, Nishitani C, Kunihisa M, Inoue E, Iwata H et al. (2014) Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). Breed Sci 64: 351–361.

LITERATURE REVIEW

Plant Genome Mapping

Physical Mapping. The essence of genome mapping is to place a collection of molecular markers onto their respective positions on the genome. According to mapping techniques, genome mapping could be divided into two main categories: genetic mapping and physical mapping. The physical map refers to a position map in which the DNA markers, genes or clones are directly located into their actual positions on the genome using molecular biological techniques. There are multiple approaches and strategies available for physical mapping, such as restriction mapping, cloned-based mapping, fluorescent *in situ* hybridization (FISH), sequence tagged site (STS), and radiation hybrid map (Yang, 2013).

Genetic Map. Genetic mapping reveals the relative positions of genetic markers on the chromosome, which is based on the recombination and exchange event of non-sister chromatids between two homologous chromosomes during the meiosis. For constructing a genetic linkage map, DNA markers or genes are required to locate into the chromosome by genetic analytical method. The likelihood of recombination event between two loci in a chromosome always depends on their physical distance. The closer they locate to each, the more likely to stay together after meiosis. The double-strand break repair model validly explain meiotic recombination event. In short, reciprocal strands break and reunion, the recombinant was obtained (Szostak et al., 1983). The relative distance between different loci could be estimated after collecting and calculating those recombinants by genetic analysis method.

Because of the uniqueness of genome structure, genetic map and physical map established for a certain organism should be consistent in their constitution. However, it is a common phenomenon that same molecular marker is located into different positions at two maps. Therefore, before carried out the assembly of genome sequence, it is absolutely essential to clarify and revise all differences between genetic map and physical map and then integrated them into a single genome map with the assistance of anchored marker. The integrate map including molecule makers from two maps and owning a higher marker density is conducive to the following genome sequencing and sequence assembly for this specific organism (Yang, 2013).

The Procedure for Constructing Genetic Map. The construction of genetic map normally includes five steps: selection of molecular marker and genotyping system; selection

of parental lines from germplasm collection which are highly polymorphism at markers loci; developing a suitable population or its derived lines with an great number of molecular markers segregated in this population; genotyping each individual/line using selected molecular markers; constructing genetic map after the linkage analysis of all genotype data.

For choosing appropriate parental lines, several factors should be taken into considerations. Firstly, the genetic polymorphism between two parents which could be determined with their geographical distribution, morphological and isozyme polymorphisms. Secondly, the parental lines should be breeding-true for autogamous plants, therefore the purification operation with further inbreeding may be required for two parents. Thirdly, the fertility of hybrid will determine whether a credible segregation population could be produced, since the cross from distantly related parents are usually accompanied with the abnormal chromosome pairing and recombination, reduced recombination frequency, and severe segregation distortion.

The common mapping materials used for genetic mapping in plant include populations of F_2, backcross 1 (BC_1), recombinant inbred lines (RILs), doubled haploid (DH), and pseudo-testcross F_1.

F_2 population has been frequently used into linkage map construction since it is easy to establish for self-pollinated or cross-pollinated plants. However, there are frequently some heterozygous genotypes in the F_2 population. And also, homozygous genotypes could not be distinguished from heterozygous genotypes by dominant markers. Therefore, a large F2 population is required to obtain more genetic information for map construction. The genetic structure of F_2 population could be easily changed after sexual reproduction. So, the vegetative reproduction and tissue culture were proposed as available techniques to prolong and extend the life of F2 population.

Only two genotypes for each locus were observed in a BC1 population, which directly reflects the segregation ratio of F1 gametes and causes BC1 to own higher mapping efficiency than other mapping materials. If two reciprocal BC1 populations can be generated from two crosses with the F1 hybrid as male or female parent, respectively, the difference of recombination frequencies between male and female gametes can be compared and analyzed. However, genetic structure of BC1 maybe be changed after further selfing. For prolonging and extending the life of BC1, two methods including vegetative reproduction and tissue culture can be used.

RILs result from continuous inbreeding such as selfing or sibmating starting from an F2

progeny until homozygosity is reached. Wu et al. (1997) supposed a minimum of 15 generations' selfings is required to obtain completely homozygous RILs for a plant with 10 chromosomes. However, researchers could not spend so much time to develop RILs in practical study. Therefore, near RILs has always been adopted in the practical studies which are established from six or seven generations' selfings started from a single F2 offspring. After six generations' selfings, the heterozygosity rate for a single locus is approximately 3% which could basically treat as homozygote. In RILs, all segregation loci owns two genotypes and with a ratio of 1:1. Therefore, the genetic structure of RILs reveals the separation percentage of F_1 gametes same as BC1. However, it is noteworthy that the recombination rate (R) obtained from linkage analysis of two loci in RILs does not equal to the recombination frequency (r) calculated in F1 gamates, because the R reveals the accumulation of recombination events of couple generations happened in the establishment process of RILs. Nevertheless, we can theoretically reckon the relationship between the R and r, and it is $R=2r/(1+2r)$. Therefore, we still estimate the recombination rate of r with RILs. As a permanent population, RILs provide a continuous supply of genetic material leading to the accumulation and of genetic information generated from different laboratories and experiments. Since inbreeding depression and barrenness, it is difficult to construct a RIL for allogamy plant (Fang et al., 2000).

Cells or plants which contain a single complete set of chromosomes are defined as haploid. Diploids produced from chromosome doubling of haploids are called doubled or double haploid (DH). The most commonly used approach for DH production is to culture the excised anther of F1 plant in vitro, induce pollen haploid plant then double chromosomes. The construction of DH plant depends on anther culture technique. Since the homozygosity of DH population, the pure lines can be obtained after its selfing. DH as another permanent population has unique advantages which make it useful in genetics and plant breeding. Since it directly reveals the segregation and combination of genes in the F1 gametes, DH has the highest mapping efficiency same as BC1. However, this technology has a selection effect on pollen with different genotypes, then destroys the genetic structure of DH population and lead to the phenomenon of serious distorted segregation, which will affect the accuracy of genetic mapping. Therefore, DH population is not an ideal mapping population for constructing of genetic linkage map.

It is extremely difficult to establish those aforementioned mapping populations for cross-pollinated woody perennials. These varieties bred from sexual propagation for the

majority of woody plants including fruit trees, forest tress, and tea plants usually show highly heterozygosity. Due to inbreeding suppression and long generation cycle, it is unlikely to develop RILs for this kind of crops. Because of high heterozygosity, it takes long time to establish BC1 for most woody perennials. For directly applying these existing F1 hybrids into gene mapping, geneticists proposed a 'double-pseudo-testcross format' based on the phenomenon of double pseudo-test cross for perennial woody plants (Hemmat et al., 1994; Grattapaglia and Sederoff, 1994).

In premise of this format is that two parents show highly heterozygous. These loci heterozygous at a parent but homozygous recessive at another will segregate with a genotype ratio of 1:1 in the F1 population. The loci heterozygous at two parents will show a segregation ratio of 3:1 in the F1 progenies. But for the loci showed homozygous dominance at any parent, there will no genotype segregation in the F1 offspring. In the first case, two separately genetic maps of parents could be constructed after severally collected genotype data from all segregation loci for each parent and analyzed with the backcross population model. In the second situation (3:1 segregating loci), these loci as shared markers could identify homologous linkage groups between two parental genetic maps. Therefore, according to this 'double-pseudo-testcross format', the F_1 progenies derived from two highly heterozygous parents could be directly treated as segregation population for map construction.

A F1 population including 56 individuals was developed from a cross of apple variety 'White Angle' and wild variety 'Rome Beauty' by Hemmat et al. (1994), then the separation conditions of DNA markers in this F1 progenies were investigated. The experimental results showed that approximately 90% molecular markers segregated in accord with the ratio of 1:1, other DNA markers' separation situation conformed to the proportion of 1:2:1 or 3:1 in the mapping population. Then two individual genetic map for 'White Angle' and 'Rome Beauty' were successfully constructed base on 'double-pseudotest-cross format', that a total of 253 loci and 256 loci were located onto tho genetic maps for two parents, respectively. In the same year, Grattapaglia and Sederoff (1994) constructed two genetic maps for *Eucalyptus grandis* Hill ex Maiden and *E. Urophylla S.T.* Blake with RAPD markers according to this format. The former covered a map length of 1,552 cM and consisted of 240 markers that distributed into 14 linkage groups, the latter owned a map length of 1,101 cM and contained 251 loci which spread into 11 linkage groups. With this same strategy, Garcia et al. (1999) also constructed two individual genetic maps for *Citrus. Volkaameriana* and *Poncirus*

Trifoliata using a 50-tree progeny as mapping materials based on 69 molecular markers, then 39 and 17 markers were included into two maps, respectively.

Nowadays, the F1 populations have been widely used for the construction of genetic maps in many perennial crops. Since the perennial characteristics of this kind of crops, more new markers could be constantly located in the genetic map between different experiments. And also, these F1 materials can be shared between different laboratories and institutes through asexual reproduction. For tea plant (*C. sinensis*), the genetic study and breeding also suffered from its highly heterogeneous genome and long generation time. Therefore, the F1 progenies could also applied for the genetic map construction of tea plant.

Analytical Approaches of Genotype Data for Linkage Mapping. We first illuminate the 'Two-point' test here. If two gene loci are located on the same chromosome and closely to each other, they will show linkage inheritance, otherwise occur genetic exchange and recombination. The 'two-point test' refers to detecting the linkage relationship between two gene loci. The recombination values are calculated based on the frequencies of different genotypes in the mapping population. The likelihood ratio was defined as the ratio of the probabilities of two gene loci with linkage relationship ($r<0.5$) to without ($r=0.5$), that the base-10 logarithm of this ratio of $L(r)/L(0.5)$ (named LOD value) has been usually used to check the reliability of the obtained recombination value. When $LOD > 3.0$, two genes could be generally considered to have linkage with each other; when $LOD<2.0$, both of genes could be usually regarded as no linkage relationship.

The two-point test is the simplest and also the most commonly used approach for linkage analysis. However, there are usually a number of markers for each chromosome during the construction of genetic map. The purpose for genetic mapping is to put those markers into right order on the chromosome and ascertain spacing distance between them. And the linkage relationships of multiple loci are required to simultaneously detect and determine. Therefore, the segregated information of these marker loci will be applied to ascertain the arrangement of them, which is named 'multi-point test'. In general case, all genetic loci are firstly separated into different linkage groups by means of two-point test, then multi-point analysis are conducted to those loci assigned into same linkage groups for determining the correct order and distance between each other. Same as two-point test, the likelihood ratio test is applied by multi-point test. Firstly, it needs to estimate the maximum likelihoods for all possible sort orders of genes; secondly, the most conceivable arrangement of genes can be determined through the likelihood ratio test. For each chromosome, the

optimal arrangement of genes and their interval distance will be ascertained after repeatedly multi-point tests. Consequently, the genetic linkage map will be constructed step by step.

With the increase of the distance of two genes, they may simultaneously happen the exchange of genetic material at different segments of this chromosome, which is named as 'double exchange'. For three linked genes, A, B, and C with r1 and r2 as single crossover frequencies between A-B and B-C, the double exchange frequency between A and C can be estimated as r1× r2 when the two single crossovers occur independently. However, the observed frequency of double crossovers is usually lower than that expected by calculating r1× r2, which means that a single crossover occurring in a particular region of chromosome will reduce the probability of a second single crossover occurring in its flanking region. This phenomenon is named as crossover interference (C). The frequency of occurrence of double exchange in a certain segment of chromosome is usually lower than theoretical value. When C=0, there is complete interference and no double crossovers, this usually means that the involved chromosome region is very short; when C=1, there is no interference occurring, indicating that the involved chromosome region is long so that two single crossovers can occur independently. Precisely since the crossover interference was taken into account, a mapping function designed by Kosambi has been widely used for calculating map distance.

Collection and Digitization of Segregated Genotypic Data. The form of DNA molecular marker genotype is usually represented as electrophoresis banding pattern. The digitization of this electrophoresis band type is the key to conduct the mathematization of the separated data of DNA molecular marker genotype. Taking RFLP marker as an example, the digitization method of the electrophoresis band type of DNA marker will be illuminated here. We suppose that two parents (P1, P2) have different bands at a RFLP loci, so the F1 hybrid will own two bands at this loci, considered as F1 band type. The individuals of F2 populations will show three band types, P1, P2, and F1. Then we arbitrarily adopt a group of numbers or symbols to record the band types of different offspring in F2 cross. For example, the band types of P1, P2, and F1 can severally correspond to 1, 2, and 3. If the band type is too ambiguous to distinguish or missing, then we can recorded it as 0. If this F2 population contains 100 individuals and 100 RFLP loci has been developed, we will get a 100 × 100 data matrix consisted of these selected simple numbers of 0, 1, 2, and 3.

When collecting and handling the genotype data of DNA molecular marker, we should pay attention to the following points: a, all uncertain data should be removed from next analysis; b, the given number for two parents should be consistent for all loci; c, if there are

multiple bands different between two parents, we should carry out the coseparation analysis to identify whether these bands belong to a single loci or not. We should record them one by one if these bands are obtained from different loci (Fang et al., 2000).

Main Computer Softwares for Constructing the Genetic Map. Nowadays, some softwares including JoinMap 4.0, Mapmarker/Exp 3.0, CRI-MAP, and MAPQTL 3.0 have been frequently used into mapping analysis. Mapmarker/Exp 3.0 has been used to construct genetic linkage map and conduct trait mapping for different segregation populations, such as BC1, F2, RILs, and F1; JoinMap 4.0 software has been mainly applied into map construction for kinds of populations including BC1, F2, RILs, DH, and F1, in which 500 loci could be simultaneously analyzed at one linkage group; CRI-MAP has been primarily used for multi-locus linkage map; MAPQTL 3.0 software can use three different methodologies including interval mapping, MQM, and non parameter analysis to conduct QTL mapping analysis for multiple populations, including BC1, F2, RILs, DH, and F1 (Ruan et al., 2002).

Application of Genetic Linkage Map in Plant. A saturated genetic map is a powerful tool for QTL mapping analysis. Numerous important agronomic traits belong to quantitative traits which are polygenic and controlled by multiple genes and easily affected by growing environment. The traditional research method in quantitative genetics can not illuminate the number of multi-genes, the contribution rate of each gene and their interactions for quantitative character, and the location of these genes at the chromosome. However, if QTL mapping analysis of quantitative character has been carried out using molecular markers, all of these issues will be solved.

The genetic linkage will provide the theoretical basis for early determining of important agronomic traits in plant and improve the precision and reliability for this assessment, which is known as marker assisted selection (MAS). For utilizing molecular marker and genetic linkage map to assist breeding, the premise is that molecular marker should be tightly linked with the important agronomic trait, then chase this gene in the study materials with the aid of this marker. Therefore, the shorter of the distance between marker and gene, the higher reliability of this selection.

Map-based cloning, also known as positional cloning, is based on the position of target gene on the chromosome. First of all, a suitable genetic segregation population should be established on the basis of this target gene existing or not, then follow the procedures: a, identifying molecular marker closely linked with target gene; b, constructing the genetic

linkage map and also physical map with high-density, then map this gene on a specific location of chromosome; c, building genome library with large fragment insert library, such as bacterial artificial chromosome (BAC) or yeast artificial chromosome (YAC); d, employing this marker as a probe, which is tightly related with target gene, to screen and identify all related clones in the established genomic library; e, obtaining large DNA fragment clone with target gene by means of chromosome walking, landing, or jumping; f, getting small DNA fragment with this gene by subcloning; g, finally ascertaining its base sequence through the genetic transformation and functional complementation assay. Based on this procedure, we can know that the high-density of linkage map is the premise and foundation for map-based cloning.

The genetic and physical maps among related species are firstly constructed using the high-conservative DNA molecular markers, then the distribution characteristics of these DNA molecular markers within different genomes were intuitively compared and the genome structures and genetic relationships of these related species were also analyzed, which is defined as 'comparative mapping'. After proceeding comparative mapping analyses, Bonierbale et al. (1988) observed that the orders of markers were highly conservative among these related species, such as tomato and potato.

Common Molecular Markers for Genetic Map Construction

As legible and special expression of genotype, genetic markers could be categorized into morphological, cytological, biochemical, and molecular (or DNA) markers. The rapid shift from the other genetic markers to DNA marker in utility is a consequence of its five inherent properties: firstly, the genotypes of molecular loci can be determined at the whole plant, organ tissue, or cellular levels without regard to the effect of surrounding environment; secondly, a relatively large number of naturally occurring alleles can be found at molecular markers; thirdly, no deleterious effects are associated with alternative allele of molecular marker; fourthly, alleles of most molecular markers are codominant, allowing all possible genotypes to be distinguished in any segregation generation; fifthly, less epistatic or pleiotropic effects are observed with molecular marker, thus a virtually limitless number of segregation markers can be monitored in a single population (Tanksley, 1983).

Nowadays, DNA markers widely used into the construction of genetic map in plant could be mainly classified into four categories. The first kind of molecular marker is based on Southern hybridization and needs the digestion treatment of DNA sample with restriction

enzyme and performing in situ hybridization such as restriction fragment length polymorphism (RFLP). The second kind is based on PCR technique that requires specific or random primer and thermostable DNA polymerase to conduct PCR amplification as with randomly amplified polymorphic DNA (RAPD), simple sequence repeat (SSR), and inter simple sequence repeat (ISSR). The third kind is based on a combination of PCR technology and restriction enzyme such as amplified fragment length polymorphism (AFLP) and cleaved amplified polymorphic sequence (CAPS). The fourth kind is based on single nucleotide polymorporphism namely single nucleotide polymorphism (SNP).

Southern Hybridization Based Molecular Marker. RFLP markers originate from DNA sequence variations including single-nucleotide mutation, insertion or deletion of small DNA fragments, and DNA rearrangements of large chromosomal fragments. RFLPs have been successfully utilized into genetic map construction in plant since early 1980s. And also, due to loci specificity, RFLP markers are suited for identifying synteny relationship between different species. Ahn and Tanksley (1993) found that RFLP markers kept similar orders in two genetic linkage maps of rice and wheat. However, this technique has been not very widely used because it is time-consuming and expensive and also demands large quantity of high quality genomic DNA. And also, the requirement of prior sequence information for designing probe makes this technique more complicated.

PCR Based Molecular Markers. The utilization of a single, random-sequence oligonucleotide primer in a low stringency PCR (35-45℃) for the simultaneous amplification of several discrete DNA fragments referred to as RAPD, AP-PCR, and DAF. These three methods differ from primer length, the stringency of PCR amplification condition, and the method of separation and detection of DNA segments. They all can be used for identifying RAPD. DAF adopts shorter primer (5 to 8 nucleotides) and produces more visible bands and reveals higher polymorphism between samples compared to RAPD which normally uses random primer with 10 nucleotides. On the contrary, AP-PCR applys longer primer (18~24 nucleotides) and higher annealing temperature which makes it more stable.

RAPD have several advantages compared with RFLP, hence have been widely used for genetic study: a, neither DNA probe nor sequence information are required to design the primer; b, the whole process of this technique does not include Southern hybridization thus making it easy, quick, and efficient; c, only small amount of DNA is required and also DNA

quality does not reach to a comparatively high level; d, because of this short primer, the competence for revealing DNA polymorphism of this technique was greatly promoted; e, development of markers is not required and the technology can be applied to virtually any organism with minimal initial development; f, the full set of RAPD primers practically cover the whole genome and have already become commercially which can be used for any species. Because of cross-specific of this marker system, it is impossible to compare and share with different populations and species. Therefore, it is generally required to transform into codominant markers such as sequence tagged site (STS), sequence characterized amplified regions (SCAR) or CAPS after cloning and sequencing RAPD products of interest, then further apply into map-based cloning (MBC), marker assisted selection (MAS) or marker assisted recurrent breeding (MARB). In the 1990s, RAPD have been prevalently used to complement those RFLP and isozyme genetic maps not only for annual crops of lettuce and kidney bean (Paran et al., 1991; Liao and Yan, 2000) but also for perennial woody plants of citrus and polar (Cai et al., 1994; Bradshaw et al., 1994).

SSR also known as microsatellite, short tandem repeats (STR), or sequence-tagged microsatellite sites (STMS), belongs to tandemly repeated units of short nucleotide motifs (one to five bp long). This kind of DNA marker occurs as interspersed repetitive elements in all eukaryotic genomes (Tautz and Renz, 1984). The universally acceptable mutation mechanism for microsatellite array is the model of ‘slipped-strand mispairing’ that it will give rise to the gain or loss of one or more repeat units if strand slippage occurs during DNA synthesis (Schlotterer and Tautz, 1992). The conservative sequence bordering SSR motif provide template for designing one pair of specific primer, which will be used to amplify the corresponding locus via PCR. After poly acrylamide gel electrophoresis and silver staining, the polymorphism of SSR loci can be observed among different individuals. On account of its codominant inheritance and highly informative, microsatellites has been broadly used to the construction of genetic maps, such as rice, wheat, maize, cotton, soybean, watermelon and so on (Rajeev et al., 2005). And furthermore, the high reproducibility of this marker makes it possible to share between different laboratories that generates consistent result, therefore cost and time for developing SSR will be greatly reduced (Saghai et al., 1994). An prevalent and easy approach for developing SSRs is to apply expressed sequence tag (EST) and other available sequence databases as sequence resource to mine SSRs using a BLAST on line. In the absence of public sequence, SSRs can be developed as the following steps: constructing a enriched small-insert clone library; screening it by hybridizing with specific

probe; sequencing positive clones; designing SSR primer pair based on the flanking regions of SSR repeats.

ISSR has been designed and developed by Zietkiewicz et al. (1994), which is biologically based on extensively widespread SSRs in eukaryotic organisms and reveals polymorphism of short DNA sequences between two contiguous SSRs. The primer for ISSR-PCR amplification usually consists of 16~18 base sequence that includes a few bases of small tandem repeat sequence. It has to be mentioned about this primer that 2~4 extra non-repeated anchoring bases are linked at its 5' or 3' end to screen those plentiful SSR loci with same motif and guarantee moderate bands obtained from PCR amplification. In recent years, most ISSR primers have been designed and developed based on di-nucleotides repeat which is most widespread SSR motif within animal and plant genome. Tsumura et al. (1996) concluded that ISSR marker based on di-nucleotides tandem repeats of (AG)n and (CT)n were most effective compared to other types in pine and cypress. ISSR markers mostly belong to dominant markers following the Mendelian inheritance rule and possess of commendable stability and assured repeatability because of the common annealing temperature (around 52°C) used for PCR amplification. It have already been widely used to cultivar identification, genetic relationship analysis, genetic diversity detection, gene mapping and genome mapping.

PCR and Restriction Enzyme Based Molecular Marker. AFLP as a great DNA fingerprint technique was invented by Zabeau and Vos (1993), which is based on the selective PCR amplification of restriction fragments from a total double-digested of genomic DNA under high stringency conditions. The AFLP technique not only owns the power of RFLP but also has the flexibility of PCR-based markers, and it provides a universal and multi-locus marker technology which could be widely applied into plant research. The primary steps are as follows (Fig. 1): the high quality of genomic DNA is usually cleaved with two restriction endonucleases; double-strand adaptors of known sequence are firstly added to the complementary and sticky ends of digested DNA segments by a ligation reaction using T4 DNA ligase; DNA amplification is carried out using specific primers which is complementary to the 5' end sequence of adaptor and cleavage site, and plus 1~3 selective base that extend into the fragment sequence and provide amplification selectivity by limiting the number of perfect sequence matches between the primer and the pool of available adaptor/DNA templates. The obtained band number of AFLP amplification simultaneously depends on adopted restriction enzymes type and number of base at the 3'

ends of primer and genome DNA for studying. This marker technique has been widely used for gene identification and expression, linkage mapping and assessing the degree of relatedness or variability among cultivars (Weng, 1996; Yong et al., 1996; Marques et al., 1998; Hayashi et al., 2001; Zewdu et al., 2001).

For generating restriction fragments with uniformly distribution at the whole genome, AFLP has always employed two restriction enzymes. One is common cutter enzyme just like *Eco*R Ⅰ which could generate a great number of small DNA segments, another is rare cutter enzyme such as *Mse* Ⅰ, which could decrease the total amount of DNA segments for the following PCR amplification. After digested with two restriction enzymes of genomic DNA template, a great number of different polymorphic DNA segments could be produced through selective-amplification of PCR with different primer combinations that just derived from a few primers. The analysis procedure of AFLP analysis generally includes three steps: firstly, after digested DNA template with selected restriction enzymes that usually contain a frequent cutter and a rare cutter, all produced DNA segments try to link with specific adaptors using T4 DNA ligase under the suitable condition; secondly, only DNA fragments that have been digested with two restriction enzymes are amplified by PCR; finally, the resulting amplicons of PCR (50~400 bp size range) are observed by silver staining followed by DNA segments separation on the denatured poly-acrylamide gel to determine polymorphisms among study objects.

The success of AFLP technique depends on completely digestion of DNA template with restriction enzyme as contaminants could inhibit its ability to entirely digest DNA sample. Most often, it takes two steps to accomplish PCR amplification for organism with large genome, which are respectively named as pre-amplification and selective-amplification. On the one hand, a good deal of templates are provided for following analysis by this two step approach. On the other hand, templates are selectively purified after twice PCR amplification, which will produce clearer DNA fingerprint and support better repeatability for this technique. For AFLP primer, the number and type of selective bases at the 3' end decide the amount of bands of AFLP amplification. The higher of G and C ratio within primer, the less products obtained from PCR amplification in AFLP analysis. In addition, the reaction recipe of PCR, especially the concentrations of Mg^{2+} and dNTP, remarkably influences its result. Hence before starting AFLP analysis, the optimization of reaction condition is necessary and required. Theoretically, AFLP assays can be carried out using relatively small DNA samples (50~500 ng per individual) compared to RFLP. It has a very high a multiplex ratio and

genotyping throughout the genome and also owns higher reproducibility than RAPD. Simple off-the-shelf technique could be used to study any organism with no or limited molecular marker development. AFLP markers belongs to classic Mendelian inheritance that can be applied to the construction of high-density molecule linkage map (Xu, 2009).

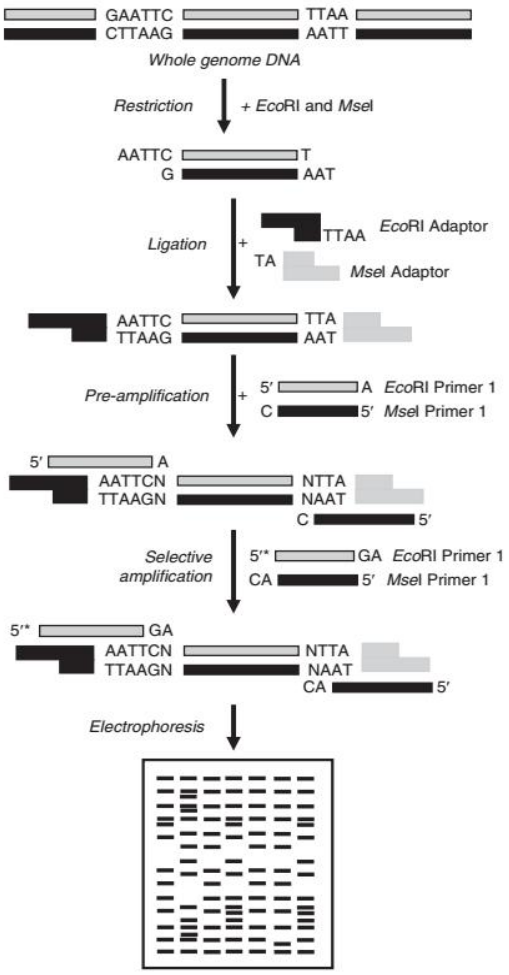


Fig. 1. AFLP flowchart. Adaptor DNA= short double strand DNA molecules, 18-20 in length, representing a mixture of two types of molecules. Each type is comparable with one restriction enzyme generated DNA end. Pre-amplifications uses selective primers, which contain an adaptor DNA sequence plus one or two random bases at the 3' end for reading into the genomic fragments. Primers for re-amplification (selective amplification) have the pre-amplification primer sequence plus one or two more bases at the 3'ends. A tag (*) is attached at the 5' end of one of the re-amplification primers for detecting amplified molecules (Xu, 2009).

CAPS is a form of genetic variation in the length of DNA fragment obtained from the restriction digestion of PCR products (Konieczny and Aububel, 1993), which is actually a remedial measure to DNA markers with specific primer. When electrophoresis bands of PCR products with specific primers do not show length variations, it can digest with restriction enzyme then detect polymorphism band patterns again by agarose-gel electrophoresis or poly acrylamide gel electrophoresis. This remedial DNA technique reveals the restriction enzyme sites variations within the generated specific PCR product and also shows length polymorphism of restriction fragments. CAPS as a useful skill have been frequently used in genetic study. Williamson et al. (1994) developed a RAPD marker named as REX-1 linked with nematode resistance gene. After transformed REX-1 into a SCAR marker, only the same band was amplified from all resistant and sensitive lines by PCR. However, after digested with restriction enzyme Taq Ⅰ, length polymorphism was observed between resistant lines and sensitive lines. Further more, the pure lines and heterozygous lines can also be discriminated.

Singe Nucleotide Polymorphism Based Marker. A SNP is an individual nucleotide base change of certain DNA nucleotide sequence between two study objects. This kind of marker generally include two categories which are transtions (C/T or G/A) and transversions (A/C, A/T, G/C or G/T), respectively. For example, two DNA sequences of 'AAGC**C**TA' and 'AAGC**T**TA' were obtained from two individuals, respectively, contain a transitional SNP (C/T). SNPs do not evenly distribute in the whole genome and most of them exists in the non-transcribed sequences. This kind of molecular marker owns some unique advantages compared with other DNA markers: a, shows high resolution power and directly reaches to single base difference; b, highly abundant in most genomes and can be automatically detected; c, its genetic stability is high; d, sibling species with close genetic relationship can be distinguished using SNP marker. Three sibling species of dragon spruce were discriminated from each other by means of a few chloroplast and nuclear SNPs (Germano and Klein, 1999). Raja et al. (2001) analyzed and compared the potential of EST-derived SNPs and SSRs for fingerprinting seven barley genotypes. There are several popular methods for detecting SNPs: a, molecular hybridization-based methodologies such as allele-specific oligonucleotide analysis and gene chip technology; b, public database-based assays, two softwares of POLYBAYES and SNP pipelines are widely used for mining SNPs; c, the conformation analysis-based inspection methods including the temperature gradient gel electrophoresis (TGGE), single strand conformation polymorphism (SSCP), denaturing

gradient gel electrophoresis (DGGE), and denaturing high performance liquid chromatography (DHPLC); d, enzymatic cleave or PCR based approaches, such as RFLP and mismatch amplification mutation assay (MAMA); e, directly sequencing methods, it is the mostly widespread adopted approach for detecting SNPs at those positioned STSs or ESTs at human genome; f, capillary electrophoresis technique (Zhou, 2005).

Development of Genetic Linkage Map in Tea Plant

The construction of genetic linkage map has been considered as a significant procedure for mapping of important agronomical important characteristics, known as QTLs mapping, also known as genetic mapping or gene tagging. Once QTLs underlying an agronomical trait of interest is fine identified, markers tags will be the most effective tools for tracing the alleles affecting this character from donor lines to new breeding materials in the crop improvement project. Therefore, the MAS that combined DNA molecular marker with conventional breeding technique was designed as a new strategy for crop breeding (Collard et al., 2005). The high-density genetic linkage map is the foundation of MAS and an powerful tool for the mapping of QTLs or other genetic researches. Several QTLs for important agronomic traits have been successfully detected and characterized for perennial woody plant not only in fruit trees such as apple and citrus (Weber et al., 2003; Kunihisa et al., 2014), but also in tea plant (Ma et al., 2014).

Up to now, more than ten partial or whole genetic linkage maps have been reported for tea plant using different molecular markers, which mainly selected F1 population as mapping materials according to double pseudo-testcross theory (Tanaka, 1996; Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2004; Huang et al., 2006; Taniguchi et al., 2007, 2012; Kamunya et al., 2010; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015). Initially, RAPD marker as a pioneer PCR-based marker system has been a preferentially appropriate approach for genetic map construction, because it can produce a large scale of reliable markers and quickly increase the density of genetic map in plant breeding. The first genetic linkage map for tea plant was just constructed with a small quantity of RAPD markers (Tanaka, 1996). Whereafter, this marker technique has been frequently adopted into genetic mapping or other genetic researches because of its unique advantages with no need specific sequence information for designing primer and easy operation. Soon afterwards, the AFLP technique as the most efficient marker approach was frequently applied to genetic study and map construction for tea plant (Vos et al., 1995; Sun et al., 2004). Hackett et al. (2000)

primarily applied AFLP markers into genetic map construction of a diploid clone SFS150 (a Malawian 'assam' type local selection of unknown pedigree). The obtained map included 106 AFLPs and 6 RAPDs distributed into 17 linkage groups (LGs), covered 1,349.7 cM and with an average distance of 11.7 cM between adjacent loci. Then, Huang (2004) constructed two high-coverage genetic maps for Chinese tea cultivars 'Qi Men No.4' and 'Chao An Da Wu Ye' (2,457.7 cM and 2,545.3 cM, respectively) based on 485 AFLP markers, which were mined from 22 primer combinations. Hu et al. (2013) constructed an integrated genetic map for tea plant based on 92 individuals produced from a cross of two heterozygous diploid varieties 'TTES 19' and 'TTES 8', in which comprised 250 AFLPs, 36 SSRs, 3 CAPSs, 1 STS, 13 ISSRs, and 64 RAPDs spread into 18 LGs, and covered 4,482. 9 cM with a map density of 12.2 cM.

SSR marker has been playing an important role for genetic study either annual or perennial plants during recent years because of its codominant nature, high reproducibility and transferability among different species (Varshney et al., 2005). A reference map for tea plant was reported by Taniguchi et al. (2012), which mainly consisted of 674 RAPDs and 441 SSRs but with a small segregation population (54 F1 hybrids of 'Sayamakaori' × 'Kana-Ck17'). Afterwards, 26 SSR loci selected from this reference map were successfully mapped into another moderately saturated genetic map using a set of F1 183 seedlings derived from two *C. sinensis* varieties with very different catechins composition as mapping materials, and homologous linkage groups of two maps were also identified through these anchor markers and presented good collinearities. In addition, nine stable QTLs associated with catechins content were identified and clustered into four main chromosome regions with LG03, LG11, LG12, and LG15 (Ma et al., 2014). Bali et al. (2015) developed a genetic linkage map using RAPDs and AFLPs for mapping drought tolerance trait for Indian tea variety.

Compared with traditional Sanger sequencing, next generation sequence (NGS)-based technologies are revolutionary techniques that can generate millions of sequence information with a relatively low cost and time (Bolger et al., 2014). For tea plant, there were two studies to develop genetic linkage maps after mined polymorphic DNA markers using NGS techniques. Tan et al. (2013) analyzed the floral transcriptome of *C. sinensis* (cv. Fudingdabaicha) using RNA seq on the platform of Illumina HiSeqTM 2000 that a total of 26.9 million clean reads were assembled into 75,531 unigenes with an average length of 402 bp, in which 10,290 (16.67%) contained one or more SSRs and finally 2,439 candidate SSR

markers were successfully mined from them. Subsequently, 720 (29.5%) were selected and experimentally checked and 431 (59.9%) SSR loci showed polymorphism between two tea cultivars ‘Longjing43’ and ‘Baihaozao’. A consensus linkage map established based on these SSR markers included 237 SSR loci evenly spread into 15 LGs and covered a total length of 1,156.9 cM with a map density of 4.9 cM. Furthermore, SNP has been acknowledged as a well-considered marker choice for the construction of high-density linkage map because of its highly abundant across the genome and available for automatic detection (Collard et al., 2005), and the fast development of NGS technologies have greatly facilitated the discovery of large scale of SNPs. The first genetic map for tea plant based on a mass of SNPs was constructed using 183 F1 offspring generated from a cross of two Chinese tea cultivars ‘Yingshuang’ and ‘Beiyue Danzhu’ by Ma et al. (2015), which contained 6,042 SNPs and covered a length of 3,965 cM with a average interval of 1.6 cM between two neighboring markers.

Association mapping, also known as linkage disequilibrium (LD) mapping, is a powerful tool for genetically dissecting complex quantitative characters, which analyzed the relationship between phenotypic variation and genetic polymorphism on the basis of linkage disequilibrium (Flint-Garcia et al., 2003). LD mapping approach replaces a biparental population required by QTL mapping that is usually time-consuming and costly, with naturally occurring populations that may cover wide genetic diversity and likelihood for higher resolution mapping because of more recombination events included. The main steps for association mapping in tea plant were presented by Bandyopadhyay (2011), including choosing individual seedlings with wide genetic diversity, checking and recording their morphological characteristics under different environments, conducting genotype analysis using appropriate DNA molecular markers, evaluating population structure for selected seedlings, correlating analysis between phenotypic and genotypic data on the basis of quantification of LD and population structure with a suitable statistical approach. Until now, two successful association-mapping studies for tea plant were reported for all we know. Yao et al. (2010) reported that five EST-SSR markers were significantly associated with the weight of 100 shoots with one bud and two leaves, leaf length, content of tea polyphenol, and content of caffeine, respectively. Jin et al. (2016) investigated the genetic relationship between tea caffeine synthase 1 (TCS1) gene and caffeine content of tea based on this mapping strategy and identified one SNP4318 significantly correlated with caffeine contents which can explain a total of 4.0%-7.0% of the phenotypic variance.

There is an increasing demand for environment friendly and pesticide free tea in the international market. DNA molecular marker have already proved its potency as genetic tool to benefit the genetic improvement of crops. But there are still lots of work to be carried out. MAS supports a opportunity for increasing efficiency in the selection of genotypes with expected agronomical traits. It is important to conduct the fine QTL mapping of desirable agronomical trait not only for tea plant but also for other crops, which is key of MAS in plant breeding. Therefore, the larger population size (more recombinational events involved), more accurate phenotypic data, different genetic backgrounds, as well as adequate validation to obtain reliable markers may guarantee its value in practice. Moreover, a new mapping strategy proposed for cross-pollinated plants by Wu et al. (2002) that the integration of the advantages of linkage mapping with LD mapping may be more effective for fine QTL mapping.

Table 1. A list of some reported linkage maps in tea plant worldwide.

Cultivar name	Mapping population type & size	No. of marker and type	No. of linkage group	Map lengths (cM)	Reference
Yabukita,	F1, 46	23 RAPDs,	6	-	Tanaka 1996
Shizu-Inzatsu131		36 RAPDs	6	-	
Sayamakari,	F1, 54	126 RAPDs,	14	1550	Ota and Tanaka 1999
Kana-CK17		140 RAPDs	17	1640	
SFS150	F1, 90	6 RAPDs, 106 AFLPs	17	1,349.7	Hackett et al. 2000
Qimen 4,	F1, 69	208 AFLPs,	17	2,457.7	Huang et al. 2004
Chaoan Dawuye		200 AFLPs	16	2,545.3	
Fuding Dabaicha	BC1, 94	46 RAPDs, 16 ISSRs	7	1,180.9	Huang et al. 2006
TR12043,	F1, 141	136 SSRs,	15	1,018	Mewan et al. 2007
TR12023		173 SSRs	15	1,192.9	
Sayamakari,	F1, 64	304 SSRs, 266 RAPDs, 1 CAPS	17	3,091	Taniguchi et al. 2007
Kana-CK17		281 SSRs, 351 RAPDs	15	3,314	
TRFCA SFS150, AHP S15/10	F1, 42	22 RAPDs, 74 AFLPs, 4 SSRs	30	1,411.5	Kamunya et al. 2010
Sayamakari	F1, 54	363 SSRs, 330 RAPDs, 7 CAPS, 1 SST	15	1,305	Taniguchi et al. 2012
Kana-CK17		354 SSRs, 348 RAPDs, 3 CAPSs, 1 SST	15	1,298	
TTES 19,	F1, 92	36 SSRs, 3 CAPSs, 1 STS,	18	4,482.9	Hu et al. 2013

TTES 8		250 AFLPs, 13 ISSRs, 64 RAPDs			
Longjing43, Baihaozao	F1, 188	237 SSRs	15	1,156.9	Tan et al. 2013
Yingshuang, Beiyue Danzhu	F1, 183	406 SSRs	15	1,143.5	Ma et al. 2014
Yingshuang, Beiyue Danzhu	F1, 183	406 SSRs, 6042 SNPs	15	3,965	Ma et al. 2015

Other Applications of Molecular Marker in Tea Plant

Genetic Diversity and Phylogenetic Relationship. The comprehensive and accurate study of genetic diversity is necessary to design a reasonable strategy for germplasm conservation programme (Bandyopadhyay, 2011). Multiple molecular markers have been frequently used for assessing genetic diversity and phylogenetic study for their own indigenous tea accessions such as RFLP, RAPD, AFLP, ISSR, SSR, and EST-SSR (Mainaak et al., 2016). Moreover, one SNP from the polyphenol oxidase coding regions of different tea varieties was identified by Huang (2004), which can successfully differentiate the processing suitabilities of all tea samples. A wonderful work have been done by Yang et al. (2016). A total of 15,444 SNPs from 18 cultivated and wild tea accessions were developed using a high-throughput genome-wide restriction site-associated DNA sequencing (RAD-Seq) technique. After carried out the phylogeny inference, principal component and genetic structural analyses for all samples based on these newly markers, six clusters corresponded to six *Camellia* species/varieties were finally obtained. In addition, *C. taliensis* var. Bangwei as a semi-wild or transient landrace was ascertained because of its phylogenetic position was located between those wild accessions and cultivated accessions.

In addition, three key genes of phenylalanine ammonia-lyase (PAL), chalcone synthase (CHS), and dihydroflavonol 4-reductase (DFR) have been applied to successfully classify all *assamica* and *sinensis* tea genotypes studied (Kaundun and Matsumoto, 2003). 5S rDNA gene has been also proven to be useful to study phylogenetic relationship for tea plant by Singh and Ahuja (2006). Katoh et al. (2003) evaluated genetic diversity of 180 tea varieties based on the nucleotide sequence polymorphisms from ribosomal RNA maturase (matK) regions of chloroplast DNA. Liu et al. (2012) studied the phylogeography of 21 natural populations of *C. taliensis* (a related species of *C. sinensis*) on the basis of sequences variations of nuclear PAL gene fragment and chloroplast rpl32-trnl intergenic space.

Core Collection. In many tea-producing countries, tea planters have pulled out the seed tea populations and displaced them with improved high yielding and quality cultivars. However, some genotypes maybe own valuable agronomical traits which are resistant to various biotic and abiotic stresses and might be lost in the present cultivated gene pool. Therefore, it is quite important to conserve these tea pants with unique characteristics from old seed tea populations to expand the germplasm pool and provide valuable materials for tea breeding program in the future. The collection, research, and utilization of tea germplasm has drawn great attention from scientists all over the world. Over than 2,500 tea accessions

were conserved at three main centers of Jorhat, Toclai Experimental Station, India (Singh, 2006). More than 2700 tea accessions had already been maintained at two nurseries of Tea Research Institute, Chinese Academy of Agricultural Science, China, in 2004 (Wang et al., 2004). For reasonable evaluation and utilization of these germplasm, core collection is required to establish from each gene repository and managed as a new working collection for increasing utilization efficiency and cutting down cost, that would be necessary to introduce marker-assisted acquisition for avoiding duplicates and enhancing accuracy. Some core collections have been successfully obtained exclusively based on DNA molecular markers for other crops (Gunasekare, 2007). Spooner et al. (2005) proposed that DNA molecular markers should supplement the characterization obtained from morphological or biochemical assessments that will supply more accurate and detailed information for constructing core collection. Before him, Wang et al. (2004) had already proposed a tentative three-step strategy for establishing a core collection from the whole tea collection of China that justly integrated phenotype and chemical data with DNA molecular markers.

Functional Genomics and Transcriptome Analysis. The transcriptome was known as the complete set of transcripts for certain type of cell or tissue in a specific developmental stage or under a particular physiological condition. Transcriptome analysis can provide a comprehensive understanding of molecular mechanisms involved in specific biological processes or diseases (Velculescu et al., 1997). For identifying differentially expressed gene in various tissues, some classical technologies have been widely adopted such as subtractive hybridization, differential screening, cDNA representative difference analysis, and mRNA differential display. In addition, serial analysis of gene expression (SAGE), cDNA microarray, DNA chip, and RNA sequence have been applied for carrying out massively analysis of gene differential expression in plant (Li and Chen, 2000). The first cDNA microarray for tea plant worldwide was developed on the basis of 1680 genes, which consisted of 6,912 dots, 6,720 ESTs, 160 positive controls, and 32 negative controls. After hybridized it with three tea clones which have different contents of tea polyphenols, the whole gene expression profiles were obtained, respectively (Zhao et al., 2006). The NGS based on RNA seq technique has strengthen whole transcriptome analysis during recent years, especially for non model perennial plants such as tea plant. Successively, Shi et al. (2011) and Wang et al. (2013) applied this new revolutionary technique to conduct functional genomic analysis of different tea tissues. The cDNA-AFLP technique, as a more inexpensive and maneuverable approach for analyzing whole genome expression profiles,

had been designed by Bachem et al. (1996), which integrated the easy operation of differential display-PCR (DD-PCR), a small quantity of mRNA required together with unique advantages of AFLP marker such as high resolution ratio, high abundance, and no requirement for sequence information, to intensively display the genome expressed sequences polymorphisms. This combinational technology has been widely used for constructing genome transcriptome map, analyzing differential expression gene, and gene cloning (Fu and Li, 2003; Escaletts et al., 2006; Henriquez and Daaf, 2010). Wu et al. (2011) established and optimized a cDNA-AFLP amplification reaction system for analyzing differentially expressed genes during white period of tea cultivar “Anji Bacha”, a temperature-sensitive rare tea germplasm. Gupta et al. (2013) identified totally 108 transcript-derived fragments which were differentially expressed in the leaves of drought tolerant tea cultivars justly based on this high efficient technique.

Literature Cited

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980-7984.

Bachem CWB, Hoeven RS, Bruijin SM, Bruijin SM, Vreugdenhil D, Zabeau M, Visser RGF (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. The Plant J 9(5):745-753.

Bandyopadhyay T (2011) Molecular marker technology in genetic improvement of tea. International J of Plant Breeding and Genetics 5(1):23-33.

Bali S, Mamgain A, Raina SN, Yadava SK, Bhat V, Das S, Pradhan AK, Goel S (2015) Construction of a genetic linkage map and mapping of drought tolerance trait in Indian beveragial tea. Mol Breed 35:112-132.

Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer K (2014) Plant genome sequencing-application for crop improvement. Current Opinion in Biotechnology 26:31-37.

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103.

Bradshaw Jr HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. Ⅲ. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89: 167-178.

Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in Citrus using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. Theoretical and Applied Genetics 89(5):606-614.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169-196.

Escalettes VSL, Hullot C, Wawrzy'nczak D, Mathieu E, Eyquard JP, Gall QL, Decroocq V (2006) *Plum pox* virus induces differential gene expression in the partially resistant stone fruit tree *Prunus armeniaca* cv. Goldrich. Gene 374:96-103.

Fang XJ, Wu WR, Tang JL (2000) DNA marker assisted selection for crop, Science press.

Fu FL, Li WC (2003) Genome transcriptome map construction with cDNA-AFLP techniques. Molecular Plant Breeding 1(4):523-529.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54:357-374.

Garcia R, Asins MJ, Forner J, Carbonell EA (1999) Genetic analysis of apomixis in *Citrus* and *Poncirus* by molecular markers. Theoretical and Applied Genetics 99(3):511-518.

Germano J, Klein AS (1999) Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. Theor Appl Genet 99: 37-49.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137.

Gunasekare MTK (2007) Applications of molecular markers to the genetic improvement of *Camellia sinensis* L. (tea) - a review. J of Horticultural Science & Biotechnology 82 (2) 161-169.

Gupta S, Bharalee R, Bhorali P, Das S, Bhagawati P, Bandyopadhyay T, Gohain B, Agarwal, Ahmed P et al. (2013) Molecular analysis of drought tolerance in tea by cDNA-AFLP based transcript profiling. Mol Biotech 53:237-248.

Hackett CA, Wachira FN, Paul S (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85(4):346-355.

Hayashi E, Kondo T, Terada K, Kuramoto N, Goto Y, Okamura M, Kawasaki H (2001) Linkage map of Japanese black pine based on AFLP and RAPD markers including markers linked to resistance against the pine needle gall midge. Theor Appl Genet 102: 871-875.

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J of Heredity 85:4-11.

Henriquez M, Daayf F (2010) Identification and cloning of differentially expressed genes involved in the interaction between potato and phytophythora infestans using a subtractive hybridization and cDNA-AFLP combinational approach. J of Integrative Plant Biology 52(5):453-467.

Huang JA (2004) Construction of molecular genetic map and studies on the SNP of polyphenol oxidase gene in tea plant. Hunan Agricultural University, Doctor dissertation.

Huang FP, Liang YR, Lu JL, Chen RB (2006) Genetic mapping of first generation of backcross in tea by RAPD and ISSR markers. J Tea Sci 26(3):171-176.

Hu CY, Lee TC, Tsai HT (2013) Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). Euphytica 191:141-152.

Jin JQ, Yao MZ, Ma CL, Ma JQ, Chen L (2016) Association mapping of caffeine content with TCS1 in tea plant and its related species. Plant Physiology and Biochemistry 105: 251-259.

Kamunya SM, Wachira FN, Pathak RS, Korir R, Sharma V, Kumar R, Bhardwaj P, Chalo R, Ahuja PS et al. (2010) Genomic mapping and testing for quantitative trait loci in tea (*Camellia sinensis* (L.) O. Kuntze). Tree Genetics & Genomes 6:915.

Katoh Y, Katoh M, Takeda Y, Omori M (2003) Genetic diversity within cultivated teas based on nucleotide sequence comparison of ribosomal RNA maturase in chloroplast DNA. Euphytica 134:287-295.

Kaundun SS, Matsumoto S (2003) Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, *Camellia sinensis* (L.) O. Kuntze, and differentiation between *assamica* and *sinensis* varieties. Theor Appl Genet 106:375-383.

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. The Plant Journal 4(2): 403-410.

Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, Kim H, Nishitani C, Terakami et al. (2014) Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. Breed Sci 64: 240–251.

Li ZY, Chen SY (2000) Progress in functional genomics in plant. Hereditas 22(1):57-60.

Liao H, Yan XL (2000) QTL mapping of root type of kidney bean under P-deficiency stress. J of Agricultural Biotechnology 8(1):61-70.

Liu Y, Yang SX, Ji PZ, Gao LZ (2012) Phylogeography of *Camellia talensis* (Theaceae) inferred from chloroplast and nuclear DNA: insights into evolutionary history and conservation. BMC Evol Biol 1:92-117.

Ma JQ, Yao MZ, Ma CL (2014) Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (*Camellia sinensis*). PLoS ONE 9(3): e93131.

Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, Zheng HK, Yao MZ, 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.

Mainaak M, Tapan KM, Pradeep KC (2016) Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): a review. Plant Cell Rep 35: 255–287.

Marques CM, Arau JA, Ferreira JG, Whetten R, Malley DM, Liu BH, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globules* and *E. tereticornis*. Theor Appl Genet 96: 727-737.

Ota S, Tanaka J (1999) RAPD-based linkage mapping using F1 segregating population derived from crossings between tea cultivar ‘Sayamakaori’ and strain ‘Kana-Ck17’. Breeding Res (1):16.

Paran I, Kesseli R, Michelmore R (1991) Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34: 1021-1027.

Raja K, Rajeev KV, Thomas T, Klaus JD (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135: 145-151.

Rajeev KV, Andreas G, Mark ES (2005) Genic microsatellite markers in plants: features and applications. TRENDS in Biotech 23(1):48-55.

Ruan CJ, Heng ZX, Qin P (2002) Research progress of plant genetic linkage map in China. J of Northwest Plants 22 (6): 1526-1536.

Saghai MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466-5470.

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20: 2211-2215.

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T, Wan XC (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics 12:131.

Singh ID (2006) Selection and testing field performance of candidate tea clones in India. Inter J Tea Sci 5:1-12.

Singh D, Ahuja PS (2006) 5S rDNA gene diversity in tea (*Camellia sinensis* (L.) O. Kuntze) and its use for variety identification. Genome 49:91-96.

Spooner D, Teruren R, Vicente MC (2005) Molecular markers for genebank management. IPGRI Technical Bulletin 10:6-73.

Sun QL, Liang YR, Ding ZT, Lu JL (2004) AFLP molecular marker technique and its application to tea genetic and breeding research. J of Tea 30(4):203-206.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33 25-35.

Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM, Liang JB (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). PloS ONE 8(11): e81611.

Tanaka J (1996) The possibility of using RAPD analysis to create genetic map of tea. Tea Res J 84:44-45.

Tanksley SD (1983) Molecular markers in plant breeding. Plant molecular biology reporter 1: 3-8.

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12(10): 4127-4138.

Taniguchi F, Tanaka J, Kono I (2007) Construction of genetic linkage map of tea using SSR markers. The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science (ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, Fukuoka H, Tanaka J (2012) Construction of a high-density reference linkage map of tea (*Camellia sinensis*). Breeding Sci 62:263-273.

Tsumura Y, Ohba K, Strauss SH (1996) Diversity and inheritance of inter-simple sequence repeat polymorphisms in Douglas-fir (*Pseudotsuga menziesii*) and sugi (*Cryptomeria japonica*). Theor Appl Genet 92: 40-45.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. TRENDS in Biotechnology 23(1):48-55.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett Jr D, Hieter P, Vogelstein B, Kinzler KW(1997) Characterization of the yeast transcriptome. Cell 88:243-251.

Vos P, Hogers R, Bleeker M, Reijans M, de Lee TV, Homes M, Frijters A, Pot J, Peleman J et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21): 4407-4414.

Wang XC, Yang YJ, Chen L, Yao MZ (2004) Establishment of core collection in tea germplasm. Subtropical Plant Science 33(4):52-56.

Wang XC, Zhao QY, Ma CL, Zhang ZH, Gao LZ, Kong YM, Yue C, Hao XY, Chen L, et al. (2013) Global transcriptome profiles of *Camellia sinensis* during cold acclimation. BMC Genomics 14:415.

Weber CA, Moore GA, Deng ZA, Gimitter Jr FG (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* × *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. J Amer Soc Hort Sci 128: 508–514.

Weng YJ (1996) AFLP - a new DNA molecule technology. Heredity 18(6): 29-31.

Williamson VM, Ho JY, Wu FF, Miller N, Kaloshian I (1994) A PCR-based marker tightly linked to the nematode resistance gene, Mi, in tomato. Theor Appl Genet 87: 757-763.

Wu RL, Ma CX, Casella G (2002) Joint linkage and linkage dis-equilibrium mapping of quantitative trait loci in natural populations. Genetics 160:779-792.

Wu WR, Li WM, Lu HR (1997) Number of selfings to produce a population of recombinant inbred lines. J of Fujian Agricultural University 26(2):129-132.

Wu Y, Deng TT, Li J, Ou QL, Huang JA (20e11) Establishment of cDNA-AFLP sliver staining technical system in tea plant (*Camellia sinensis*). Chinese Agricultural Science Bulletin 27(19):94-99.

Xu, YB (2009) Molecular plant breeding. MPG Books Group, UK, pp 28-31.

Yang, JS (2013) Genomics. Higher education press, Beijing.

Yao MZ, Qiao TT, Ma C, Jin JQ, Chen L (2010) The association analysis of phenotypic traits with EST-SSR markers in tea plant. J of Tea Sci 30(1):45-51.

Yang H, Wei CL, Liu HW, Wu JL, Li ZG, Zhang L, Jian JB, Li YY, Tai YL et al. (2016) Genetic divergence between *Camellia sinensis* and its wild relatives via genome-wide SNPs from RAD sequencing. PloS ONE 11(3):e0151424.

Yong GC, Blair MW, Panaud O, McCouch S (1996) Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained poly-acrylamide gels. Genome 39: 373-378.

Zaubeau M, Vos P (1993) Selective restriction fragment amplifications: a general method for DNA. EP534859A1.

Zewdu T, Seppo K, Kristin L (2001) AFLP fingerprinting as a tool to study the genetic diversity of *Rhizobium galegae* isolated from *Galega orientalis* and *Galega officinalis*. J of Biotechnology 91: 169-180.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176-183.

Zhou, YQ (2005). DNA molecular marker technique in plant research. Chemical Industry Press, Beijing, pp 204-205.

Zhao LP, Gao QK, Chen L, Wang XC, Yao MZ (2006) Development and preliminary application of cDNA microarray of tea plant (*Camellia sinensis*). J of Tea Sci 26(3):166-170.

CHAPTER Ⅰ. CONSTRUCTION OF PRELIMINARY GENETIC LINKAGE MAP WITH RAPD AND SSR MARKERS FOR CHINA TYPE TEA PLANTS

Abstract

According to the double pseudo-testcross theory, a set of 79 F1 seedlings crossed from two China type tea plants (*Camellia sinensis* var. *sinensis*), with a Japanese tea plant cultivar 'Fushun' as female parent and a Korean tea plant cultivar 'Kemsull' as male parent, were used to construct the preliminary genetic linkage maps with RAPD and SSR markers for two tea plant cultivars. A total of 660 decamer random primers and 60 pairs of SSR primers were screened against two parents and a few offspring, and then 41 decamer random primers and 11 SSR loci were selected for next genotyping analysis of all F1 offspring, respectively. Two preliminary genetic linkage maps for tea cultivars 'Fushun' and 'Kemsull' were constructed based on these markers. The map of 'Fushun' included 37 loci that spread into eight linkage groups (LGs), covered a total length of 255.7 cM and with an average distance of 6.9 cM between two adjacent markers. The map of 'Kemsull' located 30 loci that evenly distributed into five LGs, spanned a length of 196.1 cM and with a map density of 6.5 cM. The collinearities of LGs at two maps were unambiguously identified by common markers of two parents, such as f1 to k2, f3 to k3, f4 to k1, f5 to k4, and f8 to k5.

***Additional key works*:** *Camellia sinensis*, genetic mapping, DNA markers.

Introduction

Tea plant [*Camellia sinensis* (L.) O. Kuntze] belongs to the family Theaceace, genus *Camellia*, section Thea, and could be mainly divided into China type (*C. sinensis* var. *sinensis*) and Assam type (*C. sinensis* var. *assamica*) according to its morphological, biochemical, and physiological characteristics (Chen et al., 2012). The whole harvest region of tea plant is 4.2 million ha. with the production of approximately 4.9 million metric tons in the world (ITC, 2013). It is a very important commercial crop and contributes a mass of job opportunities and revenue for some developing countries, including China, India, Sri Lanka, and Vietnam (Chen et al., 2012).

Recently, tea has attracted attention from scientists and consumers because of its components, which have the health-promoted function and protective effect against diseases, such as cancer, cardiovascular, hepatic cirrhosis, and neurological disease (Zaveri, 2006; Basu and Lcas, 2007; Weinreb et al., 2009; Yang et al., 2009). Therefore, there has been an imperious demand for breeding new tea cultivars with the high content of certain secondary metabolite or high resistance to pest and disease. The genetic study and breeding of this species have been impeded by highly heterogeneous genome resulted from self-incompatible reproductive system and its long generation cycle. For the successful selection of a new tea cultivar by traditional breeding system, it usually takes more than 20 years including sexual hybridization, individual selection, varieties comparison test of lines, and regional comparison test (Jiang et al., 2009). However, if breeding materials could be screened by an employment of molecular markers tightly linked to desirable agronomic trait at the early development before or later planted in the field, it will significantly shorten breeding period and save the investment of time and cost. Therefore, the marker-assisted selection (MAS) has been developed as a new strategy for genetic improvement of crop, which combined molecular marker with conventional breeding technology (Collard et al., 2005). A saturated linkage map is the foundation for MAS in crop improvement and will be a powerful tool for the mapping analysis of qualitative or quantitative trait loci (QTLs) and other genome studies. A number of QTLs for important agronomic traits have been detected and characterized for perennial woody plants not only in fruit trees such as apple, citrus, and pear (Weber et al., 2003; Kunihisa et al., 2014; Yamamoto et al., 2014), but also in tea plant (Ma et al., 2014).

To date, more than ten genetic linkage maps have been developed for diverse tea varieties with different molecular marker types using F1 or BC1 progenies as mapping

populations (Tanaka, 1996; Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2005; Huang et al., 2006; Mewan et al., 2007; Taniguchi et al., 2007, 2012; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015). Among those applied molecular markers for constructing genetic map in tea plant, randomly amplified polymorphic DNA (RAPD) marker belongs to the dominant molecular marker, which has been preferentially appropriate approach for producing a large scale of reliable markers and quickly increasing the density of genetic map to discover specific genes/QTLs. Simple sequence repeat (SSR), also known as microsatellite, has been widely applied into various genetic studies in plant, because of its high abundance, locus specificity, codominant inheritance, and transferability among related species. The first reference map for tea plant was mainly composed of RAPDs and SSRs (Taniguchi et al., 2012).

In this study, two preliminary linkage maps were constructed with RAPD and SSR marker, for which 79 F1 seedlings crossed from female parent Japanese tea cultivar ‘Fushun’ and male parent Korea tea cultivar ‘Kemsull’ were selected as mapping population.

Materials and Methods

Plant Materials

A F1 segregation population consisting of 79 clones was obtained from a cross of maternal 'Fushun' and paternal 'Kemsull'. The female parent of 'Fushun' introduced from Japan was selected for its high vigor and resistance to insect and disease; the male parent of 'Kemsull' being susceptible to disease and insect pest, but having good adaptation to Korean climate and excellent flavor for traditionally handmade green tea, belongs to a Korean landrace (Fig. 1). This mapping population was planted at the field of Sulloc Cha R & D center, Jangwon Co., Ltd., Jeju, Korea. After primer screening, three F1 seedlings (No.13, No.22, and No.23) were dead during this experiment.

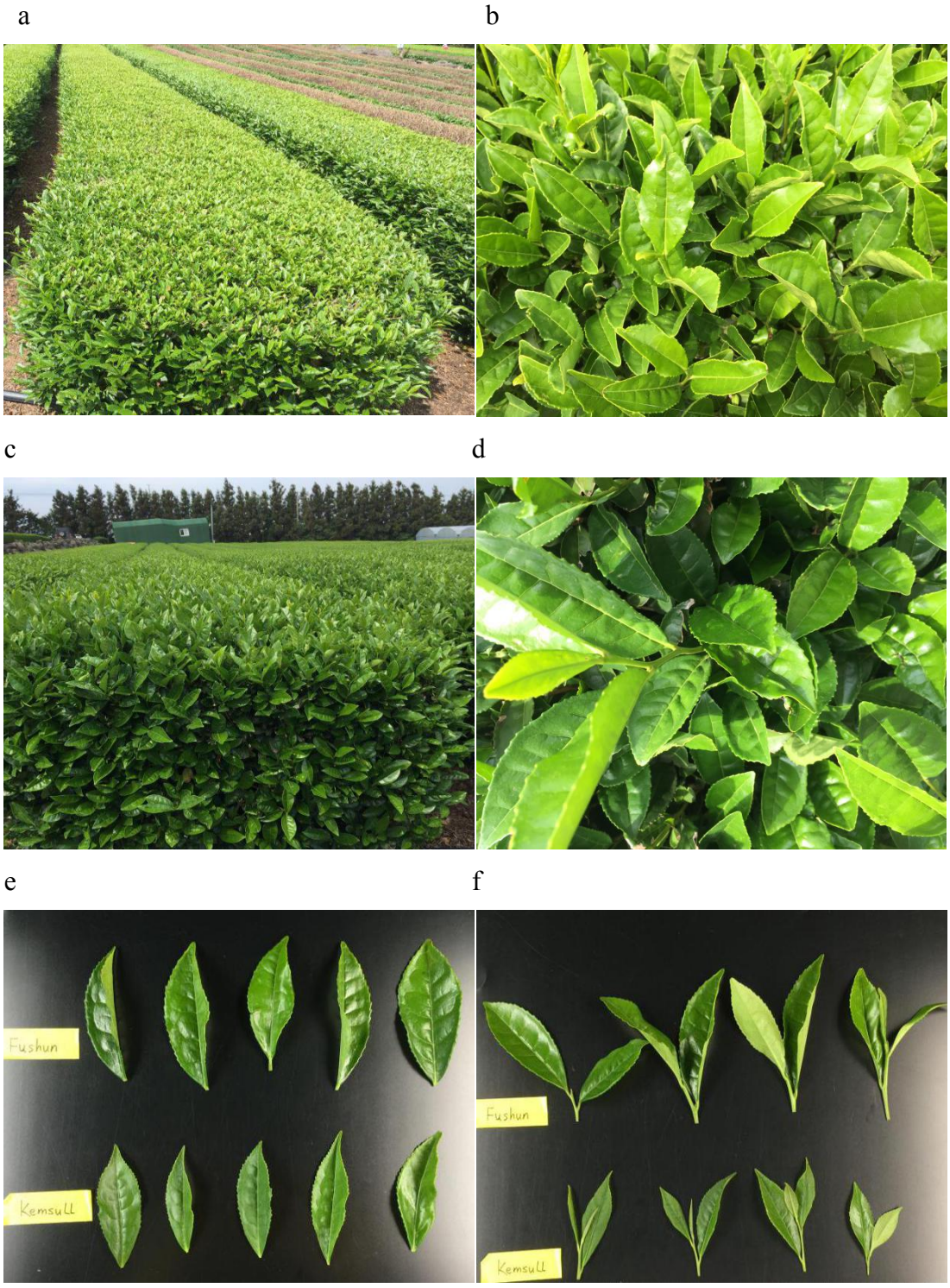


Fig. 2. Characteristics of tea cultivars of 'Fushun' (*Camellia sinensis*) and 'Kemsull' (*C. sinensis*). a and b, 'Kemsull' planted in the field; c and d, 'Fushun' planted in the filed; e, the mature leaf of new shoot; f, two leaves and a bud.

DNA Isolation

Young fresh leaves from two parents and all F1 seedlings were harvested, packed by aluminum foil and temporarily placed in the ice box. After back to laboratory, immediately sealed all samples into a zip bag and stored at 4°C. The leaf tissues were frozen in liquid nitrogen and thoroughly ground using a chilled mortar and pestle. The produced fine powder (around 100 mg) was transferred into a 1.5 mL eppendorf tube with a metal spatula. 700 μL CTAB isolation solution (98% 2 × CTAB + 2% β-mercaptoethanol, v/v) was added into the tube and softly mediated into slurry. The microtube was incubated at 65°C for one hour with gentle inversion every 30 min. Next, kept the tube under room temperature for 20 min or placed at 4°C for 5 min until the sample cooled down. An equal volume of phenol: chloroform: isoamylol (25: 24: 1) was added and mixed gently to form an emulsion, then centrifuged for 15 min at 13, 000 rpm. The resulted aqueous phase was transferred into a new tube and added an equal volume of phenol: chloroform: isoamylol (25: 24: 1), thoroughly mixed and centrifuged. This same operation was repeated with different volumes of phenol: chloroform: isoamylol (25: 24: 1). Finally, 200 μL supernatant liquor was obtained and kept into a new tube. For precipitating DNA, 200 μL chilled isoprophy alcohol was added into the tube, stored at -20°C for 30 min, then centrifuged and discarded the supernatant. The left DNA pellets precipitated at the bottom of tube was gently mixed with 500 μL of 70% ethanol, then centrifuged and discarded supernatant. 500 μL of 100% ethanol was added into the tube. After centrifuged and removed supernatant liquor, opened the lid of tube and kept it under the laminar flow cabinet for 10 min. 200 μL of TE buffer and 1 μL of RNase A (100 mg/mL) were together added into the tube and incubated for 30 min at 37°C.

Following the determination of the concentration and quality of DNA samples using a UV-1650 spectrophotometer and 1% agarose-gel electrophoresis, the DNA solutions were diluted to a working concentration with sterile ultra-pure water and stored at -80°C.

RAPD Analysis

A total of 660 random decamer primers (OPA ~ OPZ, OPAA ~ OPAD, OPAG, OPAM and OPAZ) were firstly screened against ‘Foushun’ and ‘Keumsull’ and six progenies. Only this kind of primer that could steadily amplify and show high polymorphism among our samples were selected for next genotyping analysis. The PCR reaction solution for RAPD analysis (total 20 μL) contained: 20 ng genomic DNA, 20 μM primer, 1.25 × PCR buffer, 0.25 mM per each dNTP, and 0.5 U Taq DNA polymerase (Dongsheng Biotech, Guangzhou,

China). The PCR program was as follows: 4 min at 94°C; 40 cycles of 1 min at 94°C, 1 min at 45°C, 1 min at 72°C; 5 min at 72°C to allow double-strand DNA synthesis. The PCR amplifications were performed by model 9902 Veriti TM 96-well Thermal Cycler (Applied Biosystems, CA, USA). The PCR amplicons were separated with electrophoresis in 1.0% agarose gel, visualized and taken picture under an UV transilluminator with Kodak EDAS 290 after being stained with ethidium bromide. The nomenclature of RAPD marker consisted of primer number and fragment size, for example "OPA-07-800" represents a 800 bp band amplified from the primer OPA-7. The RAPD primers were purchased from Eurofins MWG Operon (AL, USA).

SSR Analysis

The 60 pairs of SSR primers were chosen from reported papers (Ma et al., 2010; Taniguchi et al., 2012) and screened against two parents. These SSR loci were characterized with the names beginning of "MSE", "MSG", or "TM". SSR PCR reaction solution (20 μL) contain: 20 ng genome DNA, 0.5 μM each primer, 1× PCR buffer, 0.2 mM per each dNTP, and 0.5 U Taq DNA polymerase (Dongsheng Biotech, Guangzhou, China). SSR PCR program was as follows: 5 min at 94°C; 34 cycles of 30 s at 94°C, 60 s at annealing temperature, 30 s at 72°C; 10 min at 72°C to allow double-strand DNA synthesis. The all annealing temperatures were checked by gradient PCR with T100TM Thermal cycler (BioRad, CA, USA). The information together with the corresponding annealing temperature for each SSR primer were listed in Table 3. The PCR amplicons of primers were resolved in 3% agarose gel and were observed and taken picture as RAPD marker. If the polymorphic bands among our samples were too close to distinguish with 3% agarose gel, then PCR amplicons were diluted with 0.5 volume of loading dye [98% (v/v) formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene xyanole FF] and completely denatured 5 min at 95°C, the denatured samples were loaded on a 8% denaturing poly-acrylamide gel and visualized after silver staining. The electrophoresis was carried out with the double and triple wide mini-vertical gel system (MGV-202-33, C.B.S Scientific, CA, USA).

Map Construction

The linkage maps were constructed using JoinMap 4.0 version (Stam, 1993; Van Ooijen, 2006) using a 'CP' population type. The segregation data were sorted into two data sets derived from each parent before the linkage analysis. Both sets of genotyping data were

loaded into JoinMap 4.0 then the two maps of parents was separately constructed. Chi-square analysis was performed for goodness of fit to the expected Mendelian segregation ratio for each marker. Framework linkage groups were established by omitting the skewed markers from the data. These framework groups were treated as fixed orders for each map construction that next included other markers if they did not notably affect surrounding marker positions. The linkage groups were generated at a LOD score of 3.0 and a recombination frequency $\leq$ 0.40. The Kosambi mapping function was chosen to transfer recombination frequency into map distance.

Results and Discussion

DNA Quality

As shown in Fig. 3, all DNA samples of tea plants were good quality and essentially no degradation. The ratios of OD260/OD280 of all DNA samples were around 1.8 (data not shown). We can assume that those DNA samples were good enough to conduct the following work.

RAPD Analysis

A total of 660 random decamer primers were firstly screened against 'Fushun', 'Kemsull', and six F1 offspring. The candidate primers were selected on the basis of the number of easily scorable segregating bands. The screened 41 primers were used to amplify all F1 76 seedlings and only polymorphic bands that were steadily amplified and easily distinguished were selected as a potential source for mapping analysis.

Table 2 lists the total number of amplified DNA bands, polymorphic bands, and all sorable markers with the allelic configurations of Aa × aa or aa × Aa, and Aa × Aa for two highly heterozygous parents which are adoptable by JoinMap 4.0. The data summarized from 76 F1 seedlings revealed that a total of 399 DNA bands were amplified from these selected primers and with an average number of 9.7 bands, which was slighter smaller than a RAPD study in another perennial plant (Grattapaglia and Sederoff, 1994). A high degree of polymorphism was observed within these primers where the polymorphic ratios of 31 primers (75.6%) were more than 50%. For these polymorphic markers, 143 (54.8%) of them were selected for the genetic map construction of two parents. Parts of the agarose gel profiles amplified from five primers were shown in Fig. 4.

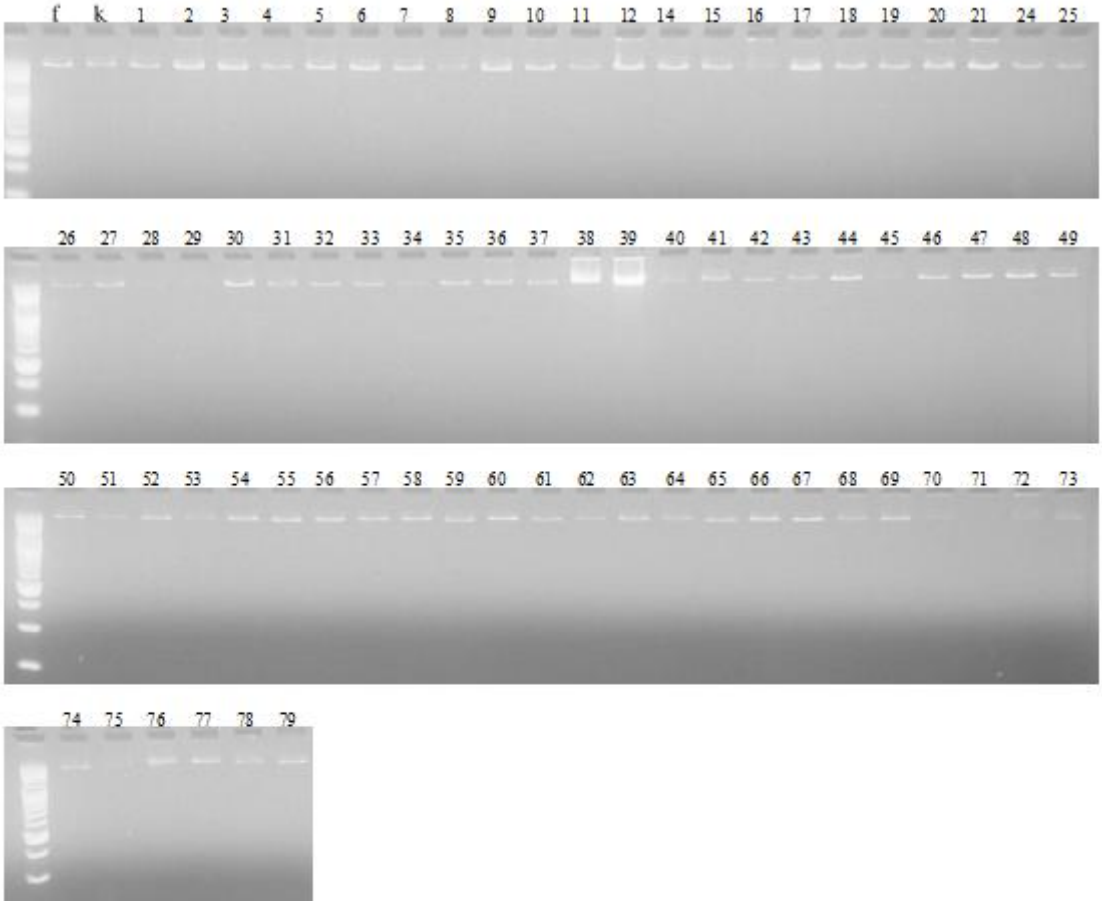


Fig. 3. Electrophoresis results of DNA samples of 'Fushun', 'Kemsull', and all F1 offspring. f, 'Fushun' (*Camellia sinensis*), female parent; k, 'Kemsull' (*C. sinensis*), male parent.

Table 2. PCR amplification of 41 primers selected from 660 random decamer primers.

Primer	No. of bands			Primer	No. of bands		
	Total	Polymorphic	Marker		Total	Polymorphic	Marker
OPA-03	5	5	1	OPG-13	9	6	4
OPA-07	11	11	1	OPG-14	8	2	1
OPA-08	6	6	4	OPG-15	14	10	10
OPA-12	2	2	2	OPG-19	8	1	1
OPA-15	6	6	2	OPI-03	11	5	3
OPA-17	8	8	5	OPI-04	14	7	4
OPA-18	4	3	1	OPI-07	9	4	1
OPB-01	15	9	3	OPI-14	10	3	1
OPB-03	11	8	5	OPI-15	10	10	6
OPB-12	16	8	7	OPI-16	7	7	2
OPB-15	15	11	3	OPL-15	11	5	1
OPB-20	16	10	5	OPL-18	12	7	2
OPC-04	8	8	6	OPM-04	11	7	5
OPC-06	14	9	6	OPM-06	5	4	1
OPC-19	12	8	8	OPM-12	12	8	1
OPC-20	11	4	2	OPV-16	14	9	2
OPD-06	8	8	5	OPV-17	10	6	2
OPE-02	6	2	2	OPW-05	11	10	10
OPE-07	10	4	1	OPX-17	9	9	8
OPF-09	12	5	3	OPAB-3	5	3	3
OPF-11	3	3	3	Total	399	261	143

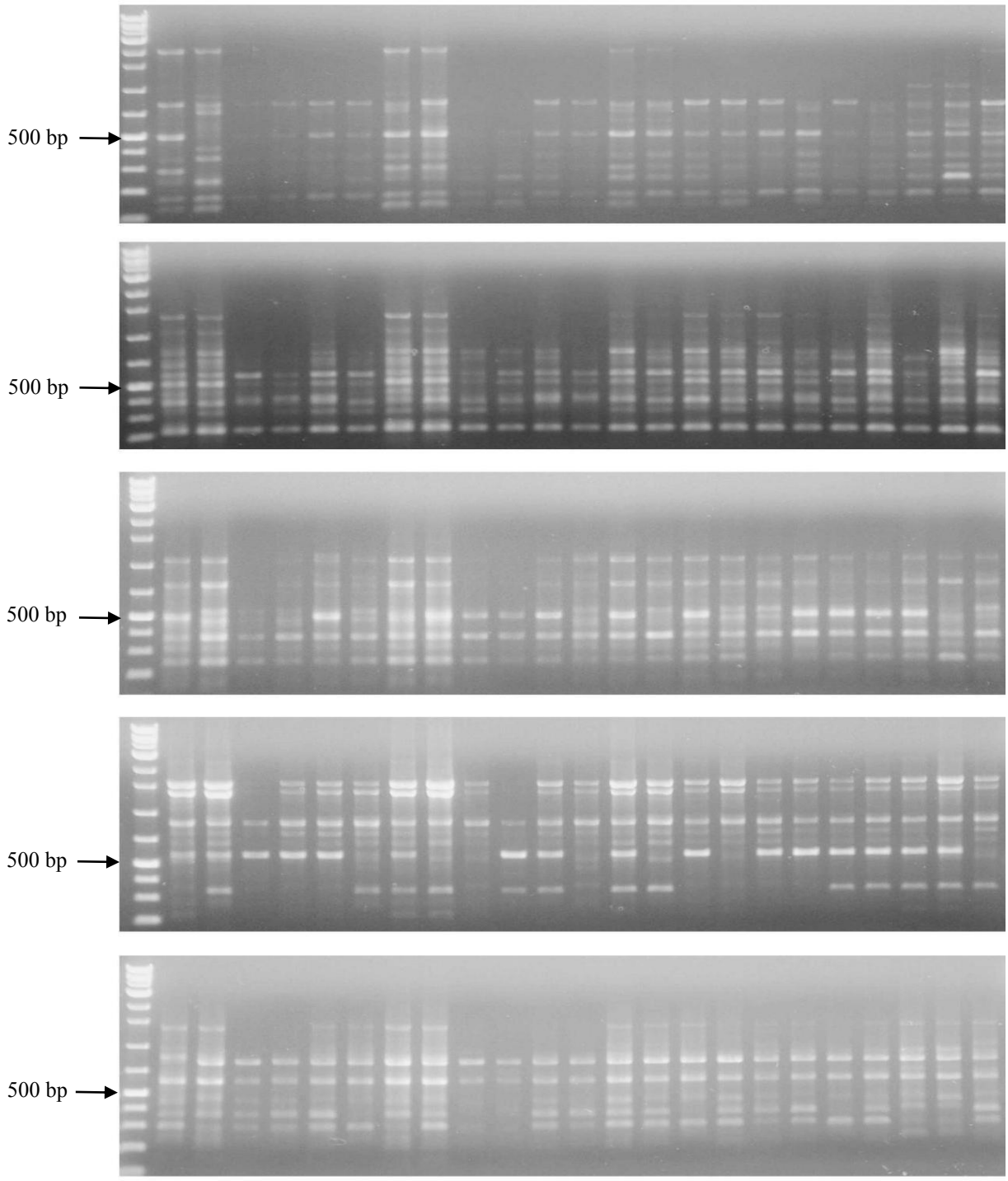


Fig. 4. Agarose profiles of amplicons of PCR with five random decamer oligonucleotide primers from two parents ‘Fushun’ (lane 2) and ‘Kemsull’ (lane 3) and 21 F1 hybrids (lane 4-24) that were seedlings from No. 1 to No. 24. DNA samples were successively amplified with OPB-01, OPC-19, OPC-20, OPG-15, and OPV-17 from up to down. All amplified products were resolved in a 1% agarose gel and electrophoresed with 0.5 × TBE buffer at 100 V for 1.5 h. Lane 1 was DNA ladder marker 1000.

SSR Analysis

60 public SSR primer pairs for tea plant were selected in the offspring of 'Fushun' and 'Kemsull' (Ma et al., 2010; Taniguchi et al., 2012). PCR amplifications were conducted using genomic DNA from both parents and six progenies with these SSR primer pairs. 15 SSR primer pairs could amplify polymorphic DNA products between two parents, which were clearly interpreted in agarose gel or poly-acrylamide gel and listed in Table 3. Two loci of MSG0232 and MSG0699 were shared between two parents, eight loci were just heterozygous at female parent 'Fushun', one loci was heterozygote at male parent 'Kemsull'. Figure 5 displayed partial agarose gel profiles of five pair primers of MSG0232, MSG0240, MSG0316, MSG0327, and MSG0811.

The dominant nature of SSR marker makes it greatly efficient to authenticate true hybrid from mapping population, because the allele contribution from each parent to their offspring could be unambiguously distinguished and affirmed. In this study, 60 published SSR primer pairs were selected and screened, and 15 of them (25.0%) showed polymorphism between two parents and different genotypes in the offspring. Five SSR loci of MSG0206, MSG0232, MSG0255, MSG0699, and TM103, which were heterozygous at both parents or one parent, could together ascertained that 74 individuals of this F1 population owned unique alleles from each parent and belonged to true hybrids. Although the authenticity of two seedlings (No. 20 and 62) could not be guaranteed right now, we still treated them as real hybrids and applied to linkage analysis and map construction, because only parental bands were amplified and observed at all five SSR loci. Undoubtedly, we will further determine their identify with more codominant markers in the near future. Here, the final F1 mapping population of 76 individuals was comparable to the other genetic mapping studies of tea plant, in which the applied mapping population sizes ranged from 46 to 188 (Tanaka, 1996; Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2005; Huang et al., 2006; Mewan et al., 2007; Taniguchi et al., 2007, 2012; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015).

Table 3. Genotypes of polymorphic SSR loci between ‘Fushun’ and ‘Kemsull’.

Name[a]	Forward primer sequence	Reverse primer sequence	Expected motif	Ta (℃)[b]	(excepted, observed)[c]	Genotype (f-k)[d]	Reference
TM068	cttcgtcctcctcctctttg	tcacatcatcagccttgggt	(ctcaat)5	62.9	195-215, 200-210	aa-bb	Ma et al. (2010)
TM103	gtccccattgctcttagttt	atcattgaccaccacatcat	(tatgtg)4	58.9	210-225, 160-220	aa-bc	
MSE0316	ggaaagagggaaccaccaaagaat	ccattcagatagccaaacaaagcc	(tc)14	63.2	313, 300-600	ab-aa	Taniguchi et al. (2012)
MSG0240	acgcgatttcttgaactagctctc	gtttcgatcgcaaacataaacatccttg	(ct)3, (tc)26, (tc)5	62.1	203, 160-200	ab-bb	
MSG0233	atttctcctccagatacgacggct	gtttatgggctctgacagtatggg	(agg)3, (ag)27, (aag)3, (gca)3	62.9	273, 250-280	aa-ab	
MSG0232	atttcgatctcccagacgtgaact	gtttacagaactagaagcaccaaaccgc	(tc)3, (tc)3, (tc)3, (tc)3, (tc)24	64.0	214, 200-240	ab-ac	
MSG0255	acccttgtctattctccatcgctc	gtttctagggtttgcgttgagaccttc	(tc)21, (ac)3, (ac)3, (ac)5	62.9	245, 210-260	ac-bb	
MSG0258	actcatcaccatgccttctccatc	gtttagctcaactggtggaacctcaact	(tc)25, (ta)4, (ta)3	69.5	289, 260-290	ab-aa	
MSG0327	acgttcaaaaccgtaatttggtcc	gtttgccacagttgcaatcaagtggtag	(tc)18, (tc)3, (cgt)3, (tc)3	62.0	273, 260-290	ab-bb	
MSG0403	atgatcgccggtttagagatgaat	gtttaagctggctaacctacacggagc	(tc)25	54.7	298, 270-300	ab-aa	
MSG0811	acaccacaccacaccacactttct	gtttggtctgaagctccaaagtgaa	(tc)15	66.1	151, 140-160	aa-ab	
MSG0699	atgcgacagtgttgctgagatttt	gtttcaaaaatggggtgtctacagaggg	(ag)18	63.0	249, 240-260	bc-ac	
MSG0206	atgagaaggtcatccaccatgaaa	gtttaaaaatccacatcccctcgttcc	(ag)8, (ag)10, (ag)3, (ag)19, (tat)3	63.2	260, 240-250	ac-bb	
MSG0473	atgatgttgatggattggtgtgga	gtttatgtgcgggactttgtgtttttg	(ag)3, (ag)20	64.0	216, 200-220	ab-aa	
MSE0159	aaatcttgatcgcaatcttgctcc	caaattgatttctcacatagggcca	(tc)20, (tg)5	54.0	298, 410-420	be-acd	

[a]: SSR loci underlined were applied into mapping analysis. [b]: Ta means the annealing temperature. [c]: The PCR amplicons comparison between expected band size and observed in our study. [d]: f, ‘Fushun’, female parent; k, ‘Kemsull’, male parent. For ‘TM103’, there are plentiful combined genotype ‘abc’ observed in the F1 population. For ‘MSG0233’, the PCR amplification effect was not good.

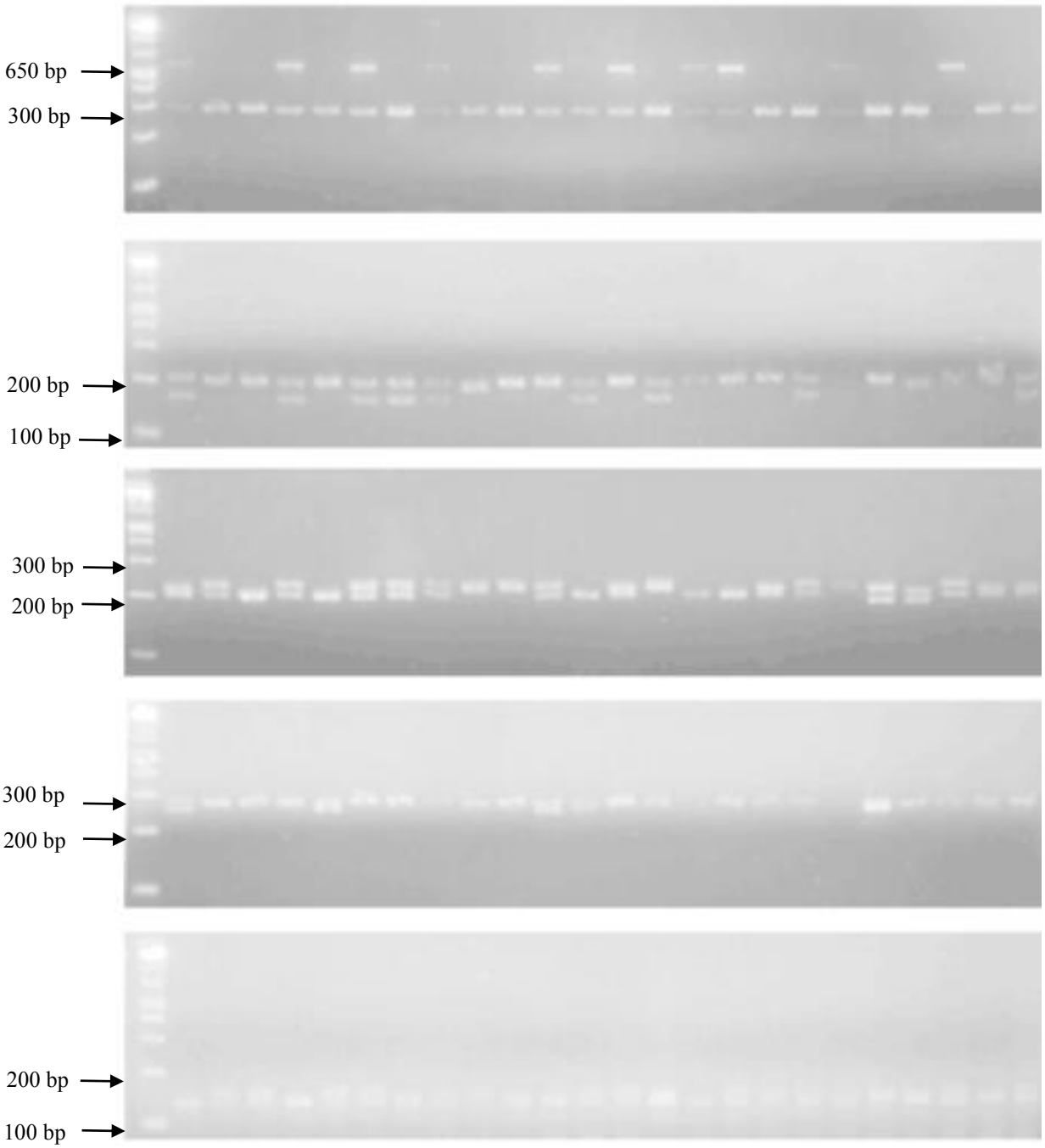


Fig. 5. Agarose profiles of amplicons of PCR with five SSR primer pairs for two parents 'Fushun' (lane 2) and 'Kemsull' (lane 3) and 21 F1 hybrids (lane 4-24) that were seedlings from No. 1 to No. 24. DNA samples were successively amplified from five pairs primer of MSG0232, MSG0240, MSE0316, MSG0327, and MSG0811 from up to down. All amplified products were resolved in 3% agarose gel and electrophoresed with 0.5 × TBE buffer at 100 V for 1.5 h. Lane 1 was DNA ladder marker 1000.

Genetic Map Construction with RAPD and SSR Markers

For 143 RAPD markers, 83 (58.0%) or 87 (60.8%) were significantly deviated from expected Mendelian ratios at the level of $p < 0.01$ and $p < 0.05$, respectively; for 11 SSR loci, 10 markers fitted with anticipated Mendelian segregation proportions at $p < 0.01$ or $p < 0.05$.

107 RAPD markers together with 10 SSR loci were used to develop a local genetic linkage map for tea cultivar 'Fushun'. The obtained map located 35 RAPD markers and 2 SSR loci that evenly distributed into eight linkage groups (LGs), covered a total length of 255.7 cM and with an average distance between two adjacent markers of 6.9 cM (Fig. 6). And 109 RAPDs with three SSRs were applied for constructing a partial genetic linkage map for Korea tea cultivar 'Kemsull'. The generated map included 30 RAPD markers spread into five LGs, spanned a map length of 196.1 cM and with a smaller density of 6.5 cM. The collinearities of LGs between two map were clearly identified with the help of common markers, such as f1 to k2, f3 to k3, f4 to k1, f5 to k4, and f8 to k5 (Fig. 6, 7). Undoubtedly, more molecular markers are required to perfect these two preliminary linkage maps. Because of different strategies applied, two maps obtained here showed no serious clustering phenomenon of markers, which was observed in an early report that two map were established without the initial omitting of distorted segregation markers (Chang et al., 2015).

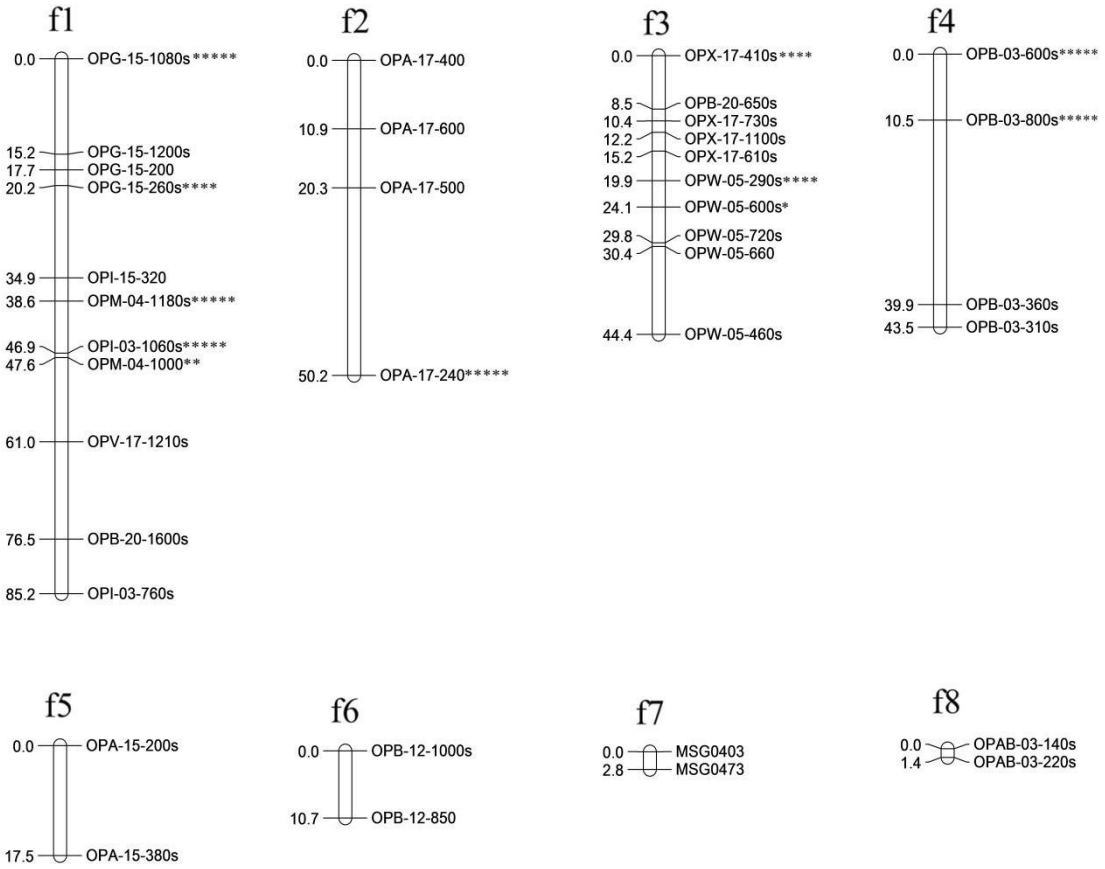


Fig. 6. A preliminary genetic linkage map based on RAPD and SSR markers for Japanese tea cultivar ‘Fushun’. Loci names were given to the right of each linkage group; genetic distance (cM) were to the left. Markers shared between two parents were postfixed with “s”. Markers with distorted segregation ratios were marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001).

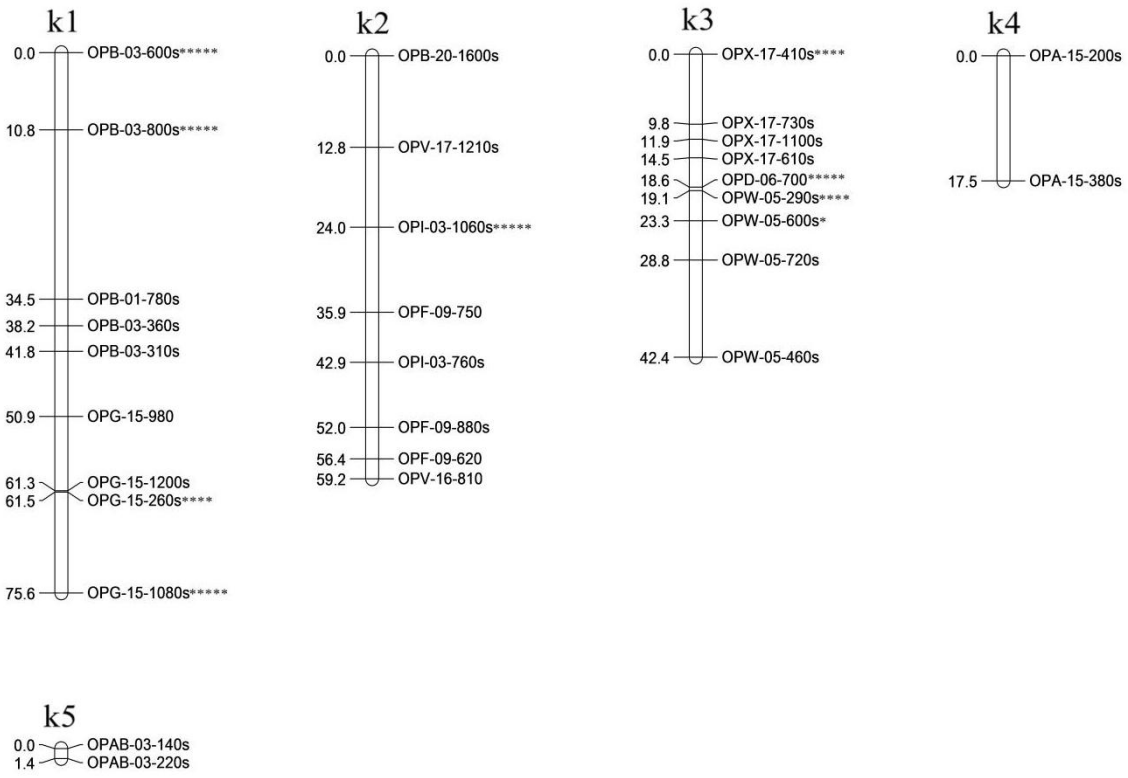


Fig 7. A preliminary genetic linkage map based on RAPD and SSR markers for Korean tea cultivar ‘Kemsull’. Notes were same as Fig. 6.

Literature Cited

Basu A, Lcas EA (2007) Mechanisms and effects of green tea on cardiovascular health. Nutr Rev 65:361-375.

Chang YL, Ou EN, Min SL, In KS, Joo Y, Kwan JS (2015) Preliminary linkage mapping of tea plants with RAPD and SSR markers. J Kor Tea Soc 89-94.

Chen L, Zeno A, Chen ZM (2012) Global tea breeding: achievements, challenges and perspectives. Hangzhou: Zhejiang University Press.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics. 137:1121-1137.

Hackett CA, Wachira FN, Paul S (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85(4):346-355.

Hu CY, Lee TC, Tsai HT (2013) Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). Euphytica 191:141-152.

Huang JA, Li JX, Huang YH, Luo JW, Gong ZH, Liu ZH (2005) Construction of AFLP molecular markers linkage map in tea plant. J Tea Sci 25 (1):7-15.

Huang FP, Liang YR, Lu JL et al. (2006) Genetic mapping of first generation of backcross in tea by RAPD and ISSR markers. J Tea Sci 26(3):171-176.

International Tea Committee. Annual bulletin of statistics 2013. London: ITC, PP48-53.

Jiang CJ, Liang YR, Luo JW (2009) Science of tea breeding. Beijing: China agriculture press.

Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, Kim H, Nishitani C, Terakami S et al. (2014) Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. Breed Sci 64: 240–251.

Ma JQ, Zhou YH, Ma CL, Yao MZ, Jin JQ, Wang XC, Chen L (2010) Identification and haracterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). American J Botany e153-156.

Ma JQ, Yao MZ, Ma CL, Wang XC, Jin JQ, Wang XM, Chen L (2014) Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (*Camellia sinensis*). PLoS ONE 9(3): e93131.

Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, Zheng HK, Yao MZ, 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.

Mewan KM, Saha MC, Konstantin C (2007) Construction of a genomic and EST-SSR based saturated genetic linkage amp of tea (*Camellia sinensis* L.). The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science (ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Ota S, Tanaka J (1999) RAPD-based linkage mapping using F1 segregating population derived from crossings between tea cultivar 'Sayamakaori' and strain 'Kana-Ck17'. Breeding Res (1):16.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739-744.

Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM et al. (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). PloS ONE 8(11): e81611.

Tanaka J (1996) The possibility of using RAPD analysis to create genetic map of tea. Tea Res J 84:44-45.

Taniguchi F, Tanaka J, Kono I (2007) Construction of genetic linkage map of tea using SSR markers. The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science(ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, Fukuoka H, Tanaka J (2012) Construction of a high-density reference linkage map of tea (*Camellia sinensis*). Breeding Sci 62:263-273.

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Weber CA, Moore GA, Deng ZA, Gmitter Jr FG (2003) Mapping freeze tolerance quantitative trait loci in a *Citrus grandis* × *Poncirus trifoliata* F1 pseudo-testcross using molecular markers. J Amer Soc Hort Sci 128: 508–514.

Weinreb O, Amit T, Mandel S, Youdim MBH (2009) Neuroprotective molecular mechanisms of (-)-epigallocatechin-3-gallate: a reflective outcome of its antioxidant, iron chelating and neuritogenic properties. Genes & Nutrition 4:283-296.

Yang CS, Wang X, Lu G, Picinich SC (2009) Cancer prevention by tea: animal studies, molecular mechanisms and human relevance. Nat Rev Cancer 9:429-439.

Yamamoto T, Terakami S, Takada N, Nishio S, Onoue N, Nishitani C, Kunihisa M, Inoue E, Iwata H et al. (2014) Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia Nakai*). Breed Sci 64: 351–361.

Zaveri NT (2006) Green tea and its polyphenolic catechins: medicinal uses in cancer and noncancer applications. Life Sci 78:2073-2080.

CHAPTER II . CONSTRUCTION OF GENETIC LINKAGE MAPS WITH RAPD, SSR, AND AFLP MARKERS FOR CHINA TYPE TEA PLANTS

Abstract

Based on the double pseudo-testcross theory, a population of 76 F1 offspring derived from a cross of two China type tea plants (*Camellia sinensis* var. *sinensis*), with a Japanese tea variety 'Fushun' as female parent and a Korean tea variety 'Kemsull' as male parent, were used as segregation population for constructing two genetic linkage maps for both parents based on AFLP markers. For the AFLP study, 2,439 bands were obtained from 27 primer combinations and with an average number of 90.3 per each. For 495 AFLP markers, 400 (80.8%) polymorphic loci with Mendelian segregation ratios ($p < 0.01$) were found, of which 136 (34.0%) were 3:1 segregation ratio and 264 (66.0%) were 1:1 segregation ratio. The map of 'Fushun' consisted of 92 loci that distributed into eight linkage groups (LGs), covered a total length of 470.8 cM and with an average distance of 5.2 cM between two adjacent markers; the map for 'Kemsull' included 17 LGs and totally located 152 loci, owned a map length of 1,053.1 cM with a mean interval of 6.9 cM among two neighbouring markers. Clear collinearities of six LGs at two maps were clearly identified through shared markers that were heterozygous at both parents, such as f1 to k1, f2 to k3, f3 to k6, f4 to k5, f6 to k2, and f7 to k14.

After combining 143 RAPD markers and 11 SSR loci with 495 AFLP markers as one data set, a single integrated genetic map was constructed using this strategy of 'One-step method' for tea plant by JoinMap 4.0. This developed genetic map contained 295 loci (76 RAPD, 5 SSR, and 214 AFLP markers) which evenly distributed into 15 LGs, covered a total length of 1,376.9 cM with an average interval of 4.7 cM between two adjacent markers, in which eight candidate segregation distorted regions were identified. This genetic map will lay a foundation for identifying quantitative or qualitative trait loci (QTLs) of important agronomic characters for genetic improvement of tea plant in the next study.

***Additional key works*:** *Camellia sinensis*, genome map, molecular marker, tea.

Introduction

The AFLP technique was initially exploited by Zabeau and Vos (1993) which combines the reliability of RFLP with the flexibility of PCR technique and provides a universal and multi-locus marking technique for various genetic studies in plant, such as genetic diversity analysis, phylogenetic relationship research, and genetic map construction.

Studying genetic diversity with new molecular biological technology will lay a theoretical foundation for reasonably protecting and utilizing the existing tea germplasm. Using the AFLP marker system, Paul et al. (1997) studied genetic diversity and difference between some tea populations planted in Kenya and India and also revealed the genetic variation differences between intra-populations and inter-populations. They concluded 79% of genetic variation was from intra-populations while the rest of 21% from inter-populations. From the dendrogram based on AFLP bands, all studied tea plant populations could be clearly classified into China type (*sinensis*), Assam type (*assamica*), and Cambod type (*assamica* ssp. *lasiocalyx*). The PCO analysis results not only revealed that China type tea accessions dispersedly distributed in the plot and owned widest genetic diversity, but also supported the hypothesis that tea clones in Kenya had been initially selected from India region. Wachira et al. (2001) studied the genetic diversities of 40 tea varieties around the world using RAPD and AFLP markers, and mainly subdivided them into three groups after carried out the clustering analysis, in which one group contained all Assam variants, another group included all China type, and the last group mixed some cultivation tea varieties with wild tea plants. This analysis results showed that all intra-populations owns high level of genetic variations and the degree of genetic variation of these tea plants descended as the following order: wild tea plants > India tea varieties > China tea varieties > Kenya tea plants > Sri lanka tea varieties > Vietnam tea varieties > Japan tea plants > China Taiwan tea varieties. Balasarvavanan et al. (2003) analyzed genetic diversities of 49 tea varieties with AFLP markers which were widespreadly planted in Southern India. This results revealed that a lower genetic diversity existed between these broadly cultivated tea varieties in South India. Hou et al. (2015) studied the genetic diversity and phylogenetic relationship of 28 tea varieties or new breeding materials and concluded there was a certain extent between the genetic relationships and their geographical original places.

A saturated genetic linkage map is the foundation for marker assisted selection (MAS) which will greatly shorten breeding period and save cost especially for perennial plants just

as tea plant. The construction of genetic linkage maps for tea plant has been quickly developed using various molecular markers from different countries (Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2005; Huang et al., 2006; Mewan et al., 2007; Taniguchi et al., 2007, 2012; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015) after Tanaka (1996) reported the initial linkage maps based on a small quantity of RAPD markers using a F1 generation population (46 progenies) for two tea varieties 'Yabukita' and 'Shizu-Inzatsu 131'. Up to now, the development of four genetic linkage maps for tea plants have adopted the AFLP marker. Hackett et al. (2000) firstly applied the AFLP marker technique into the map construction for tea plant. They established a F1 segregation population consisting of 90 seedlings as mapping population which was deemed to be derived from a cross of two tea varieties SFS150 and TN14/3, then developed a genetic linkage map for female parent with AFLP and RAPD markers. The generated map consisted of 126 loci spread over 15 linkage groups (LGs), spanned a total length of 1,349.7 cM and with an average distance of 11.7 cM between two adjacent markers. Huang et al. (2005) reported two genetic maps for Chinese famous tea varieties of 'Qi Men No.4' and 'Chao An Da Wu Ye' based on AFLP markers using a set of 69 F1 offspring as mapping materials. The map of 'Qi Men No.4' totally located 208 loci distributed into 17 LGs, covered a length of 2,457.7 cM and with an average interval of 11.9 cM between two neighbouring markers; for male parent 'Chao An Da Wu Ye', the obtained map consisted of 200 AFLP loci evenly distributed into 16 LGs, spanned over a overall length of 2,543.3 cM with an slightly higher average distance between markers of 12.8 cM. Kamunya et al. (2010) developed a consensus map based on 22 RAPDs, 4 SSRs and 74 AFLPs for tea plant using a set of 42 F1 offspring crossed from two tea varieties with significant different agronomic traits as mapping population, which included 30 LGs, covered a total map length of 1,411.5 cM and with a mean interval between markers of 14.1 cM. Hu et al. (2013) built an integrated map including for tea plant 36 SSRs, 3 CAPSs, 1 STS, 250 AFLPs, 13 ISSRs, and 64 RAPDs.

In this chapter, we used a set of 76 F1 progenies as mapping population, which was derived from a cross of Japanese tea cultivar 'Fushun' (*Camellia sinensis*) as female parent and Korean tea cultivar 'Kemsull' (*C. sinensis*) as male parent, to construct genetic linkage maps for both parents based on AFLP markers. Meanwhile, 143 RAPD and 11 SSR markers which was developed from previous study together with AFLP markers obtained here were applied to establish a combined genetic linkage map for China type tea plant.

Materials and Methods

Plant Materials

The same mapping materials in chapter Ⅰ were used into this study.

DNA Isolation

The preparation and isolation of total genomic DNA was same as chapter Ⅰ.

RAPD and Reported SSR Analysis

The approaches for developing RAPD and SSR markers have been reported previously in chapter one. A total of 660 decamer random primers were first screened against two parents and six F1 seedlings and then 41 were selected for the subsequent genotyping analysis. RAPD markers consist of primer number and band size, such as the "OPA-07-800" marker. 60 SSR primer pairs were chosen from previous reports (Ma et al., 2010; Taniguchi et al., 2012) and screened against two parents. Those SSR loci were characterized with the name beginnings of "MSE," "MSG," or "TM."

AFLP Analysis and Screening Primers

AFLP Technique System. AFLP analysis was performed according to the method of Zhang (2009) with some modifications. Firstly, the digestion-ligation reaction of genomic DNA with restriction enzymes and adaptors was carried out at the same time within a 20 μL reaction solution containing 200 ng of genomic DNA, 15 U *Eco*RⅠ and 15 U *Mse*Ⅰ, 5 pmol *Eco*RⅠ- adaptor and 50 pmol *Mse*Ⅰ- adaptor, T4 DNA ligase 35 U, 0.5 mM ATP (TAKARA, Shiga, Japan), and 1 × PCR buffer (1.6 mM Mg^{2+}) (DONGSHENG BIOTECH, Guangdong, China) for 16 h at 37°C. Secondly, 2 μL of this digestion-ligation product was used as template for next preamplification. The reaction system for PCR amplification contained that 1 × PCR buffer (1.6 mM Mg^{2+}), 1 pmol *Eco*RⅠ-primer (E0), 1 pmol *Mse*Ⅰ-primer (M0), 0.2 mM dNTP (per each), and 0.5 U Taq DNA polymerase (DONGSHENG BIOTECH, Guangdong, China). The primers of E0 and M0 both had one selective nucleotide on the 3' terminals (Table 4). The PCR conditions were as follows: 5 min at 94°C; 30 cycles of 30 s at 94°C, 1 min at 56°C, 1 min at 72°C; 10 min at 72°C to allow perfect double-strand DNA extension. Thirdly, the amplicons from PCR preamplification was firstly diluted 10 times with sterile ultra-pure water, then 2 μL of this diluted product was utilized as template for the following selective amplification. This

solution included 1 × PCR buffer (1.6 mM Mg^{2+}), 1 pmol *Eco*R Ⅰ-primer (E1-E4), 1 pmol *Mse* Ⅰ-primer (M1-M11), 0.2 mM dNTP (per each), and 0.5 U Taq DNA polymerase. The primers of E1-E4 had two selective nucleotides (E0+NN) and M1-M11 had three (M0+NNN, Table 4). The PCR program as follows: 5 min at 94°C; 30 s at 94°C, 1 min at 65°C decreased with 0.7°C in each cycle for next 13 cycles then continued at 56°C for the remaining 24 cycles, 1 min at 72°C; finally 10 min at 72°C to allow double-strand DNA extension. The products of PCR selective amplification were diluted with 0.5 volume of loading dye [98% (v/v) formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene xyanole FF] and denatured 5 min at 95°C. Next, the samples were loaded on a 8% denaturing poly-acrylamide gel and the following electrophoresis was carried out on the Sequi-Gen® GT Nuclei Acid Electrophoresis cell (Bio-Rad, South Korea). A total of 44 AFLP primer combinations (PCs) were screened against 'Kemsull', 'Fushun', and a few offsprings. The name of an AFLP marker refers to two primers codes and band size. For example, "E1/M5-218" represents the fragment of 218 bp obtained with PC of E1/M5.

Screening Primers. 44 PCs obtained from the pair combination between four primers of E-NN and eleven primers of M-NNN (Table 4), were screened against two parents and a few individuals for the following genotyping analysis. After electrophoresed with 8% denaturing poly-acrylamide gel and silver staining, those PCs amplified with clear bands, more polymorphic ratio, and high reproducibility were selected to carry out the PCR amplification for the whole F1 offspring.

Map Construction

The PCR amplifications were conducted for two parents 'Fushun', 'Kemsull' and all F1 individuals with those selected PCs. All electrophoresis banding patterns were gathered together. Then we proceeded the digitization step of all electrophoresis bands with the numbers of '1' and '0' based on double-pseudo testcross format strategy. Before conducting linkage analysis with JoinMap 4.0, three loci types of (1,0), (0,1) and (1,1) obtained from this AFLP analysis were transformed into suitable codes of (lm, ll), (nn, np) and (h-, kk), respectively. Linkage maps were constructed for each parent based on AFLP marker with JoinMap 4.0 version (Stam, 1993; Van Ooijen, 2006) using a 'CP' population type. Firstly, the segregation data were sorted into two data sets derived from each parent before the linkage analysis. Both sets of genotyping data were loaded into JoinMap 4.0 then two maps of parents was constructed for two parents, respectively. Chi-square analysis was performed

for goodness of fit to the expected Mendelian segregation ratio for each marker. Framework linkage groups were established by omitting the skewed markers from the data. These framework groups were treated as fixed orders for each map construction that next included other markers if they did not noticeably affect surrounding marker positions. The linkage groups were generated at a LOD score of 3.0 and a recombination frequency ≤ 0.40. The Kosambi mapping function was chosen to transfer recombination frequency into map distance.

Two genetic linkage maps were subsequently established for two parents using RAPD, SSR, and AFLP markers with JoinMap 4.0 version (Stam, 1993; Van Ooijen, 2006) followed the same strategy as above. Secondly, a consensus genetic map was constructed by merging all the markers heterozygous in both or either of the parental plant as one set of data with the same strategy, which was known as 'One-way method' Tavassolian et al (2010).

Table 4. Sequences and codes of primers used for PCR amplification in AFLP analysis.

Adaptors or primers	Sequences
*Eco*R Ⅰ adaptor	5’ -CTCGTAGACTGCGTACC- 3’
	3’ -CTGACGCATGGTTAA- 5’
Mse Ⅰ adaptor	5’ -GACGATGAGTCCTGAG- 3’
	3’ -TACTCAGGACTCAT- 5’
*Eco*R Ⅰ primer	GACTGCGTACCAATTCA (**E0**)
	5’ -GACTGCGTACCAATTCA+AG- 3’ (**E-AG, E1**)
	5’ -GACTGCGTACCAATTCA+AT- 3’ (**E-AT, E2**)
	5’ -GACTGCGTACCAATTCA+AA- 3’ (**E-AA, E3**)
	5’ -GACTGCGTACCAATTCA+AC- 3’ (**E-AC, E4**)
Mse Ⅰ primer	5’ -GATGAGTCCTGAGTAA+A- 3’ (**M0**)
	5’ -GATGAGTCCTGAGTAAA+AAC- 3’ (**M-AAC**, **M1**)
	5’ -GATGAGTCCTGAGTAAA+AGA- 3’ (**M-AGA**, **M2**)
	5’ -GATGAGTCCTGAGTAAA+AAG- 3’ (**M-AAG**, **M3**)
	5’ -GATGAGTCCTGAGTAAA+AAT- 3’ (**M-AAT**, **M4**)
	5’ -GATGAGTCCTGAGTAAA+AGT- 3’ (**M-AGT**, **M5**)
	5’ -GATGAGTCCTGAGTAAA+CTA- 3’ (**M-CTA**, **M6**)
	5’ -GATGAGTCCTGAGTAAA+CTT- 3’ (**M-CTT**, **M7**)
	5’ -GATGAGTCCTGAGTAAA+CTC- 3’ (**M-CTC**, **M8**)
	5’ -GATGAGTCCTGAGTAAA+CAA- 3’ (**M-CAA**, **M9**)
	5’ -GATGAGTCCTGAGTAAA+CAG- 3’ (**M-CAG**, **M10**)
	5’ -GATGAGTCCTGAGTAAA+CAT- 3’ (**M-CAT**, **M11**)

Results and Discussion

Primer Screening for AFLP analysis

For 44 primer combinations (PCs), PCR amplifications were carried out with DNA samples of 'Fushun', 'Kemsull', and six their F1 descendants. After electrophoretic separation and silver-staining detection, 27 PCs were chosen for next PCR amplification for the whole segregation population. The screening results of E-AT/M-NNN were shown in Fig. 8, from which six PCs were selected for next study. Table 5 listed these 27 PCs and their corresponding codes adopted in our study.

Distribution of AFLP Markers in F1 offspring

As shown in Table 6, all AFLP markers accorded with Mendelian segregation ratios were 382 ($p < 0.05$) or 400 ($p < 0.01$), which severally accounted for 15.7% ($p < 0.05$) or 16.4% ($p < 0.01$) of total bands amplified from 27 PCs, and occupied for 77.3% ($p < 0.05$) or 80.8% ($p < 0.01$) of all polymorphic AFLP markers. There were 251 ($p < 0.05$) or 264 ($p < 0.01$) AFLP markers separated with a 1:1 Mendelian ratio, took a proportion of 65.7% ($p < 0.05$) or 66% ($p < 0.01$) for all markers fitted with the Mendelian separation rate, and possessed a percentage of 50.7% ($p < 0.05$) or 53.3% ($p < 0.01$) in the all polymorphic AFLP markers (Table 7). This kind of markers were considered as an important potential for building a single genetic map per each parent. In addition, 131 ($p < 0.05$) or 136 ($p < 0.01$) markers were accorded with a Mendelian separation rate of 3:1 that severally accounted for 22.8% ($p < 0.05$) or 19.2% ($p < 0.01$) for total polymorphic markers (Table 7), which will be greatly useful to identify homologous linkage groups between two parental genetic maps.

In Table 6, 113 ($p < 0.05$) or 95 ($p < 0.01$) AFLP markers belonged to the distorted segregation (DS) markers, which severally occupied 4.6% ($p < 0.05$) or 3.9% ($p < 0.01$) of total bands, and accounted for 22.8% ($p < 0.05$) or 19.2% ($p < 0.01$) of all AFLP polymorphic markers. In our study, partial of DS markers were used for genetic map construction after established the framework map with non DS markers by JoinMap 4.0.

Table 5. Primer combinations and their codes for selective amplification of PCR.

Primer combination	Code	Primer combination	Code
E-AG/M-AGT	E1/M5	E-AA/CTA	E3/M6
E-AG/M-AGA	E1/M2	E-AA/AGT	E3/M5
E-AG/M-AAC	E1/M1	E-AA/CTC	E3/M8
E-AG/M-CTT	E1/M7	E-AA/CAA	E3/M9
E-AG/M-CTC	E1/M8	E-AC/CTC	E4/M8
E-AT/AGA	E2/M2	E-AC/AAT	E4/M4
E-AT/AGT	E2/M5	E-AC/CAT	E4/M11
E-AT/CTA	E2/M6	E-AC/CTA	E4/M6
E-AT/CTT	E2/M7	E-AC/CAA	E4/M9
E-AT/CAT	E2/M11	E-AC/AGT	E4/M5
E-AT/AAG	E2/M3	E-AC/CAG	E4/M10
E-AA/CTT	E3/M7	E-AC/AGA	E4/M2
E-AA/CAG	E3/M10	E-AC/AAG	E4/M3
E-AA/AGA	E3/M2		

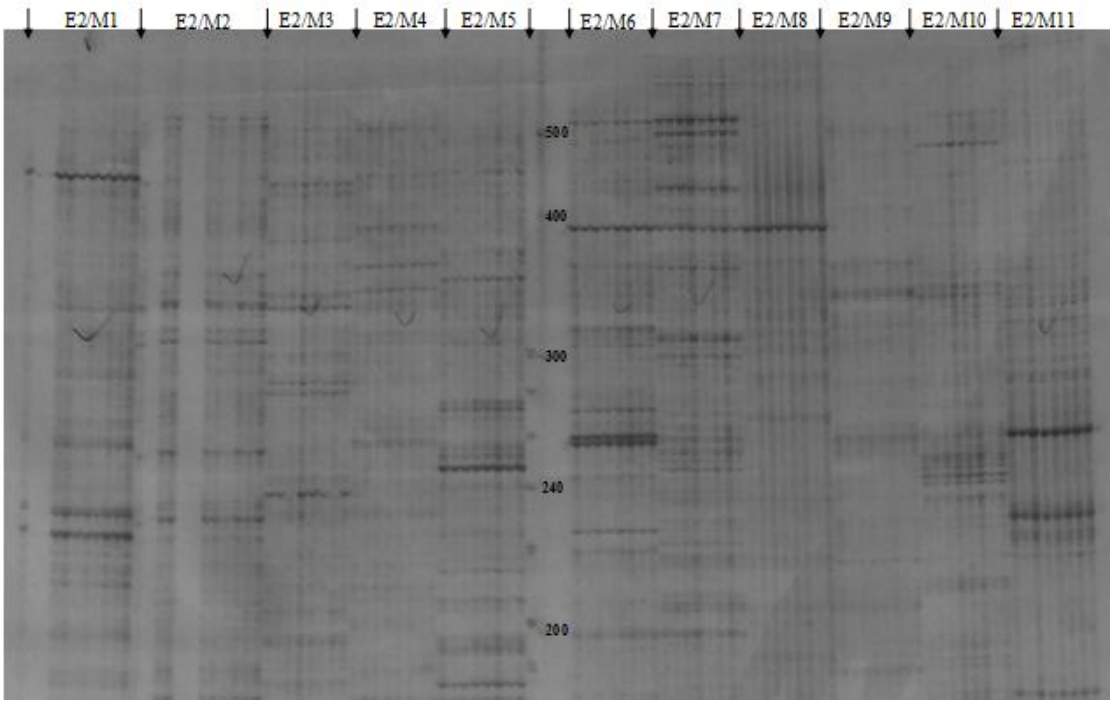


Fig. 8. PCR amplification results of two parents and six offspring with eleven primer combinations (E-AT/M-NNN). Six PCs including E2/M2, E2/M3, E2/M5, E2/M6, E2/M7, and E2/M11 were selected to conduct the following genotyping study for all F1 seedlings.

Table 6. Distribution of AFLP polymorphic markers in F1 progenies.

Code	Total markers	Non DS marker		Frequency (%)		DS marker		Frequency (%)	
		$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$
1	11	8	8	72.7	72.7	3	3	27.3	27.3
2	23	21	22	91.3	95.7	2	1	8.7	4.3
3	16	14	15	87.5	93.8	2	1	12.5	6.2
4	12	7	7	58.3	58.3	5	5	41.7	41.7
5	16	13	13	81.3	81.3	3	3	18.7	18.7
6	17	14	14	82.4	82.4	3	3	17.6	17.6
7	13	10	10	76.9	76.9	3	3	23.1	23.1
8	13	6	7	46.2	53.8	7	6	53.8	46.2
9	16	13	14	81.3	87.5	3	2	18.7	12.5
10	13	10	11	76.9	84.6	3	2	23.1	15.4
11	18	13	13	72.2	72.2	5	5	27.8	27.8
12	16	9	10	56.3	62.5	7	6	43.7	37.5
13	26	20	21	76.9	80.8	6	5	23.1	19.2
14	22	15	16	68.2	72.7	7	6	31.8	27.3
15	28	23	23	82.1	82.1	5	5	17.9	17.9
16	24	21	22	87.5	91.7	3	2	12.5	8.3
17	18	13	15	72.2	83.3	5	3	27.8	16.7
18	18	17	18	94.4	100	1	0	5.6	0
19	16	13	13	81.3	81.3	3	3	18.7	18.7
20	20	18	18	90	90	2	2	10	10
21	14	7	9	50	64.3	7	5	50	35.7
22	21	16	16	76.2	76.2	5	5	23.8	23.8
23	15	14	15	93.3	100	1	0	6.7	0
24	18	9	10	50	55.6	9	8	50	44.4
25	45	36	38	80	84.4	9	7	20	15.6
26	15	11	11	73.3	73.3	4	4	26.7	26.7
27	11	11	11	100	100	0	0	0	0
Total	495	382	400	/	/	113	95	/	/
Average	18.3	14.1	14.8	77.2	80.8	4.2	3.5	22.8	19.2

Table 7. Distribution of AFLP Mendelian segregation markers in F1 progenies.

Code	Polymorphic markers	1:1 segragation marker		Frequency (%)		3:1 segregation marker		Frequency (%)	
		$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$
1	11	7	7	63.7	63.7	1	1	9.1	9.1
2	23	15	15	65.2	65.2	6	7	26.1	30.4
3	16	9	10	56.3	62.5	5	5	31.3	31.3
4	12	6	6	50	50	1	1	8.3	8.3
5	16	8	8	50	50	5	5	31.3	31.3
6	17	12	12	70.6	70.6	2	2	11.8	11.8
7	13	6	6	46.2	46.2	4	4	30.8	30.8
8	13	4	5	30.8	38.5	2	2	15.4	15.4
9	16	10	11	62.5	68.8	3	3	18.8	18.8
10	13	8	9	61.5	69.2	2	2	15.4	15.4
11	18	8	8	44.4	44.4	5	5	27.8	27.8
12	16	3	3	18.8	18.8	6	7	37.5	43.8
13	26	13	13	50	50	7	8	26.9	30.8
14	22	10	11	45.5	50	5	5	22.7	22.7
15	28	11	11	39.3	39.3	12	12	42.9	42.9
16	24	12	13	50	54.2	9	9	37.5	37.5
17	18	11	13	61.1	72.2	2	2	11.1	11.1
18	18	7	7	38.9	38.9	10	11	55.6	61.1
19	16	6	6	37.5	37.5	7	7	43.8	43.8
20	20	15	15	75	75	3	3	15	15
21	14	6	7	42.9	50	1	2	7.1	14.3
22	21	8	8	38.1	38.1	8	8	38.1	38.1
23	15	9	10	60	66.7	5	5	33.3	33.3
24	18	6	7	33.3	38.9	3	3	16.7	16.7
25	45	29	31	64.4	68.9	7	7	15.6	15.6
26	15	6	6	40	40	5	5	33.3	33.3
27	11	6	6	54.5	54.5	5	5	45.5	45.5
Total	495	251	264	/	/	131	136	/	/
Average	18.3	9.3	9.8	50.7	53.3	4.9	5	26.5	27.5

For 27 PCs, the number of AFLP fragments varied from 65 to 128 and with an average of 90.3 per each, which was similar the average visible bands of 88.7 presented by Hackett et al. (2000) and 87.5 reported from Huang et al. (2005). Totally, 2,439 bands and 495 (20.3%) AFLP markers were identified in our study. The polymorphic percentage was slightly lower than 25.19% reported from Huang et al. (2005). The polymorphic ratios varied greatly from 11.7% to 35.2% among those 27 PCs, in which E1/M5 and E4/M10 amplified the lowest and highest polymorphic frequencies, respectively. For 495 AFLP markers, 187 (37.8%) and 118 (23.8%) were generated from male parent 'Kemsull' and female parent 'Fushun', respectively, and the remaining 190 AFLP markers belonged to shared markers which existed at both parents and also showed different band patterns in the F1 progenies (Table 8).

According to Chi-square analysis, 400 (80.8%) of 495 AFLP polymorphic markers fitted the expected Mendelian ratio ($p < 0.01$), 136 (34.0%) showed 3:1 segregation ratio while other 264 (66.0%), 1:1 segregation ratio. When Huang et al. (2005) constructed linkage maps for China type tea varieties 'Qi Men No.4' and 'Chao An Da Wu Ye' based on AFLP markers, 26.6% of all tested markers showed distorted segregation ($p < 0.01$). This percentage was slighter higher than our result of 19.2%, which could be partially attributed to the smaller population size (69 F1 seedlings) used in their study. Furthermore, just distinct and reproducible polymorphic bands from total produced bands were applied for next genotypic analysis in our study that may also explain the decrease of distorted marker rate to a certain degree extent.

The polymorphic figure of parents and 76 F1 progenies amplified with the PCs of E3/M9 was shown in Fig. 9.

Table 8. Polymorphism AFLP markers obtained from PCR amplifications of 27 PCs.

Primer combination	No. of polymorphic band				No. of total band	Polymorphic ratio
	Marker from Fushun	Marker from Kemsull	Shared marker	Total		
E1/M5	3	6	2	11	94	11.7%
E1/M2	9	7	7	23	95	24.2%
E1/M1	2	8	6	16	95	16.8%
E1/M7	2	7	3	12	95	12.6%
E1/M8	2	6	8	16	92	17.4%
E2/M2	11	4	2	17	99	17.2%
E2/M5	0	8	5	13	68	19.1%
E2/M6	2	5	6	13	65	20.0%
E2/M7	2	10	4	16	103	15.5%
E2/M11	2	8	3	13	89	14.6%
E2/M3	2	9	7	18	94	19.1%
E3/M7	1	2	13	16	73	21.9%
E3/M10	3	10	13	26	94	27.7%
E3/M2	2	12	8	22	92	23.9%
E3/M6	2	11	15	28	113	24.8%
E3/M5	3	12	9	24	108	22.2%
E3/M8	6	7	5	18	93	19.4%
E3/M9	6	1	11	18	118	15.3%
E4/M8	2	5	9	16	71	22.5%
E4/M4	4	12	4	20	90	22.2%
E4/M11	5	5	4	14	74	18.9%
E4/M6	2	8	11	21	77	27.3%
E4/M9	4	6	5	15	100	15.0%
E4/M5	3	5	10	18	72	25.0%
E4/M10	28	8	9	45	128	35.2%
E4/M2	8	1	6	15	81	18.5%
E4/M3	2	4	5	11	66	16.7%
Total	118	187	190	495	2,439	20.3%

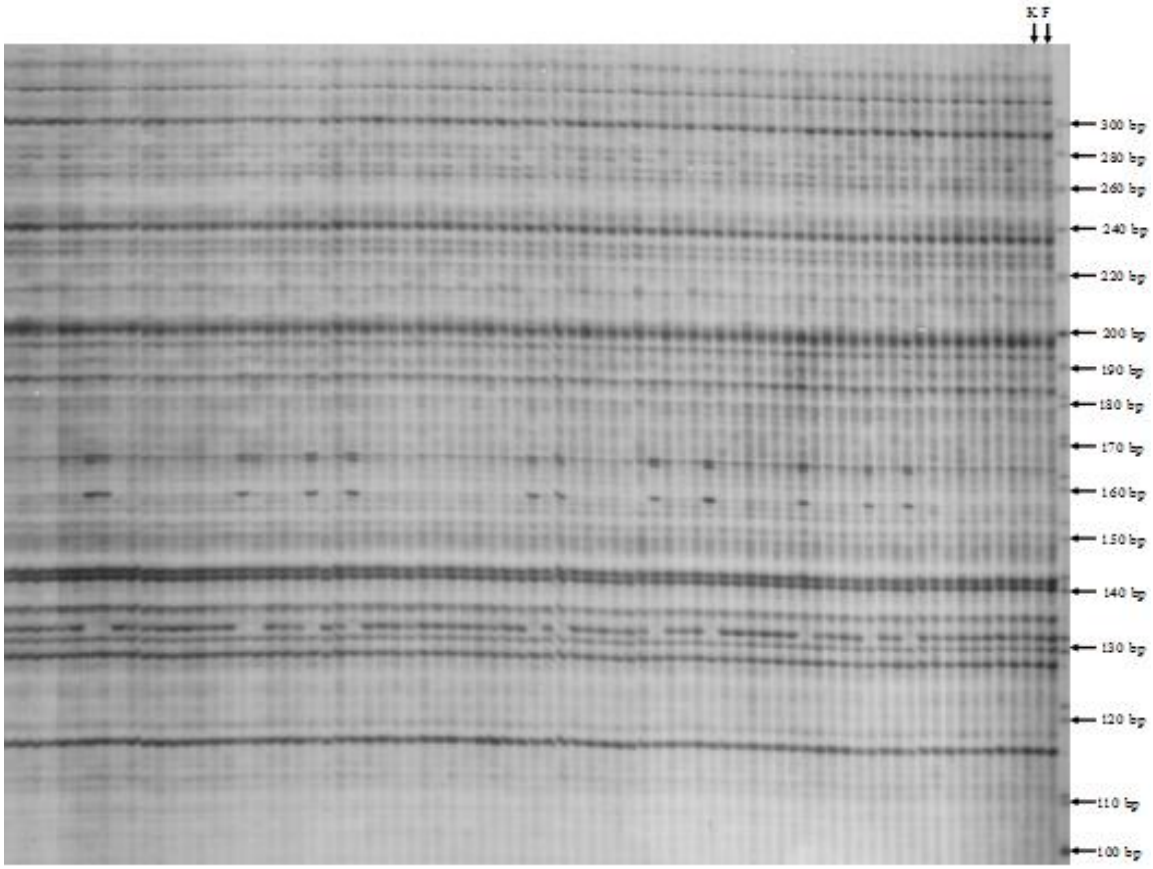


Fig. 9. AFLP fingerprinting patterns of parents and F1 offspring amplified by the primer combination of E3/M9. Notes: lane 79, 'Fushun' (F), female parent; lane 78, 'Kemsull' (K), male parent; Lane 80, DNA marker; lanes 1-76, 76 F1 individuals, from No. 1 to No. 79. Three seedlings including No.13, No.22, and No.23 were dead.

Linkage Analysis and Genetic Map Construction of Female Parent 'Fushun'

For 307 AFLP markers used for map construction of 'Fushun', 175 (57.0%) belonged to non DS markers and were firstly applied for establishing the framework LGs, then DS markers were added into these LGs step by step. The obtained genetic map consisted of 92 markers that evenly distributed into eight LGs, covered a total length of 470.8 cM and with an average distance of 5.2 cM between two adjacent markers. Twelve DS markers were located into three LGs of f1, f2, and f3 ($p < 0.05$). Over half of all mapped markers were located in the first LG which rendered it to own the highest marker density of 1.7 cM. The LG of f2 mapped 10 loci and covered the longest length of 110 cM. A total of fifty-seven shared markers were mapped into six LGs in this map, which could be used for identifying those LGs with collinearity in the map of 'Kemsull'.

This local genetic map contained four small LGs that are only located two or three loci, the number of eight LGs was not equal to the basic chromosome number of 15. More markers were required to develop and saturate this map.

Table 9. Summarized data of each linkage group of genetic map for ‘Fushun’.

LG	No. of marker				Length (cM)	Average distance (cM)
	1:1 marker	3:1 marker	Total	DS marker ($p < 0.05$)		
f1	22	31	53	8	88.1	1.7
f2	1	10	11	3	110.0	10
f3	2	6	8	1	83.8	10.5
f4	6	4	10	0	81.1	8.11
f5	3	0	3	0	63.5	21.2
f6	0	2	2	0	29.7	14.9
f7	0	2	2	0	10.6	5.3
f8	2	0	2	0	4.0	2.0
Total	34	57	91	0	470.8	5.2

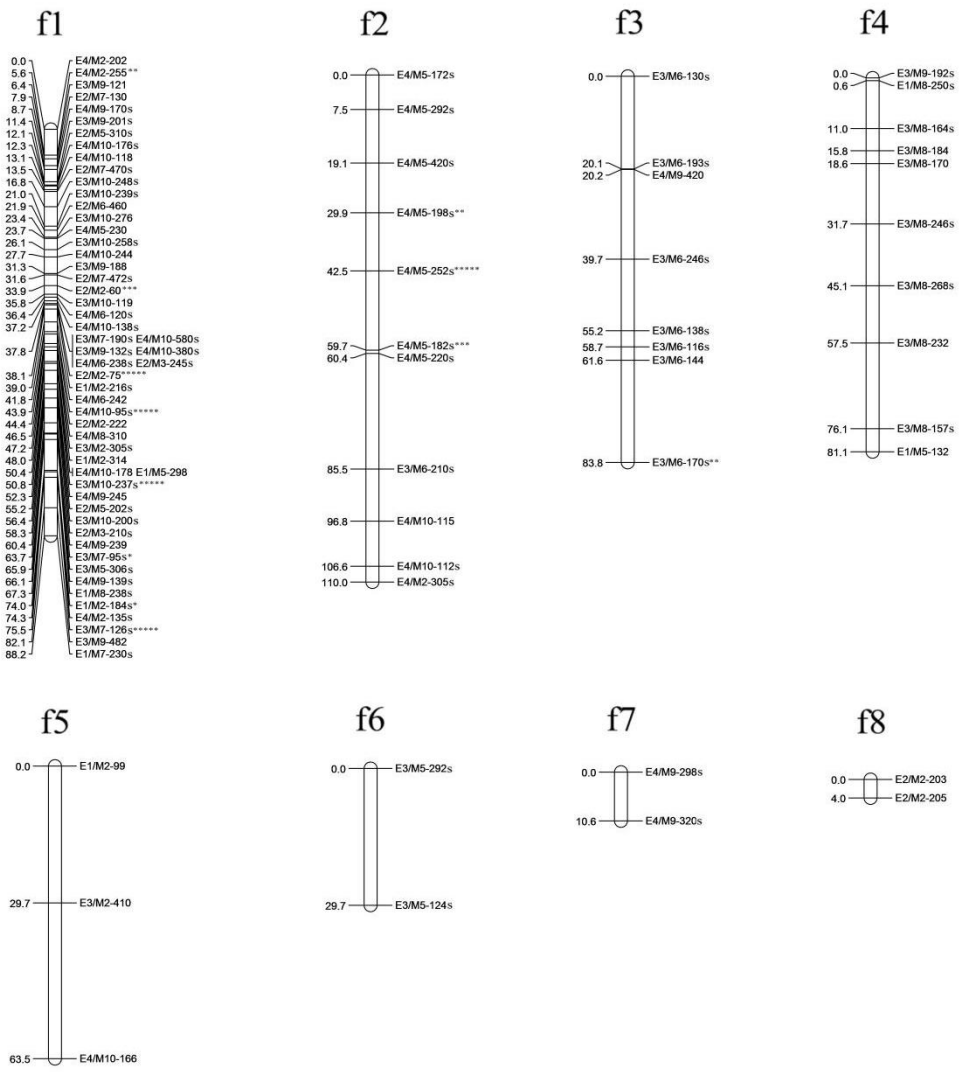


Fig. 10. The genetic linkage map of female parent ‘Fushun’ with AFLP markers. Loci names were given to the right of each linkage group; genetic distances (cM) were to the left. Markers with suffix “s” were heterozygous at both parents. Markers with distorted segregation ratios were marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001).

Linkage Analysis and Genetic Map Construction of Male Parent 'Kemsull'

Following the same mapping strategy as 'Fushun', a genetic linkage map for 'Kemsull' was constructed with AFLP markers which included 152 markers and 15 DS markers, which accounted for 9.9% of all mapped loci and mainly distributed into 4 LGs of k1, k2, k3, and k7. This map covered a total length of 1,053.1 cM and with an average distance of 6.9 cM between adjacent markers. The length of 17 LGs ranged from 17.2 cM (k16) to 130.3 cM (K1) and the number of loci at LGs varied from 2 (k12, k13, k15 and k17) to 53 (k1). The first LG owned the highest marker density (2.5 cM) and also had the longest map distance (130.3 cM). (Fig. 11).

The average distances for more than half of LGs were higher than 10.0 cM. The number of 17 LGs was slight higher than the haploid chromosome number of 15 in tea plant and several small LGs just consisted of two or three loci. Therefore, additional molecular markers were required to incorporate into this genetic map for saturating small LGs and simultaneously concentrating the number of LGs into 15 (Cai et al., 1994; Sankar and Moore, 2001).

Table 10. Summarized data of each linkage group of genetic map for ‘Kemsull’.

LG	No. of marker				Length	Average distance
	1:1 marker	3:1 marker	Total	DS marker ($p < 0.05$)	(cM)	(cM)
k1	24	29	53	3	130.3	2.5
k2	18	5	23	2	124.1	5.4
k3	7	10	17	4	111.6	6.6
k4	4	2	6	1	102.1	17.0
k5	1	7	8	1	79.5	9.9
k6	2	5	7	1	78.1	11.2
k7	5	2	7	3	72.6	10.4
k8	1	7	8	0	79.5	9.9
k9	3	0	3	0	48.0	16.0
k10	1	2	3	0	43.2	14.4
k11	1	2	3	0	38.1	12.7
k12	2	0	2	0	33.0	16.5
k13	2	0	2	0	29.3	14.7
k14	1	2	3	0	27.0	9.0
K15	2	0	2	0	24.1	12.1
k16	2	1	3	0	17.2	5.7
k17	1	2	2	0	15.4	7.7
Total	77	75	152	15	1,053.1	6.9

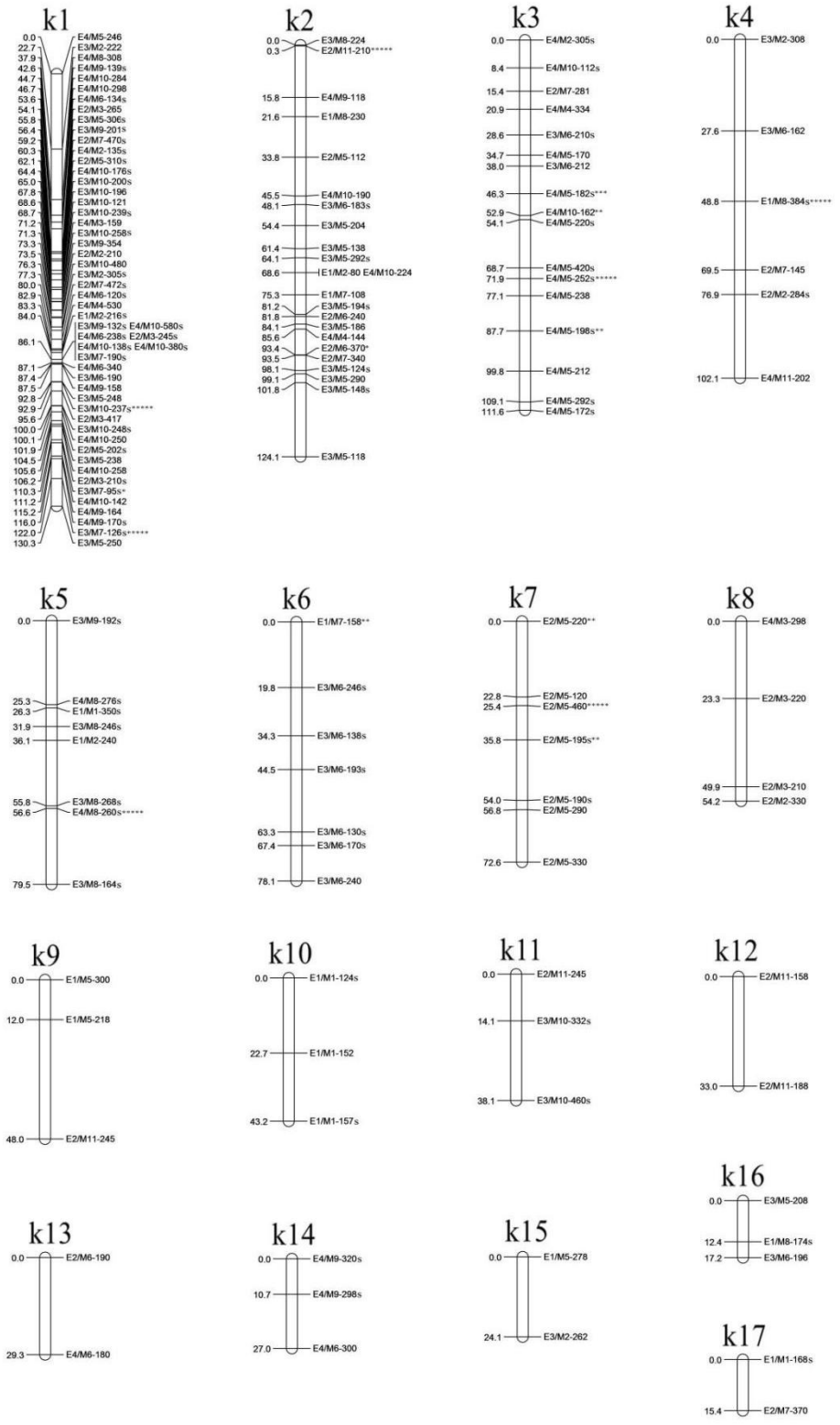


Fig. 11. The genetic linkage map of male parent 'Kemsull' with AFLP markers. Labels and notes were same as Fig. 9.

Genetic Map Construction with RAPD, SSR, and AFLP Markers

For ‘Kemsull’, 379 AFLPs, 109 RAPDs, and 2 SSRs were incorporated into mapping analysis; 308 AFLP markers together with 107 RAPD and 10 SSR markers were applied for the map construction of ‘Fushun’; a total of 649 loci including 495 AFLPs, 143 RAPDs, and 11 SSRs were collectively participated into the construction of a combined map based on ‘One-way method’ (Tavassolian et al., 2010).

Three-line linkage maps of tea varieties ‘Kemsull’ and ‘Fushun’ were constructed using RAPD, SSR, and AFLP markers (Fig. 12). The distribution of markers on three maps was presented in the Table 11. The vast majority of LGs separately obtained from RAPD and SSR data (Fig. 5, 6), and AFLP markers (Fig. 9, 10), were integrated into a combined map after merging all markers as a data set and proceeding linkage analysis with same parameters and mapping strategy. In this combined map, a total of 295 loci including 214 AFLPs, 76 RAPDs, and 5 SSRs were successfully mapped into 15 LGs which corresponded to the basic chromosome number of tea plant (2n=2x=30). The length of 15 LGs ranged in size from 27.0 cM (c15) to 136.9 cM (c6) and the number of loci on each group varied from 3 at c14 or c15 to 86 at c1. The LG of c1 owns the highest marker density with an average interval between contiguous markers of 1.4 cM. For other LGs, the marker densities were mostly less than 10.0 cM with three exceptions of c8, c13, and c14. The vast majority of spaces between two neighboring markers were relatively small, but two large marker intervals higher than 30 cM severally existed in LGs of c11 and c12 (gray-filled LG segments in Fig. 12).

According to 15 LGs of this combined linkage map, some small LGs separately established for each parent could be arranged into one, just as LGs of k1 to c1, k9 to c9. Finally, 228 loci including 165 AFLPs, 62 RAPDs and 1 SSRs were successfully assembled into 15 LGs in the map of ‘Kemsull’, which covered a length of 1,437.5 cM with an average distance of 6.3 cM among two neighboring markers; the map of ‘Fushun’ consisted of 107 AFLPs, 43 RAPDs, and 4 SSRs and possessed a total length of 947.2 cM with a similar average distance among two contiguous marker of 6.2 cM compared to Kemsull’s map (Fig. 12, Table 11).

During recent decades, several genetic maps for tea plants have been developed based on different molecular marker systems around the world. Two preliminary linkage maps for ‘Kemsull’ and ‘Fushun’ appeared a severe clustering phenomenon, which were mainly developed based on a small quantity of RAPD markers (Chang et al., 2015). However, after we incorporated the large number of newly-mined AFLP loci into mapping analysis and

constructed the framework LGs on the basis of non DS markers, those clustered markers mostly separated and positioned together into two or more LGs and with other markers. The generated combined genetic map here, which was built with RAPD, SSR, and AFLP markers using a F1 population of 76 offsprings derived from the cross of two *C. sinensis* var. *sinensis*, owned a length of 1,376.9 cM with 295 markers that was close to the assessments for tea plant genome from some tea varieties, such as 1,550 cM for 'Sayamakari' (Ota and Tanaka, 1999), 1,349.7 cM for 'SFS150' (Hackett et al., 2000), but far shorter than other studies reported 2,545.3 cM – 4,482 cM (Huang et al., 2005; Taniguchi et al., 2007; Hu et al., 2013; Ma et al., 2015). Since some factors including size and type of mapping populations, number and type of mapped markers, mapping strategies, statistical algorithm, and computer packages codetermine the final genetic distance assessment, it is impossible to directly compare these genetic maps generated from different studies. The average interval between two adjacent markers less than 10 cM has been recommended for genome-wide QTL analysis (Doerge, 2002). The marker density of genetic map obtained in this study was 4.7 cM, which might be suitable for QTL mapping study. However, there were three big gaps of larger than 30 cM which respectively existed on the c11 and c12 in the map. The presence of these gaps on the map may be explained as follows. Firstly, the individuals in our mapping population are too finite to reveal the recombination event happened in the gap region. Secondly, two restriction enzymes of EcoR Ⅰ and Mse Ⅰ were selected for carrying our AFLP analysis and then there is a chance that the developed AFLP markers are directed toward AT-rich regions, leaving the absence of markers in GC-regions. Thirdly, the genome regions corresponding to these gaps may be homologous between two parents of mapping population, hence there were no recombination events occurred. Therefore, additional different types of marker were required to try to fill these gaps and saturate in this present map.

The mapping strategy used for map construction of two parents of 'Fushun' and 'Kemsull' was same as the combined map (Fig. 12, Table 11). The majority of located markers presented consistent mapping orders among three maps, such as LGs of k5-c5-f5, k8-c8-f8, k12-c12-h12, k13-c13-f13, k14-c14-h14, and k15-c15-h15 which were completely collinear with each other by comparing common shared markers. The rest of LGs had one or more markers positioned in different map order, and mostly appeared small differences. However, some markers showed severe divergence in orders, such as E2/M5-202s in k1-c1-f1, OPC-06-550s in k4-c4-f4, E3/M8-164s in k7-c7-f6, and E3/M5-114s in k9-c9-f9.

This phenomenon of the divergence of marker orders on genetic maps were also reported in another studies (Tavassolian et al., 2010; Taniguchi et al., 2012; Tan et al. 2013).

For establishing the relationship with a reference map (Taniguchi et al., 2012), 11 anchor markers were utilized for mapping analysis in our study. Among these markers, two loci of MSG0232 and MSG0255 linkaged together as expected and located in the LG of c1, which corresponded to LG03 in the reference map; MSE0316 and MSG0258 were simultaneously positioned at the LG of c13 that was same as in the first LG of the reference map. However, MSG0811 marker were mapped in the LG of c7, which were supposed to link together with MSE0316 and MSG0258. This discrepancy might be caused from different number of alleles obtained from our study, or two LGs of c7 and c13 were just partial segments from a single chromosome, which need to invest through further study.

In the combined genetic map, all 81 distorted segregated markers ($p<0.05$) were annotated with asterisks based on their distortion level. The majority of these markers (75.3%) were clustered on seven LGs of c1, c2, c3, c4, c6, c11, and c12 and the rest of markers were dispersedly distributed into six LGs, c5, c7, c8, c9, c10 and c13. If an allele or alleles at a locus is or are detrimental to gamete or zygote fitness then that locus will cause markers linked to it to derivate from the expected Mendelian segregation ratios. Therefore, any region with a cluster of DS markers showing similar segregation pattern will be considered as a candidate segregation distortion region (SDR) in which one or more genes causing segregation distortion may be contained. Eight candidate SDRs were separately identified on five LGs of c3, c4, c5, c6, c9, c11 and c12 (black-filled LG segments in Fig. 12). For c11, two SDRs severally contained three and five significant DS markers ($p<0.05$), which were located in the middle and terminal sections of this LG, respectively; other six SDRs including six, four, three, three, three, and eight distorted markers were identified at four LGs of c3, c4, c5, c6, c9, and c12 in the combined map, respectively.

Segregation distortion of marker as a common phenomenon has been reported in the mapping analysis for other perennial species, such as *Eucalyptus* (Gratthapagila and Sederoff 1994; Petroli et al., 2012), *Populus* (Yin et al., 2004), apple (Kenis and Keulemans, 2005), and *Citrus* (Cai et al., 1994). The distortion ratio of markers reached up to 77.42% was observed in cotton study (Li et al., 2007). For tea plant, this phenomenon has been frequently reported in the map construction for tea plant (Tanaka, 1996; Ota and Tanaka, 1999; Hackett et al., 2000; Huang et al., 2005; Huang et al., 2006; Mewan et al., 2007; Taniguchi et al., 2007, 2012; Hu et al., 2013; Tan et al., 2013; Ma et al., 2014, 2015). Segregation distortion

may occur by chance alone, mapping population size, or wrong scoring of markers, but the percentage of markers displaying distorted segregation based on these reasons is variable. However, distorted markers that appear in clusters, such as eight SDRs identified on LGs of c3, c4, c4, c5, c9, c11, and c12, may contain genes that affect viability (Lorieux et al., 2000; Cervera et al., 2001; Ma et al., 2014) or fitness. After comparison of homologous linkage groups of the genetic maps of cultivars from the same or different mapping experiments in apple, Kenis and Keulemans (2005) supposed that the segregation distortion of markers sometimes affect only a parent because some traits just exist in this parent. Therefore, further studies are required to investigate these SDRs on the consensus map.

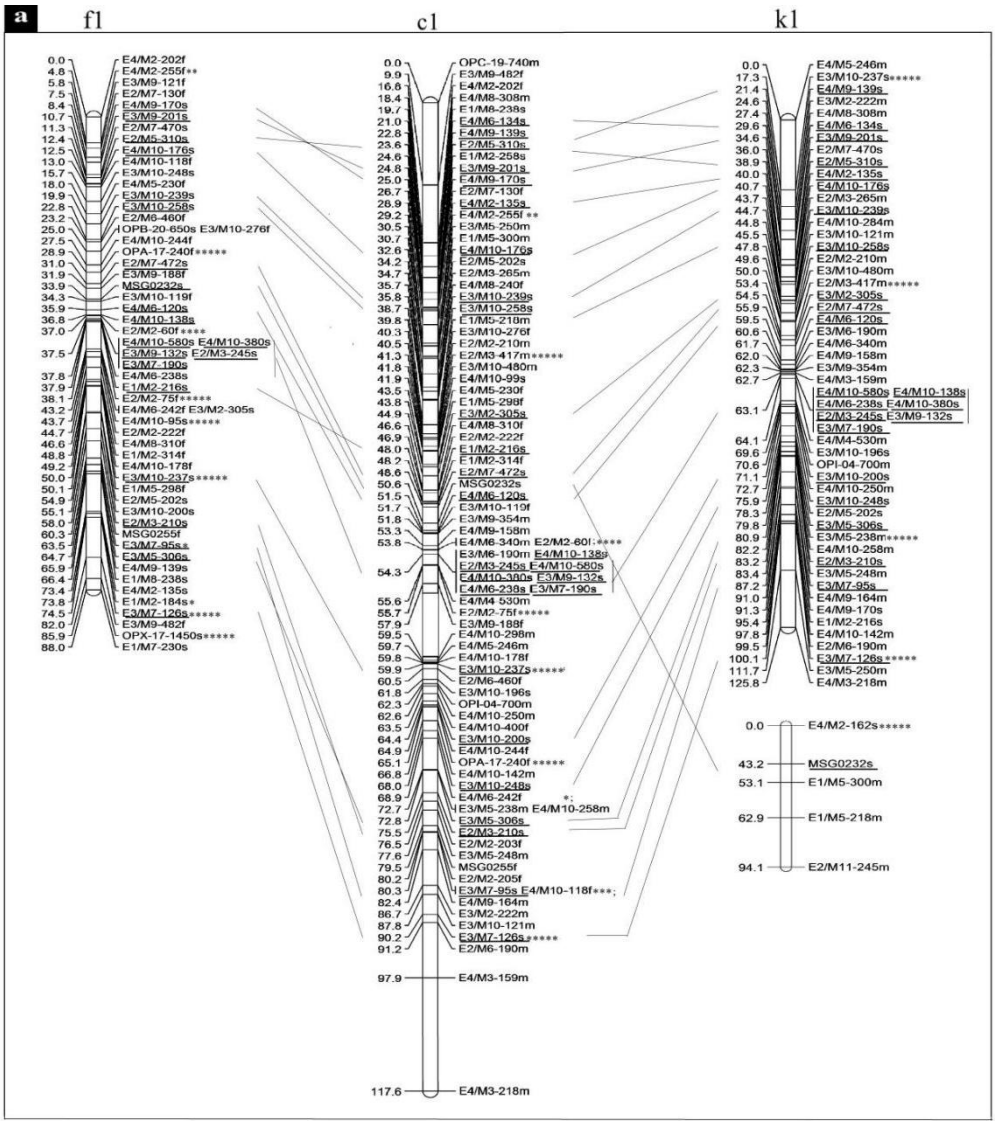


Fig. 12. Genetic linkage map based on RAPD, SSR and AFLP markers for tea plant. LGs were numbered from k1-c1-f1 to k15-c15-f15. The map in the middle (c1-c15) is the consensus map constructed from the combined dataset. The genetic linkage maps on the left (k1-k15) and on the right (f1-f15) are the paternal (Kemsull) and maternal (Fushun) maps, respectively. Loci names were given to the right of each linkage group; genetic distances (cM) were to the left. Regions of linkage groups considered to be candidate segregation distortion region (SDR) were denoted by black fill. Markers with suffix 'f' or 'm' were just heterozygous at female parent or male parent, respectively. Marker with suffix 's' belonged to shared markers of two parents, which was used for denoting collinearity and homology among three maps. Markers with distorted segregation ratios are marked with asterisks according to their significant levels (*: 0.01, **: 0.005, ***: 0.001, ****: 0.0005, *****: 0.0001).

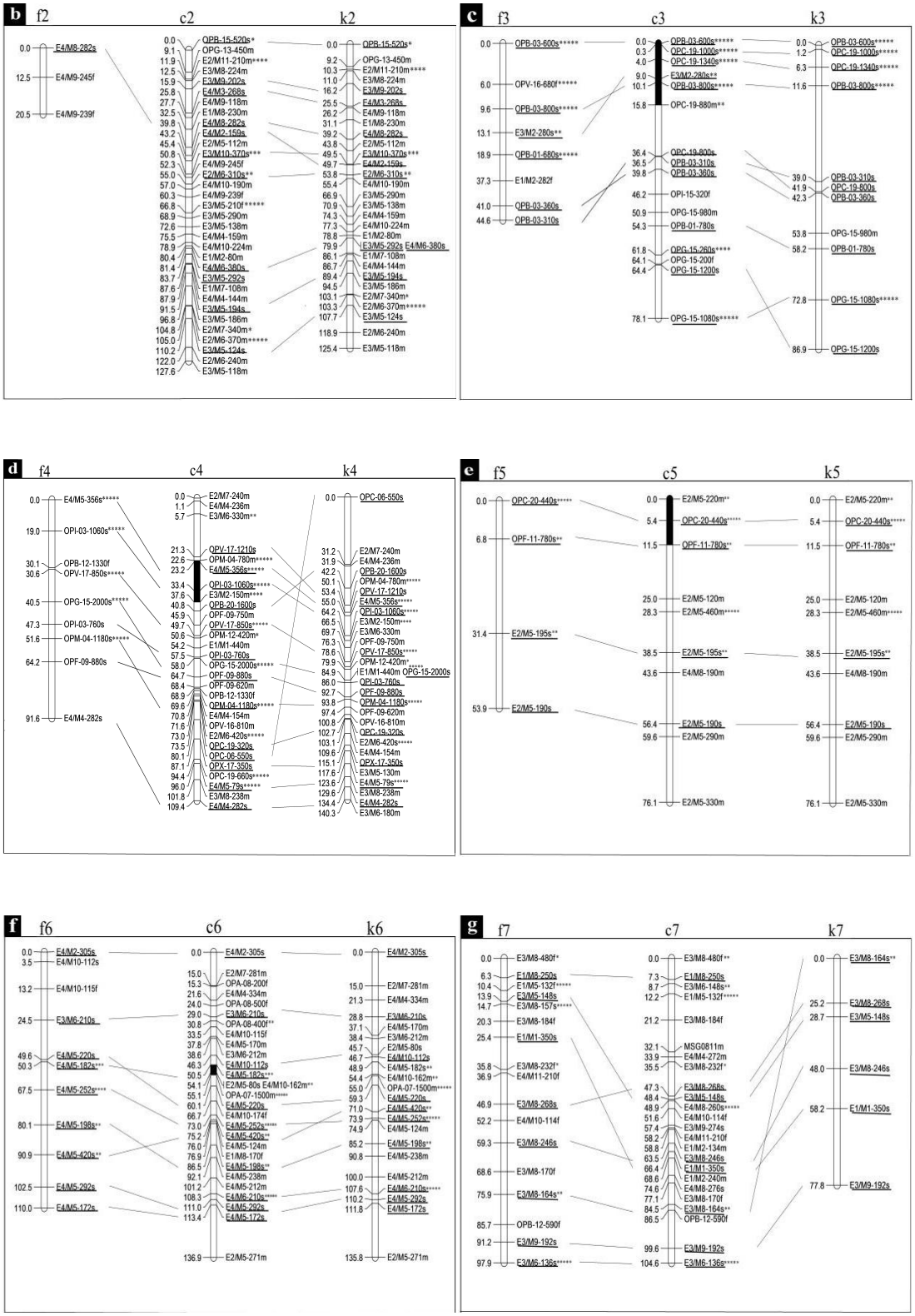


Fig. 12. (Continued)

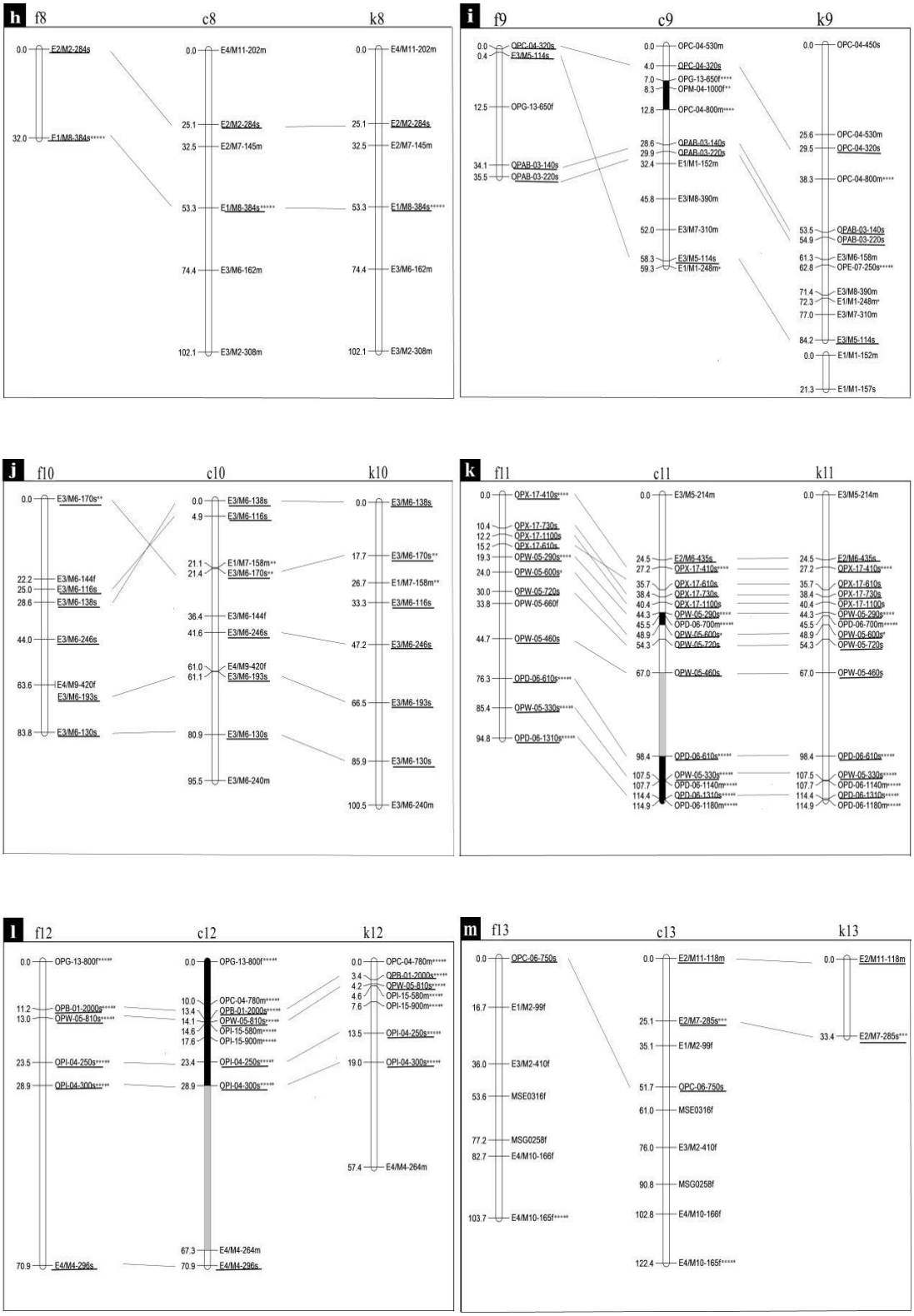


Fig. 12. (Continued)

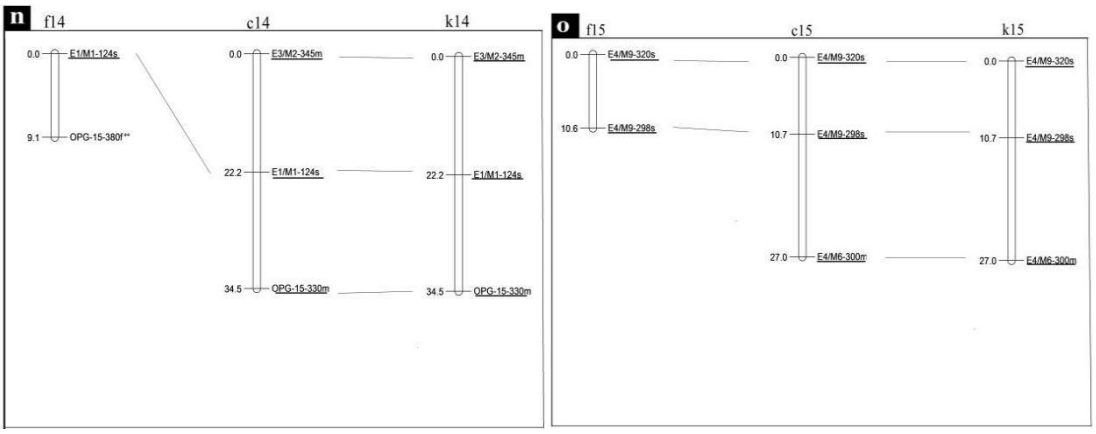


Fig. 12. (Continued)

Table 11. Summarized data of each linkage group of three linkage maps.

LG		Length (cM)	No. of marker	Density (cM)	No. of marker			
					AFLP	RAPD	SSR	DS ($p < 0.05$)
LG01	c1	117.6	86	1.4	81	3	2	9
	f1	88.0	57	1.5	52	3	2	1
	k1	219.9	60	3.7	58	1	1	5
LG02	c2	127.6	33	3.9	31	2	0	7
	f2	20.5	3	16.8	3	0	0	0
	k2	125.4	30	4.2	28	2	0	6
LG03	c3	78.1	16	4.9	1	15	0	8
	f3	44.6	8	5.6	2	6	0	5
	k3	86.9	11	7.9	0	11	0	5
LG04	c4	109.4	29	3.8	11	18	0	12
	f4	91.9	9	9.2	2	7	0	5
	k4	140.3	29	4.8	13	16	0	10
LG05	c5	76.1	10	7.6	8	2	0	5
	f5	53.9	4	13.5	2	2	0	3
	k5	76.1	10	7.6	8	2	0	5
LG06	c6	136.9	28	4.9	24	4	0	8
	f6	110.0	11	10.0	11	0	0	4
	k6	135.8	22	6.2	21	1	0	7
LG07	c7	104.6	24	4.4	22	1	1	6
	f7	97.9	17	5.8	16	1	0	6
	k7	77.8	6	13.0	6	0	0	1
LG08	c8	102.1	6	17.0	6	0	0	1
	f8	32.0	2	16.0	2	0	0	1
	k8	102.1	6	17.0	6	0	0	1
LG09	c9	59.3	12	4.9	5	7	0	4
	f9	35.5	5	7.1	1	4	0	0
	k9	105.5	14	7.5	7	7	0	2
LG10	c10	95.5	10	9.6	10	0	0	2
	f10	83.8	8	10.5	8	0	0	1
	k10	100.5	8	12.6	8	0	0	2

LG11	c11	114.9	16	7.2	2	14	0	9
	f11	94.8	12	7.9	0	12	0	6
	k11	114.9	16	7.2	2	14	0	9
LG12	c12	70.9	10	7.1	2	8	0	8
	f12	70.9	6	11.8	1	5	0	5
	k11	57.4	8	7.2	1	7	0	7
LG13	c13	122.4	9	13.6	6	1	2	2
	f13	103.7	7	14.8	4	1	2	1
	k13	103.7	7	14.8	4	1	2	1
LG14	c14	34.5	3	11.5	2	1	0	2
	f14	9.1	2	4.6	1	1	0	1
	k14	9.1	2	4.6	1	1	0	1
LG15	c15	27.0	3	9.0	3	0	0	0
	f15	10.6	2	5.4	2	0	0	0
	k15	27.0	3	9.0	3	0	0	0
Total	c	1,375.9	295	4.7	214	76	5	81
	f	947.2	153	6.2	107	42	4	48
	k	1,437.5	228	6.3	165	62	1	60

Literature Cited

Cai Q, Guy CL, Moore GA (1994) Extension of the genetic linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation responsive loci. Theor Appl Genet 89:606-614.

Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Slycken JV, Montagu MV, Boerjan W (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. Nigra* and *P. Trichocarpa*) based on AFLP and microsatellite markers. Genetics 158:787-809.

Chang YL, Ou EN, Min SL, In KS, Joo Y, Kwan JS (2015) Preliminary linkage mapping of tea plants with RAPD and SSR markers. J Kor Tea Soc 89-94.

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43-52.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137.

Hackett CA, Wachira FN, Paul S (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85(4):346-355.

Huang JA, Li JX, Huang YH, Luo JW, Gong ZH, Liu ZH (2005) Construction of AFLP molecular markers linkage map in tea plant. J Tea Sci 25 (1):7-15.

Huang FP, Liang YR, Lu JL, Chen RB (2006) Genetic mapping of first generation of backcross in tea by RAPD and ISSR markers. J Tea Sci 26(3):171-176.

Hu CY, Lee TC, Tsai HT (2013) Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). Euphytica 191:141-152.

Hou YJ, Chang YL, He Q, Tang M, Wu XH (2015) Study on tea breeds or new materials by AFLP molecular markers. Southwest China J of Agr Sci 02:514-518.

Kamunya SM, Wachira FN, Pathak RS, Korir R, Sharma V, Kumar R, Bhardwaj P, Chalo R, Ahuja PS et al. (2010) Genomic mapping and testing for quantitative trait loci in tea (*Camellia sinensis* (L.) Kuntze). Tree Genetics & Genomes 6(6): 915–929.

Kenis K, Keulemans J (2005) Genetic linkage maps of two apple cultivars (*Malus domestica* Borkh.) Based on AFLP and microsatellite markers. Mol Breeding 15: 205-219.

Li W, Lin ZX, Zhang XL (2007) A novel segregation distortion in intraspecific population of asian cotton (*Gossypium arboretum* L) detected by molecular markers. J Genetics and Genomics 34(7):634-660.

Lorieux M, Ndjiondjop MN, Ghesquiere A (2000) A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite based genetic linkage map. Theor Appl Genet 100: 593-601.

Ma JQ, Zhou YH, Ma CL, Yao MZ, Jin JQ, Wang XC, Chen L (2010) Identification and haracterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). American J Botany e153-156.

Ma JQ, Yao MZ, Ma CL, Wang XC, Jin JQ, Wang XM, Chen L (2014) Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (*Camellia sinensis*). PLoS ONE 9(3): e93131.

Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, Zheng HK, Yao MZ, 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.

Mewan KM, Saha MC, Konstantin C (2007) Construction of a genomic and EST-SSR based saturated genetic linkage amp of tea (*Camellia sinensis* L.). The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science(ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Ota S, Tanaka J (1999) RAPD-based linkage mapping using F1 segregating population derived from crossings between tea cultivar 'Sayamakaori' and strain 'Kana-Ck17'. Breeding Res (1):16.

Paul S, Wachira FN, Powell W, Waugh R (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea revealed by AFLP markers. Theor Appl Genet (94):255-263.

Petroli CD, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, Myburg AA, Silva Jr OB, Pappas JR GJ, Kilian A et al. (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. PLoS ONE 7: e44684.

Sankar AA, Moore GA (2001) Evaluation of inter-simple sequence repeat analysis for mapping in *Citrus* and extension of the genetic linkage map. Theor Appl Genet 102:206-214.

Stam P (1993) Construction of intergrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3: 739-744.

Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM et al. (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). PloS ONE 8(11): e81611.

Tanaka J (1996) The possibility of using RAPD analysis to create genetic map of tea. Tea Res J 84:44-45.

Taniguchi F, Tanaka J, Kono I (2007) Construction of genetic linkage map of tea using SSR markers. The Organization Committee of the 2007 ICOS. Proceedings of the 3rd International Conference on O-Cha (tea) Culture and Science(ICOS). Shizuoka:The Organization Committee of the 2007 ICOS.

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, Fukuoka H, Tanaka J (2012) Construction of a high-density reference linkage map of tea (*Camellia sinensis*). Breeding Sci 62:263-273.

Tavassolian I, Rabiei G, Gregory D, Mnejja M, Wirthensohn MG, Hunt PW, Gibson JP, Ford CM, Sedgley M et al. (2010) Construction of an almond linkage map in an Australian population Nonpareil × Lauranne. BMC Genomics 11:551.

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Wachira F, Tanaka J, Takeda Y (2001) Genetic variation and differentiation in tea (*Camellia sinensis*) germplasm revealed by RAPD and AFLP variation. Journal of Horticulture Science and Biotechnology 76(5):557-563.

Yin TM, Difazio SP, Gunter LE et al. (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theor Appl Gene 109:451-463.

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 94102629, 7 (Publication No: 05348A1) Paris: Europe.

Zhang SM (2009) Applications of AFLP and SSR technologies in genetic diversity analysis of Chinese bayberry. Zhejiang University, Doctoral thesis.

CHAPTER Ⅲ. DEVELOPMENT OF RNA-SEQ BASED SSR MARKERS AND INCORPORATION INTO A GENETIC LINKAGE MAP WITH RAPD, SSR, AND AFLP MARKERS FOR CHINA TYPE TEA PLANTS

Abstract

A moderately saturated genetic linkage map with 76 RAPD, 5 public SSR, and 214 AFLP markers were established for tea plant, which covered a map length of 1,376.9 cM and with an average interval of 4.7 cM between two adjacent markers. In this study, more informative markers were developed and then incorporated into this map. Using high throughput RNA sequencing, we analyzed and compared the transcriptome sequences of flowers and leaves of 'Fushun' (*Camellia sinensis*) and 'Kemsull' (*C. sinensis*). Afterwards, 1,800 potential polymorphic SSR markers were successfully developed and 296 of them were selected and experimentally validated with a subset of tea plants (including two parents and six F1 offspring), from which 75 (25.3%) were repeatably amplified and also polymorphic between two parents. From that, 29 (38.7%) newly mined SSR markers which were heterozygous in 'Fushun' and /or 'Kemsull' and showed segregated genotypes in F1 seedlings, and also adoptable by JoinMap 4.0, were attempted to incorporate into a combined genetic linkage map that consisted of 76 RAPDs, 5 public SSRs, and 214 AFLPs. Finally, 11 of them were successfully merged into this map. The new genetic map covered 1,441.6 cM and 309 markers with an average distance of 4.7 cM between two adjacent markers and included 79 RAPDs, 5 public SSRs, 214 AFLPs, and 11 new SSRs mined by RNA seq technique. This map will lay a foundation for qualitative or quantitative trait loci (QTLs) analysis of important agronomic traits for tea plant in the future study.

***Additional key works*:** *Camellia sinensis*, genome map, SSR, RNA-seq, tea.

Introduction

Tea plant [*Camellia sinensis* (L.) O. Kuntze] is one of the most popular non-alcoholic beverage crops around the world. Tea plant is self-incompatible and highly heterozygous and has a long juvenile period (Chen et al., 2012). Those characteristics have reduced the breeding efficiency of tea plant. Normally, around 20 years are required to obtain a new tea cultivar through conventional breeding methods (Chen et al., 2007). The marker assisted selection (MAS) has been designed as a new strategy for genetic improvement of crop which combines molecular marker with conventional breeding technology (Collard et al., 2005). And the key for MAS is to conduct the fine qualitative or quantitative trait loci (QTLs) mapping of desirable agronomic traits, for which a saturated genetic linkage map is required to primarily construct with a plentiful of available molecular markers and a suitable mapping population. Therefore, the need for mining codominant molecular marker quickly and inexpensively with the development of molecular biotechnology has become urgent for developed high density genetic map.

Next generation sequencing (NGS) techniques are revolutionary technologies which can generated millions of sequences at a relatively low cost compared with traditional methods (Zalapa et al., 2012). During recent years, the NGS-based RNA-seq method has greatly strengthened the transcriptome analysis, particularly for non-model perennial plants such as tea plant. The transcriptome analysis will be helpful to uncover the molecular mechanisms in the specific biological progress or under certain physiological condition. The advanced RNA-seq technique provides a fast, cost-effective, and reliable approach to generate large expression dataset for functional genomic researches. Using this technique, the transcriptome of *C. sinensis* was analyzed at an unprecedented depth (2.59 Giga base pair) and many genes for several major metabolic pathways were verified and explored from this extensive transcriptome dataset (Shi et al., 2011). Zhang et al. (2016) conducted the transcriptome analysis for two kinds of styles using this method after cross- and self- pollination treatment and offer novel insights to the biological phenomenon of self-incompatability in tea plant.

Transcriptome analysis has been conducted to analyze the molecular mechanisms underlying albino phenomenon of elite albino tea cultivar ‘Anji Baicha’ and revealed that the main differential metabolisms between the albescent stage and the green stage were proceeded in photosynthesis and the phenylpropanoid and flavonoid biosynthetic pathways (Chen et al., 2016). DNA-binding with one finger (Dof) transcription factors (TFs) have

significant roles in abiotic stress tolerance of plant. Li et al. (2016) identified a total of 29 putative Dof TFs based on transcriptome study for tea plant and concluded that abiotic stress can cause the changes of gene expression levels. To elucidate the molecular mechanisms of the defense responses to *Ectropis oblique* attack in tea plants, the different transcriptome analysis of tea leaves was performed using RNA-seq technique, which will be useful to create novel prevention measures against the *E. Oblique* (Wang et al., 2016).

Furthermore, the RNA-seq data are valuable resources for marker development such as SSR and SNP markers. These generated molecular markers have a high transferability between all related species and will be very powerful in the comparative mapping. Luro et al. (2008) reported that 41 genic-SSR markers developed from transcriptome data were mostly originated from untranslated region at the 5' terminus and showed high level of transferability between those related species in *Citrus*. Zhang et al. (2012) also reported that 32 polymorphic genic-based SSR markers were identified from the transcriptome data of *Sesamum indicum* L. and 14 of them have been successfully integrated into 9 genetic maps. For tea plant, Tan et al. (2013) studied that an analysis of *C. sinensis* floral transcriptome could successfully pre-selected 2,439 candidate SSR markers using high-throughput Illumina RNA sequencing. Of them, 720 were experimentally tested and 431 (59.9%) were validated as novel polymorphic SSR markers for tea plant. Subsequently, a consensus SSR-based linkage map was constructed which covered 1,156.9 cM and included 237 SSR markers evenly distributed into 15 linkage groups.

In our study, we tried to directly develop polymorphic SSR markers from the transcriptome sequences comparison between 'Fushun' (*C. sinensis*) and 'Kemsull' (*C. sinensis)* using high throughput RNA-seq technique and evaluate the efficiency of this strategy through verification and application steps. Finally, we tried to incorporate these SSR markers into a previously developed linkage map that consisted of 74 RAPD, 5 public SSR, and 214 AFLP markers and with a map density of 4.7 cM.

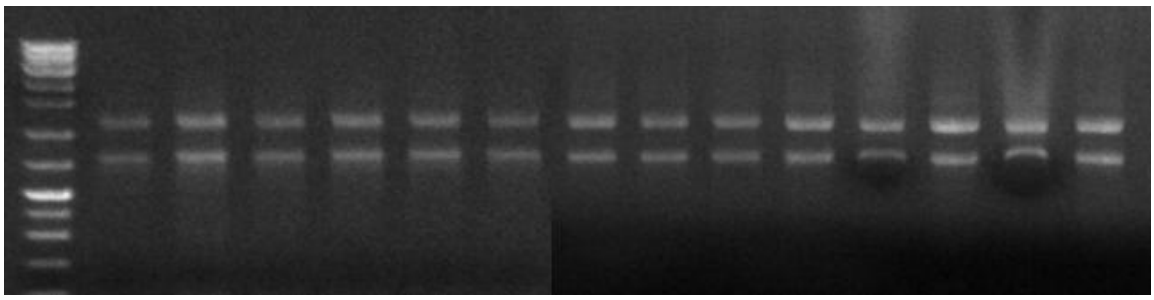
Materials and Methods

Plant Materials and RNA and DNA Isolation

For RNA isolation and RNA-seq analysis, several fresh leaves and intact flowers at the balloon stage were separately collected from female parent 'Fushun' (*Camellia sinensis*) and male parent 'Kemsull' (*C. sinensis*) and immediately snap frozen in liquid nitrogen. Total RNA was extracted from each plant organ using the Quick-RNA™ MiniPrep kit (ZYMO RESEARCH, USA) according to the manufacturer's instruction with some modifications. The quantity and quality of the purified RNAs were checked using DeNovix DS-11+ Spectrophotometer (DeNovix, Wilmington, DE, USA) and agarose gel electrophoresis (Fig. 13), respectively. The quality of RNA was checked by Bioanalyzer 2100 (Agilent, Santa Clara, CA).

For DNA marker analysis and linkage mapping, the same mapping materials from chapter Ⅰ were applied into this study. It should be noted that nine F1 individuals including No.13, No.22, No.23, No.32, No.34, No.39, No.40, No.45, and No.68 lost their whole plants during this study. The preparation and isolation of genomic DNA from two parents and F1 seedlings were same as chapter Ⅰ.

a



b

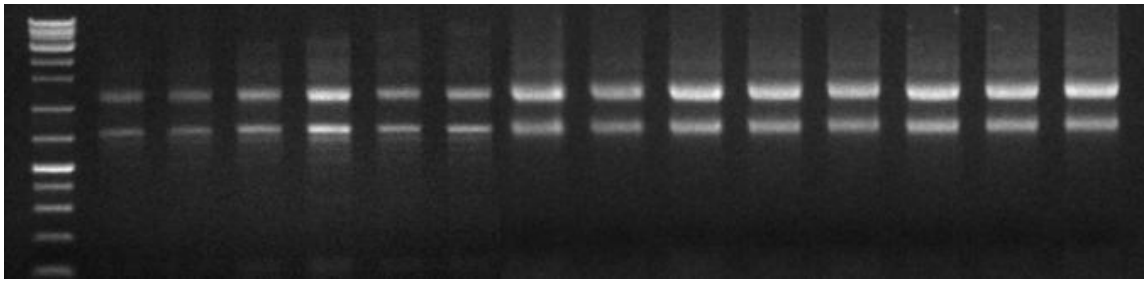


Fig. 13. Electrophoresis results of RNA samples of leaves and flowers of ‘Fushun’ (*Camellia sinensis*) and ‘Kemsull’ (*C. sinensis*). Notes: a, lane 2 to lane 7 were RNA samples of leaves of ‘Fushun’, lane 8 to lane 15 were RNA samples of flowers of ‘Fushun’; b, lane 2 to lane 7 were RNA samples from leaves of ‘Kemsull’, lane 8 to lane 15 were RNA samples of flowers of ‘Kemsull’; lane 1 was DNA ladder marker 100.

Library Preparation, RNA-seq Assembly and SSR Detection

Each 5 ug of total RNA from the leaves and flowers of each cultivar was mixed to construct library using TruSeq RNA sample preparation kit according to the manufacturer's protocol (Illumina, San Diego, CA, USA). RNA sequencing was performed to obtain approximately 3 gigabytes of sequence information for each cultivar using Illumina HiSeq 2500 sequencing platform by Theragen Etex (Theragen Etex Co., Ltd., Suwon, Korea; http://www.theragenetex.com/). For the pre-processing of the sequencing reads, the reads were trimmed to remove low quality of reads and adaptors using DynamicTrim and LengthSort programs of SolexaQA package (ver. 1.13; Cox et al., 2010). *De novo* assembly of transcriptome was performed from the cleaned reads using Velvet (ver. 1.2.08; Zerbino and Birney, 2008) and Oases (ver. 0.2.08; Schulz et al., 2012). The MIcroSAtelitte (MISA) was used to identify SSR from the assembled contigs (http://pgrc.ipk-gatersleben.de/misa/). To find the presence of SSRs, only 2 to 10 nucleotides motifs were considered, and the minimum repeat unit was defined as 6 for di-, 5 for tri-, 4 for tetra-, 3 for penta-, and 2 for hexa-, hepta-, octa-, nona- and deca-nucleotides. The information obtained by MISA was used for designing primers flanking the SSR motif. To design primers flanking the SSR loci, the software Primer3 (ver. 2.3.5) were used (Untergrasser et al., 2012). The following parameters were used: primer length 18–24 bp, with 20 bp as the optimum; primer GC% = 20–80%, with the optimum value being 50%; primer Tm 55–65℃, with 60℃ as the optimum, and product size range 150–500 bp.

***In silico* Mining of Polymorphic SSRs**

In silico polymorphism analysis of SSR markers was performed using the virtual PCR approach where pairs of primer sequences from reference were mapped onto the consensus sequence of each sample (Li and Durbin, 2009). The SSR primer was then identified as candidates for marker development if they had sufficient sequences on both sides of the SSR repeats for primer design. And the primer aligned only one was selected as candidate for SSR marker. The *in silico*-generated amplicons were compared with the expected amplicon size between samples differences were recorded. If an amplicon size differed by at least 1 bp, the SSRs was classified as polymorphic, while amplicons of identical size were considered as monomorphic.

Validation of SSR marker and SSR Analysis

296 sets of SSR primer pairs were preferably selected from all potential polymorphic SSR loci and screened against two parents and/or six their F1 seedlings. Only informative markers, which were heterozygous at one or two parents and also showed segregated genotypes in F1 seedlings, were used to genotype the whole mapping population. These SSR loci were characterized with the name beginning of "KF". The PCR reaction solution (20 μL) contain: 20 ng genome DNA, 0.5 μM each primer, 1× PCR buffer, 0.2 mM per each dNTP, and 1 U Taq DNA polymerase (Dongsheng Biotech, Guangzhou, China). The PCR program was as follows: 5 min at 94°C; 34 cycles of 30 s at 94°C, 1 min at 55°C , 30 s at 72°C; 30 min at 72°C to allow double-strand DNA synthesis. The PCR amplicons were diluted with 0.5 volume of loading dye [98% (v/v) formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene xyanole FF] and completely denatured 5 min at 95°C. The denatured samples were loaded on a 8% denatured poly-acrylamide gel and visualized after silver staining. The electrophoresis was carried out with the double and triple wide mini-vertical gel system (MGV-202-33, C.B.S Scientific, CA, USA).

Map Construction

Linkage maps for each parent were constructed based on RAPD, public SSR, and AFLP markers, and newly mined SSR markers with a 'cp' population type by JoinMap 4.0 version (Stam, 1993; Van Ooijen, 2006). Firstly, the segregation data were sorted into two data sets derived from each parent before the linkage analysis. Both sets of genotyping data were loaded into JoinMap 4.0 and then the two maps of parents was separately constructed. Chi-square analysis was performed for goodness of fit to the expected Mendelian segregation ratio for each marker. Framework linkage groups were established by omitting the skewed markers from the data. These framework groups were treated as fixed orders for each map construction and then the skewed markers were tried to be incorporated if they could not noticeably affect surrounding markers positions. The linkage groups were generated at a LOD score of 3.0 and a recombination frequency ≤ 0.40. The Kosambi mapping function was chosen to transfer recombination frequency into map distance. Secondly, a combined genetic linkage map was constructed by merging all the markers heterozygous in both or either of the parents as one set of data with the same strategy which was known as 'One-way method' Tavassolian et al. (2010).

Results and Discussion

SSR Markers Development

A total of 1,800 SSR primer pairs were successfully designed from the flanking regions of long SSRs using Primer 3. Of them, 296 primer pairs were synthesized for validation with a set of tea plants, including two parents 'Fushun', 'Kemsull', and/or six their F1 seedlings. As a result, 200 (65.6%) primer pairs yielded PCR amplification products and 75 (25.3%) amplified the repeatable and polymorphic bands which were considered as new informative markers for mapping analysis for tea plant (Table 12). The successful rate (25.3%) was much lower than the other similar study in tea plant (Tan et al., 2013), in which a total of 2,439 SSR primer pairs were successfully designed, 431 (59.9%) out of 720 SSR markers were polymorphic among eight representative tea cultivars. This could be explained with the amount of samples and the preparation of a high representative RNA pool for preparing a cDNA library. In their tudy, hundreds of flowers were collected and then dissected into three parts (petals, pistils, and stamen) and frozen into liquid nitrogen, then total RNA from three plant parts were separately extracted, respectively, equal amounts of total RNA from three plant parts were pooled together and gently mixed well to prepare a representative RNA pool. For our study, a few leaves and several flowers were collected, and then equal RNA volume isolated from tea leaves and whole flowers were pooled together as a representative RNA pool to prepare a cDNA library. Nonetheless, the new SSR markers reported here could be useful into our mapping analysis even though low efficiency. Thus, these newly developed genic-based SSR markers may help to identify candidate genes and increase the efficiency of marker-assisted selection (MAS).

Of 75 newly mined SSR markers, only 29 (38.7%) were subsequently used to genotype in the whole mapping population, because they were not only heterozygous in 'Fushun' and/or 'Kemsull' but also adoptable by JoinMap 4.0. The remaining 49 markers were excluded because they did not have corresponding codes in JoinMap even though successful PCR amplification and alleles segregation or the alleles did not segregate in the F1 seedlings (Table 12). Figure 14 displayed partial agarose gel profiles of five SSR pair primers.

Previously, 74 individuals of the whole mapping population were ascertained to be true hybrids by means of five SSR loci of MSG0206, MSG0232, MSG0255, MSG0699 and TM103, and the remaining two seedlings (No. 20 and 62) could not be confirmed. Here, the authenticity of No.20 seedling as F1 hybrid crossed from 'Fushun' (*C. sinensis*) and

‘Kemsull’ (*C. sinensis*) could be confirmed by a SSR locus KF0434. And for No.62 plant, any of seven SSR loci including KF0036, KF0043, KF0265, KF0340, KF0365, KF0434, and KF0545 could ascertain its authenticity as true F1 hybrid derived from two parents.

Table 12. RNA-seq based SSR markers being polymorphic between ‘Fushun’ and ‘Kemsull’.

ID[a]	Forward primer (5'-3')	Reverse primer (5'-3')	SSR unit type	Band size[b]	Genotype (f-k)[c]	Notes[d]
KF0007	TTTCACCAGGACGAATGCACTA	CACAGAATGCAGTGCTATGAAA	(ACAATA)2	194	aa-ab	
KF0036	TATAAACCACCTCGCGAGACTC	GGGGCTGGTGTTCTTCATAGAA	(TCA)8	362	ac-bd	
KF0043	TTTGCTCTGCATTTTGCTGC	ATGCTGCTTGAGAGACCATGAA	(ATCAAA)6	260	ab-cd	
KF0055	ATGTGGTTGGTTCTGACTCGAT	ACTTGCCGGGTTAAGCGTATTA	(TTC)5	249	abc-cc	
KF0073	CTTGTTCCCTCCTATGCCCTTT	ATGGCGCAAAATCATCATACGG	(CCACCT)4	192	ab-bcd	
KF0097	CCTCCTACACAGCATCACCATT	TAATCTCACTCCTTTCGGTGGC	(CACCGC)2	206	cc-ab	
KF0112	GTAGAAATCGCCGAAATCACCG	AACTCTTTGCTCCAGCTCATCA	(ACCCGG)2	212	ab-ac	
KF0120	TGGGGCGTTGAACAATCATCTA	CCTGCCTGAATCACAAAACCAG	(TTGATG)2	194	ab-ab	
KF0134	TACCAGCGTGTCCTAAAACTCC	CCTAACGAAGCAAAAACGAGCA	(CTAGCC)3	250	aa-bb	
KF0141	GCATGAAGAGCTCGCCTTTAAG	ACCTTCCATGGCATTCGAATCT	(AAG)7	208	ab-ab	
KF0147	CCTTGGCTACAAAGAGCAGAGA	GACCTTGCTAAAGTTGGTGCAG	(CAACAG)2	387	ac-ab	
KF0166	ATCAGCTCGAATCAAACCCCTT	TTTGGGTTTTGCGGGGTTAATC	(TCGCTC)2	189	cd-ab	PCR amplicons were unclear and most F1 seedlings were genotype ‘abcd’.
KF0191	GCAGAAAGAAAGGGGCGAAAAT	GCTTCATCTACTTCGTCTCCGT	(AAG)5	202	abcd-ac	
KF0213	AGATGTAACTCTAGCAGTGGC	GGAAAGGCTAATCGGGATAGGG	(ACC)7	279	bb-abc	
KF0217	AGCCTCATGCCCAAATCTTGTA	GCAACTGGATGATGGTGATGTG	(TTTC)9	399	aa-ab	PCR amplification failure.
KF0231	AAAACTTCCCGATTCCTCCTCC	TCAGACAATGGACGCTGATGAA	(TCT)7	398	ab-aa	PCR amplification failure.
KF0234	CACTCTCTCTCACAGAAACCCC	ACGTGTTTCTTCGCTAGTGGAT	(TACGAC)3	273	aa-ab	
KF0236	TCAAAATCGAGTGCGCCCTATA	AGGGATGACCAAACAGGATTCC	(AGG)5	268	abc-ac	

KF0251	CCAACCCAACCTAGTCTACCAC	AAAGTGGATGTGTCCCTCTTCC	(GGCCAT)4	297	ac-bc	PCR amplification failure.
KF0265	GTGTTCTTCAGTGATCCCGAGT	TCTGGTTTCGCATCTTGCTAGT	(GGT)6	373	aa-bc	
KF0269	CACCTCCTCCTCAAAAGGTCTC	GGGGACACACATGCCTATTACT	(AACACA)3	398	aa-ab	PCR amplicons were undistinguishable.
KF0271	TGCAACAGCCTTCATACTGCTA	GCCGCATGCTGTTAAAAGCTAA	(GCAG)4	344	bb-ab	
KF0288	TCATGGGTCATTCATCGTTGGT	TGGAATTCAAAGTGAGCTCCGA	(AACCCT)2	361	ab-aa	
KF0289	CCCAAATCCCAAGCAGTAGAGT	TGCTAACAATGGAAGCATCCCT	(GTA)6	228	ac-bc	PCR amplification failure.
KF0290	AGAGTGGGCTTTGATGATTCGT	CCACCCCTTCGAAAATTATAACCA	(GAGAT)3	212	bb-ac	PCR amplification failure.
KF0304	CTGGAGGCTGAATTGAGGAAGT	TTCTTCACAGGACAAGCACAGT	(CGA)5	356	ab-bb	PCR amplification failure.
KF0305	AGACAAAGCTATGACCCTGGTT	CTCCAGAAAACCCATCCACAGA	(TGAGT)3	260	ab-aa	
KF0311	GCGTTGACAATGCTAACTTGGT	ACCTGTCCATGTTCCCTTTTGT	(AATGAACT)2	214	aa-ab	PCR amplification failure.
KF0320	CTCTCCCTGGGCTTGAATGAAT	GAACTCACAGCAACAACATCCC	(TTG)7	303	ac-ab	PCR amplification failure.
KF0340	CAAAGTGGAATGCACCGATTGT	ATTGGCTTTCAATCCAGCTATGG	(TATTAC)3	207	ac-bb	
KF0350	TTTTTCATCTTCCGAATGGCGC	GGCTATTTGCTGGGTTCCAAAG	(CGG)5	228	aa-ab	
KF0353	AATTCCCACCACGACTCTCTTC	TGATTCCAACTCCGACAACGAT	(AGGGTT)4	233	cd-ab	Most F1 seedlings were ‘abcd’ genotypes.
KF0365	GTGACTCCGTAGATGGTGATCC	CTTGGCCCTCTACATCCCTTTT	(TGATGT)8	420	ab-cc	
KF0382	AGCTCAAGGATTGGTTTCCGAT	CACCCAGAGAGTGCAGTAAAGT	(TTTCT)5	388	aa-ab	
KF0387	GGTTTCCCCCTTGTATGGCTTA	AGGCCTAAACTTCAAACCCTGT	(CTGATG)2	398	ac-ab	PCR amplification failure.
KF0391	TCTACCTCCTTAGAACCCTGCA	GTCGTCGTAGTTCTCGAAGTCA	(CCA)5	319	bb-ab	PCR amplification failure.
KF0403	AAGGTGGCCATCATGTTCATCT	GCTGAGACTCTGTCGCAATTTC	(CAA)7	187	abcd-abcd	
KF0411	TTACCATCCTCCTCCACCTGAT	CACTCAAGCAAACAGAGCAGTC	(TCC)10	380	aa-bb	
KF0413	GCATCTACTGACATTGCACGTC	TGCACATTTCTTGAGATGTTTACT	(GTTGATT)2	191	abcd-cde	

KF0416	AGCAAAGAGTCAGCTGGATCAA	AGAGAGACAGAGAGATGGGCAA	(AATCCT)4	261	bb-ab	
KF0434	GGAAATTCTGTTGTTGACCGCA	ACCAGCTGAGGATGTTGAGAAG	(CCCTAA)4	344	bc-aa	
KF0453	TTCCCACCACCATACTCTTTCG	CTGAGGAGACCCAAACTGTAGG	(TGCCCT)4	187	ab-bc	
KF0468	CTAAATCGGCTTCGTCGTTGTC	TCTTACTGTTCTTCGGTTCCGG	(ATCAGA)4	315	bb-ab	
KF0502	TGAGGTTCAATCGGCTACACAA	GGAGGTTCCCCAAATGGAGAAT	(GAGTGG)4	231	ab-bcd	
KF0503	AACAATCACTGAGGTCGATGCT	AACCCATCCCAGAAACAGATCC	(GGT)9	365	ab-ac	PCR amplification failure.
KF0512	TGCACCTCCAAACCCAATTACT	CCTCCCCAAGCTCATACTTGTT	(GGAGGT)2	436	bb-ab	
KF0515	ATTCCCCATGTTTGCATACCCT	AGCTCTACATGTCACCATCGTC	(GTC)5	305	ac-bd	Most F1 seedlings were ‘abcd’ genotypes.
KF0536	ACAGTCAAGCCTCCATTCCTTT	GTTAGGGGGTTTGGAATCGCTA	(ATC)11	249	ab-ac	
KF0542	TCCCAAATGCTTTCTCTCTCCT	TCATTGGCTTGGATGGTCTCAA	(AACAAT)2	430	abc-ab	
KF0545	CTCTGTTCTTCGATCCACACCA	CGAAGAACCAAGAAGCTCGTTC	(TCTTCG)5	203	ac-ab	
KF0551	GTGAGGGAAGTTGACGTTAGGT	GCGGCATGTCAAAGTATTTCGT	(GAA)8	361	ac-bd	PCR amplicons were weak and most F1 seedlings were ‘abcd’ genotypes.
KF0587	AGGACATTGCTTCCTACTGGTG	ATATGGGTGTCAAGCAAGCAGA	(TCGTCA)3	189	bb-aa	
KF0593	AAATTCATCGTCGTTTTGCGGT	CTGAGAGAGGTGGAGAGACAGA	(TCT)5	222	ac-bd	most F1 seedlings were ‘abcd’ genotypes.
KF0595	TTGTTCCTCTTCAACCCCAACA	AACTGAACTGGAGATTCGGCAT	(AATCCG)3	179	ab-bb	
KF0606	GCAAAAATATCACAAGAGAACTGA	TGATGAGTACTGCCGTTGTTCA	(CTCATC)6	164	aa-ab	PCR amplification failure.
KF0607	TGAACAACGGCAGTACTCATCA	CATAAGGCGGCCTTTTGATTGG	(TCC)7	166	aa-ab	
KF0610	CTCCGATAGCTTCCTTTCCCTC	TCAGATTTGGGTGTGAATCGGT	(CTC)7	255	ab-abc	
KF0635	CGCACCTCAAACAGCATTCATT	GGTTGAGCAGGTTTAGGATGGA	(GCCGCCACT)2	305	ab-cd	most F1 seedlings were ‘abcd’

						genotypes.
KF0638	CCCTCCCTTCTCTTCTTTTCCA	CATCGATTGCATTGCAGCAGAA	(GGCTGT)3	129	cd-ab	PCR amplification failure.
KF0643	CCATTGTTGTCAGCAGCATCAA	TTCAAGCGGCAAAGGCAATTAA	(AACCCT)5	338	ab-cc	
KF0657	ACCTTCACATGCTGGGAATCAT	GGGGAGATTGTGTGAACCTCAT	(CAG)6	198	bc-ab	PCR amplification failure.
KF0660	TTGGCGTAATTTTTGTCGCAGT	TCGTCTTCCTCTTAAGTCGCAC	(CACAAC)2	266	bb-aa	
KF0664	GAAGTTACCTGAGTCGTCGAGG	GGTTCAACCCTCTTAGCTTCCA	(TGACTGGGT)2	300	aa-ab	PCR amplicons were weak and unclear.
KF0665	AACGAAACCAAACAGTGTTGGG	CTCACCGCCTAATTCCTCCTAC	(GAA)6	234	aa-bb	
KF0684	GTGTGATCCGGAGACAGAAACT	TCCGAAACACACTCTTCTCACG	(AAC)7	174	aa-ab	PCR amplification failure.
KF0685	ATCCTGCTGTGTGTGTCAGATT	GTGCTTCGTCTTGTTGCTTTCT	(AGGCGG)3	224	ab-ab	
KF0696	TGGTACACCAATCCAGGAACAG	AACATGTTACACTGCTGGGCTA	(CTG)5	267	ab-aa	
KF0734	CTGTACCAGTGCCTAATCAGGG	ACTGAACATTTGAGGTGCATGC	(CCT)7	194	aa-ab	PCR amplicons were weak and unclear.
KF0768	AGAGTGACTTGGCTAATGGTGG	GACGTAGGCAGCGAATTTCTTC	(CAA)9	358	abcd-ce	
KF0783	ATATACCACTAGCCGAACTCGC	AGATGTCGTGGAGTTGCAGAAT	(ACCACT)2	150	abcd-ab	
KF0784	GAGAGGAGCACAACGAAGGTAA	CCAAGACCCCTTTCCAATAGCT	(ATTGTC)3	132	aa-ab	PCR amplicons were weak and unclear.
KF0799	TCAAGTTTTGCCTGCGAATGAG	ATAATGGGGAGAAGGGGGAGAA	(TTCCCA)5	442	ab-bb	
KF0807	CCCTAGCAGCTGTCTCAAAGAA	GAAGATTTTGCGGTGAGAGACG	(ACTCCG)5	254	abc-bcd	
KF0810	GCAACACGCTTAGGGTTTGATT	AAGTTGAGTAGAGGGCTTCTGC	(AGAGA)4	257	aa-ab	

Notes: a, SSR markers with underline were used into mapping analysis; b, the anticipated band size *in silico*; c, 'f' means 'Fushun', 'k' means 'Kemsull'; d, illuminate why some SSR markers with admissible genotypes by JoinMap 4.0 could not apply into linkage analysis.

Table 13. Genic SSR markers developed by RNA-seq and BLASTX hit results in The Arabidosis Information Resource (TAIR) database.

ID[a]	TARI[b] number	TARI symbol	TAIR description
KF0007	PF03171	ACO4, EAT1,EFE	ethylene-forming enzyme
KF0036	PF04570		Protein of unknown function (DUF581)
KF0043	PF00227	PBG1	20S proteasome beta subunit G1
KF0097	PF08263, PF00560	LRX2	leucine-rich repeat/extensin 2
KF0112	PF00201	UGT73B2	UDP-glucosyltransferase 73B2
KF0120	PF00125		Histone superfamily protein
KF0141	PF00831		Ribosomal L29 family protein
KF0147	PF05047		NADH-ubiquinone oxidoreductase B8 subunit, putative
KF0234	PF00406		P-loop containing nucleoside triphosphate hydrolases superfamily protein
KF0271	PF01496	VHA-A1	vacuolar proton ATPase A1
KF0288	PF04818	PCFS4	PCF11P-similar protein 4
KF0305		NLP7	NIN like protein 7
KF0382	PF01344, PF00646		Galactose oxidase/kelch repeat superfamily protein
KF0416	PF00368	HMG, HMGR1	hydroxy methylglutaryl CoA reductase 1
KF0434	PF00097		RING/U-box superfamily protein
KF0453	PF00170	ABF2, AREB1, ATAREB1	abscisic acid responsive elements-binding factor 2
KF0468			Chaperone DnaJ-domain superfamily protein
KF0512	PF00023		XB3 ortholog 5 in Arabidopsis thaliana
KF0545			F-box family protein
KF0607	PF02893		GRAM domain family protein
KF0643	PF01535		ATP binding;nucleic acid binding;helicases

KF0685	PF00400, PF12490	ATATG18H, ATG18H	homolog of yeast autophagy 18 (ATG18) H
KF0696	PF00160	ATPUB49,PUB49	plant U-box 49
KF0799	PF00443	UBP25	ubiquitin-specific protease 25
KF0810	PF00514		ARM repeat superfamily protein

Notes: a, same as table 12.

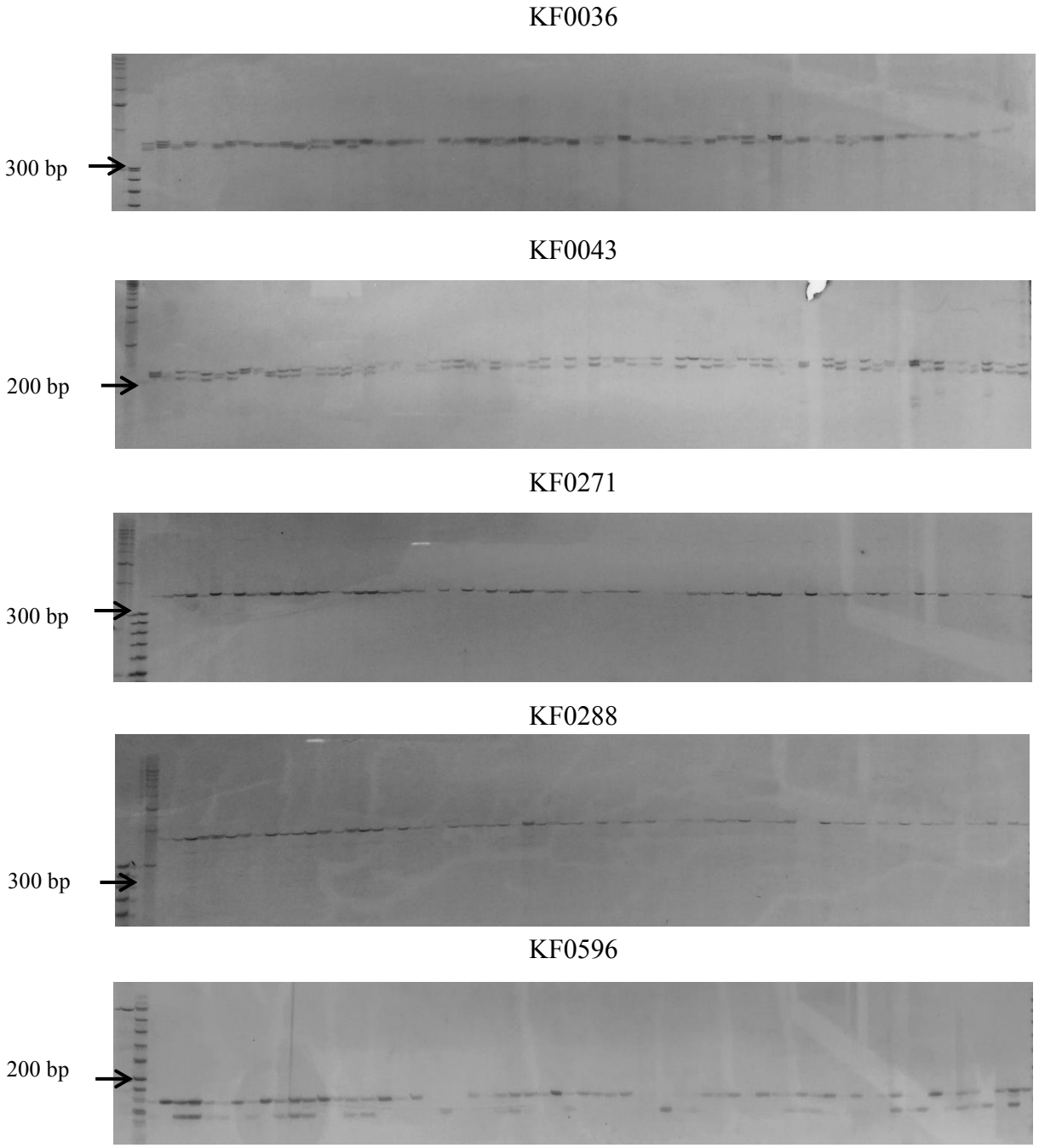


Fig. 14. 8% denatured poly-acrylamide gel profiles of PCR amplification of five SSR primer pairs for the whole mapping population.

Linkage Analysis and Genetic Map Construction

Of 29 SSR markers developed by RNA-seq, 18 were heterozygous in female parent 'Fushun' and 20 in male parent 'Kemsull'. These 18 SSR markers were incorporated with 107 RAPDs, 10 public SSRs, and 308 AFLPs into mapping analysis for 'Fushun'; 20 SSR markers were combined with 109 RAPDs, 3 public SSRs, and 379 AFLPs to perform the construction of 'Kemsull'. Totally, 678 markers including 143 RAPDs, 11 public SSRs, 495 AFLPs, and 29 newly mined SSR markers were conjointly used to construct a combined linkage map (Fig.15).

Compared with the previous map including 76 RAPDs, 5 public SSRs, and 214 AFLPs and covering a total length of 1,376.9 cM (Fig. 12), a new combined map obtained here was improved a little with an addition of 11 informative SSR marker in saturation level (309 markers) and map length (1,441.6 cM). The length of 15 LGs varied from 27.0 cM (c15) to 146.3 cM (c1) and the number of loci was changed from 3 (c14 and c15) to 100 (c1). The c1 still owned the highest marker density of 1.5 cM between two adjacent markers. The marker density of this combined map was 4.7 cM and might be suitable QTLs analysis. Of these newly located SSR markers, 7 SSR loci including KF0120, KF0141, KF0288, KF0340, KF0416, KF0595, and KF0696 were incorporated into the first linkage group (LG), two SSR loci KF0545 and KF0097 were arranged at LG7, and SSR loci including KF0810 and KF0536 were successfully grouped at LG2 and LG6, respectively. These gene-based markers are more informative compared with dominant and other codominant markers in genetic study.

The majority of molecular markers mapped in c1, c2, c6, and c7 as in Fig.12 were still retained at same LGs in Fig.15. However, we observed there were some obvious changes between LGs of two maps. Two LGs arranged for k1 were aligned in c1 (Fig.12), but 3 out of 5 markers (E1/M5-300, E1/M5-218, and MSG0232) from the small LG were incorporated into another large LG as a single LG in Fig.15. A segregation distortion region (SDR) was identified from previous study including three loci of E4/M5-182, E4/M10-162, and OPA-7-1500 in c6 in Fig.12, in which E4/M5-182 and E4/M10-162 replaced by E4/M5-252 and E4/M6-210 together with OPA-7-1500 was recognized as a SDR in c6 in Fig.15. We supposed that new and stronger linkage relationships between markers resulted from new markers grouped in the LG maybe explain these changes.

After BLASTX to annotated proteins in The Arabidopsis Information Resource (TAIR) database, KF0097, KF0120, KF0141, KF0288, KF0416, KF0545, KF0696, and KF0810

were recognized to have BLASTX hits to leucine-rich repeat extensin 2, histone superfamily protein, ribosomal L29 family protein, PCF11P-similar protein 4, hydroxy methylglutaryl CoA reductase, F-box family protein, plant U-box 49, and ARM repeat superfamily protein, respectively. The information of other adopted polymorphic SSR markers with BLASTX hits in TAIR database were listed in the table 14. Those SSR loci are more informative than other loci which will greatly useful in comparative mapping or other genetic studies.

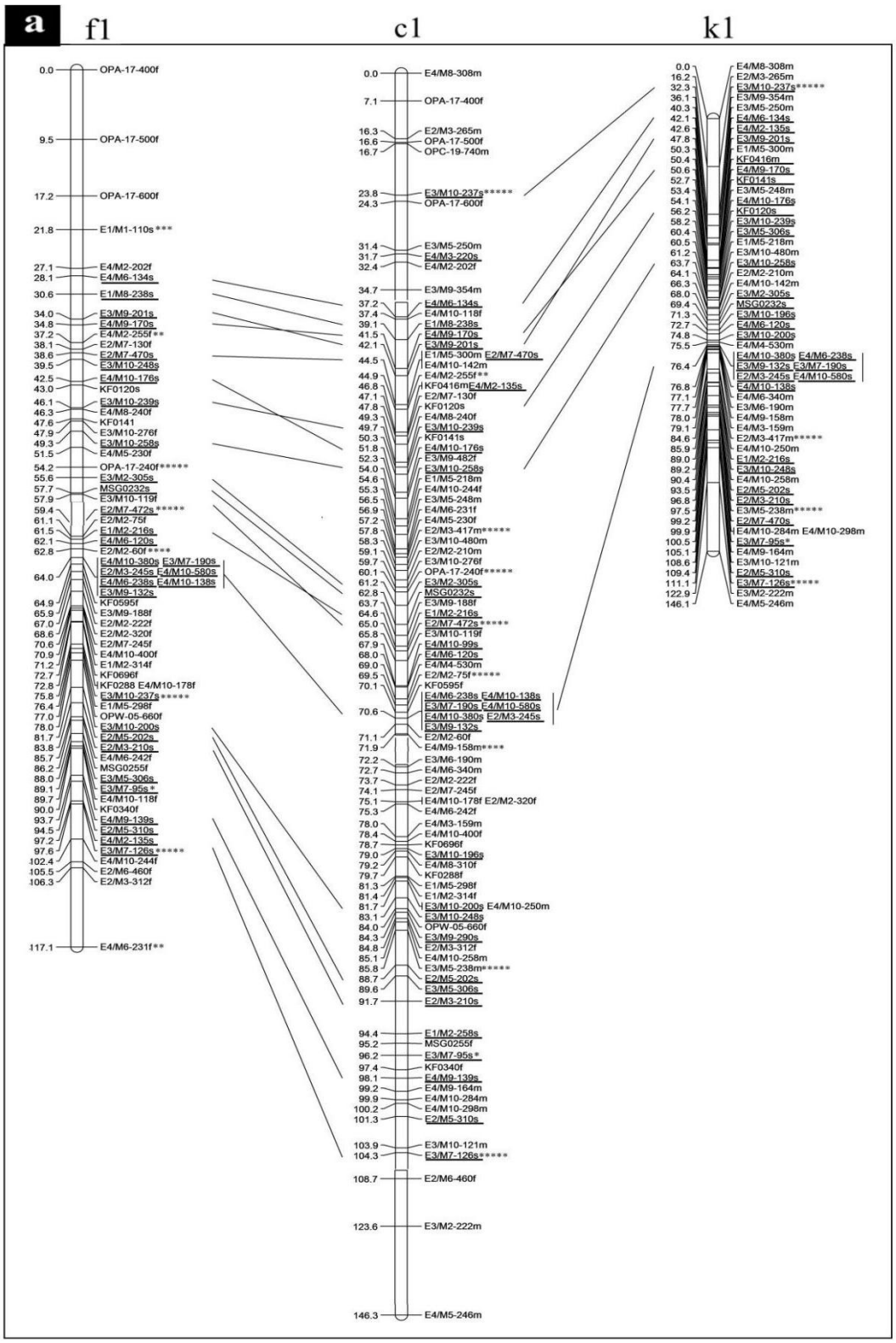


Fig. 15. Three-lined genetic linkage maps based on RAPD, AFLP, and SSR markers for tea plant. The arrangement of three maps and significance of all symbols and marks were same as Fig. 12.

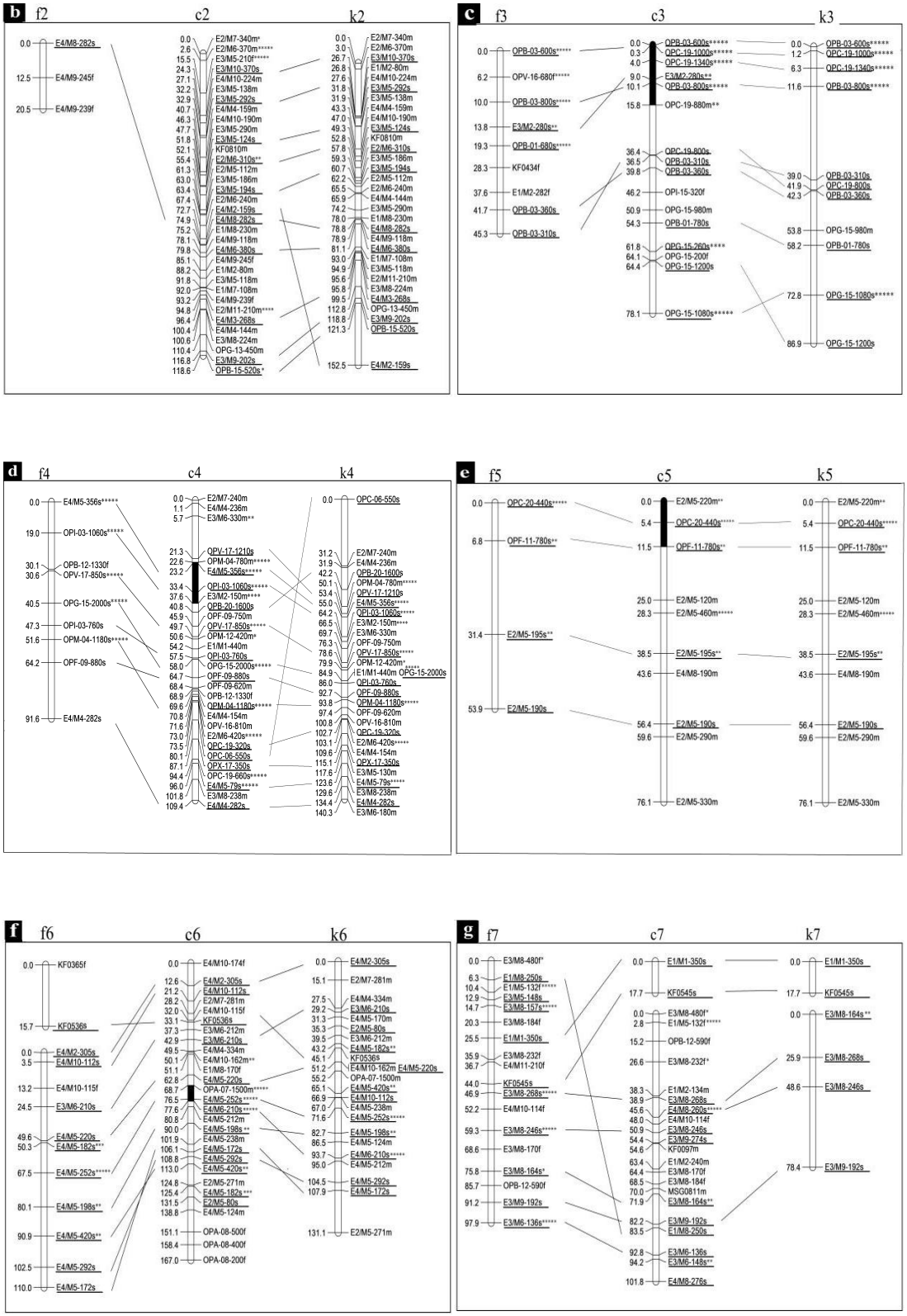


Fig. 15. (Continued)

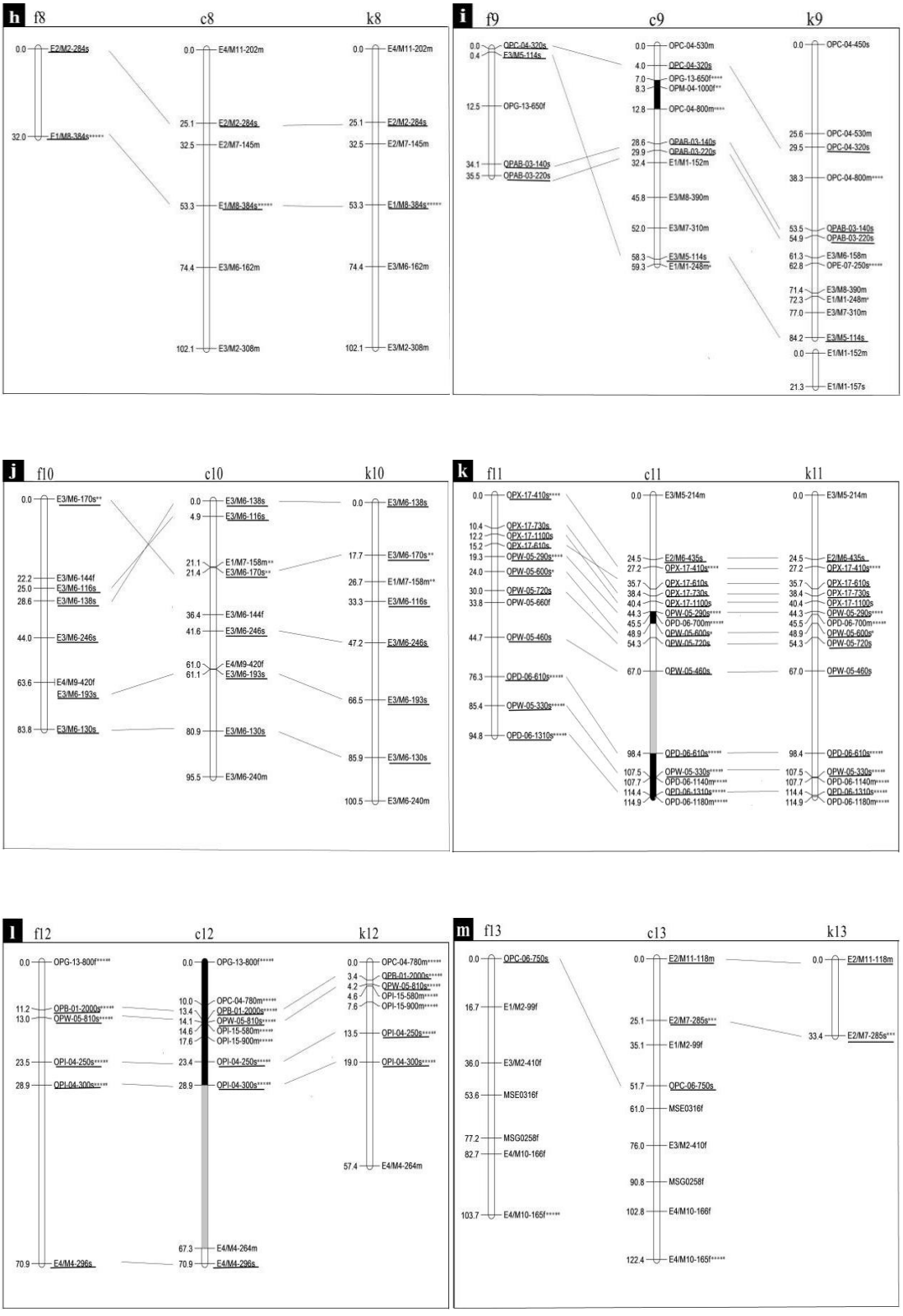


Fig. 15. (Continued)

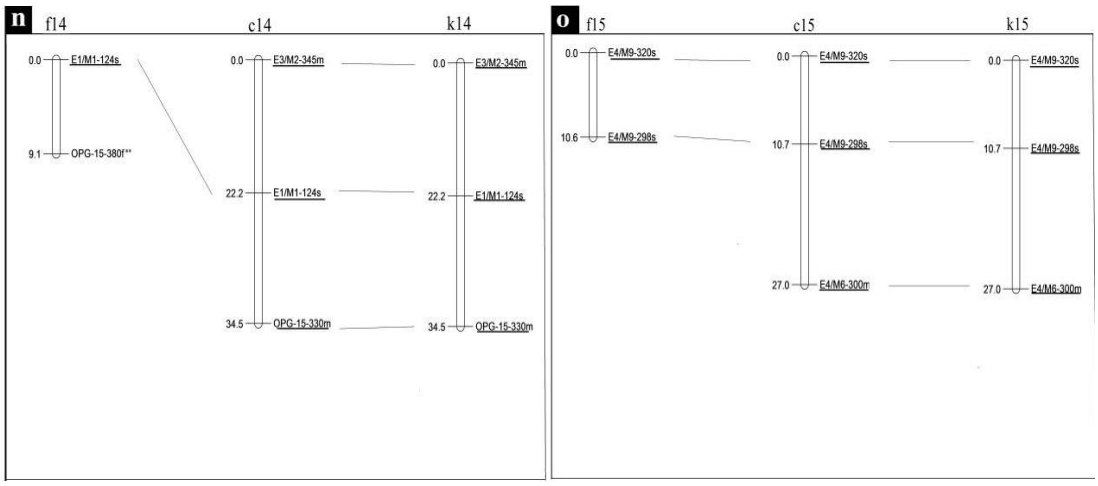


Fig. 15. (Continued)

Table 14. Summarized data of each linkage group of three linkage maps.

LG		Length (cM)	No. of loci	Average distance (cM)	No. of			
					AFLP	RAPD	SSR	DS ($p < 0.05$)
LG1	c1	146.3	100	1.5	85	6	9	10
	f1	117.1	67	1.7	54	5	8	9
	k1	146.1	57	2.6	53	0	4	5
LG2	c2	118.6	34	3.5	31	2	1	6
	f2	20.5	3	6.8	3	0	0	0
	k2	152.5	31	4.9	28	2	1	0
LG3	c3	78.1	16	4.9	1	15	0	8
	f3	45.3	9	4.1	2	6	1	4
	k3	86.9	11	9.7	0	11	0	5
LG4	c4	109.4	29	3.8	11	18	0	12
	f4	91.6	9	10.2	2	7	0	0
	k4	140.3	29	4.8	13	16	0	10
LG5	c5	76.1	10	7.6	8	2	0	5
	f5	53.9	4	13.5	2	2	0	3
	k5	76.1	10	7.6	8	2	0	5
LG6	c6	167	28	6.0	23	4	1	7
	f6	125.7	13	9.7	11	0	2	4
	k6	131.1	23	5.7	21	1	1	5
LG7	c7	119.5	23	5.2	19	1	3	6
	f7	97.9	18	5.4	16	1	1	7
	k7	96.1	6	16.0	5	0	1	1
LG8	c8	102.1	6	17.0	6	0	0	1
	f8	32.0	2	16.0	2	0	0	1
	k8	102.1	6	17.0	6	0	0	1
LG9	c9	59.3	12	4.9	5	7	0	3
	f9	35.5	5	7.1	1	4	0	0
	k9	105.5	14	7.5	7	7	0	3
LG10	c10	95.5	10	9.6	10	0	0	2
	f10	83.8	8	10.5	8	0	0	1
	k10	100.5	8	12.6	8	0	0	2

LG11	c11	114.9	16	7.2	2	14	0	9
	f11	94.8	12	7.9	0	12	0	6
	k11	114.9	16	7.2	2	14	0	9
LG12	c12	70.9	10	7.1	2	8	0	8
	f12	70.9	10	7.1	2	8	0	8
	k12	57.4	8	7.2	1	7	0	7
LG13	c13	122.4	9	13.6	6	1	2	2
	f13	103.7	7	14.8	4	1	2	1
	k13	33.4	2	16.7	2	0	0	1
LG14	c14	34.5	3	11.5	2	1	0	0
	f14	9.1	2	4.6	1	1	0	1
	k14	34.5	3	11.5	2	1	0	0
LG15	c15	27.0	3	9.0	3	0	0	0
	f15	10.6	2	5.3	2	0	0	0
	k15	27.0	3	9.0	3	0	0	0
Total/ average	c1	1,441.6	309	4.7	214	79	16	79
	f1	992.4	171	5.8	111	47	14	45
	k1	1,404.4	227	6.2	159	61	7	54

Literature Cited

Chen L, Zeno A, Chen ZM (2012) Global tea breeding: achievements, challenges, and perspectives. Hangzhou: Zhejiang University Press.

Chen L, Zhou ZX, Yang YJ (2007) Genetic improvement and breeding of tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. Euphytica 154:239-248.

Chen L (2016) Differential metabolic profiles and transcriptomic profiles during different color stages of albino tea plant (*Camellia sinensis*). Adv Crop Sci Technol 4(3): 1-41.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485.

Li H, Huang W, Liu ZW, Wang YX, Zhuang J (2016) Transcriptome-based analysis of Dof family transcription factors and their responses to abiotic stress in tea plant (*Camellia sinensis*). Intern' J Genomics 1-15.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. BMC Genomics 9: 287.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28(8):1086-1092.

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q et al. (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics 12(131): 1-19.

Stam P (1993) Construction of intergrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3: 739-744.

Tan LQ, Wang LY, Wei K, Zhang CC, Wu LY, Qi GN, Cheng H, Zhang Q, Cui QM et al. (2013) Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). Plos ONE 8(11): e81611.

Tavassolian I, Rabiei G, Gregory D, Mnejja M, Wirthensohn MG, Hunt PW, Gibson JP, Ford CM, Sedgley M et al. (2010) Construction of an almond linkage map in an Australian population Nonpareil × Lauranne. BMC Genomics 11:551.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Res 40(15):e115.

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.

Wang RJ, Gao XF, Kong XR, Yang J (2016) An efficient identification strategy of clonal tea cultivars using long-core motif SSR markers. Springer Plus 5: 1-16.

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in plant sciences. Am J Bot 99(2): 193-208.

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 8(5):821-829.

Zhang CC, Wang LY, Wei K, Wu LY, Li HL, Zhang F, Cheng H, Ni DJ (2016) Transcriptome analysis reveals self-incompatibility in the tea plant (*Camellia sinensis*) might be under gametophytic control. BMC Genomics 17:359.

Zhang H, Wei L, Miao HM, Zhang T, Wang CY (2012) Development and validation of genic-SSR markers in sesame by RNA-seq. BMC Genomics 13:316.

ABSTRACT IN KOREAN

소엽종(*Camellia sinensis* var. *sinensis*)의 일본 차 품종 '후슌'(종자친)과 한국 차 품종 '금설'(화분친)을 교배한 후 얻어진 79 개의 F1 교배실생을 이용하여 double pseudo-testcross 이론에 따라 두 품종에 대한 RAPD, AFLP 및 SSR 마커에 기반한 기초 유전적 연관지도를 작성하였다. 처음에 양친과 6 개의 F1 실생에 대하여 총 660 개의 10-염기 임의 RAPD 프라이머와 60 개의 공개 SSR 프라이머를 탐색하여, F1 실생 유전자형의 분석을 위한 41 개의 10-염기 임의 프라이머와 11 개의 공개 SSR 프라이머를 선발하였다. 이들 분자표지를 기반으로 차 재배 품종 인 '후슌'과 '금설'에 대한 기초 유전적 연관지도가 작성되었다. '후슌' 연관지도는 8 개의 연관 그룹에 고르게 분포하는 37 개의 유전자좌를 포함하는 총길이 255.7cM 으로 인접한 두 분자표지 사이의 평균 거리는 6.9 cM 이다. 금설의 유전자지도는 5 개의 연관 그룹에 30 개의 유전자좌가 포함된 총길이 196.1cM 으로 두 분자표지 사이의 평균거리는 6.5cM 이다. 두 연관지도의 연관 그룹간 비교에서 양친에 공유되는 분자표지에 의해 f1 과 k2, f3 과 k3, f4 와 k1, f5 와 k4 및 f8 과 k5 간 명확한 동등성이 확인되었다.

동일한 유전적 지도작성 집단을 사용하여 double pseudo-testcross 이론에 따른 AFLP 분자표지를 기반으로 하여 '후슌'과 '금설'에 대한 유전적 연관지도가 별도로 분석되었다. AFLP 기반 연구에서는 27 개의 프라이머 조합으로 2,439 개의 표지가 나타났고, 각각 평균 90.3 개의 표지가 얻어졌다. 이들 중 다형성을 나타

내는 495 개 AFLP 표지로부터 400 개(80.8 %)의 표지가 멘델의 분리비($p < 0.01$)에 적합한 것으로 확인되었는데, 그 중 136 개(34.0 %)는 3:1 의 분리비율을, 그리고 264 개(66.0 %)는 1:1 의 분리비율을 나타내었다. '후슌'의 연관지도는 8 개의 연관그룹에 분포하는 92 개의 유전자좌로 구성되었으며, 총 길이는 470.8cM 이고 인접한 두 분자표지 사이의 평균 거리는 5.2 cM 이었다. '금설'의 연관지도는 17 개의 연관그룹으로 구성되어 152 개의 유전자좌가 위치하고 있으며 총길이는 1,053.1 cM 이고 인접한 두 분자표지 사이의 평균 거리는 6.9 cM 이었다. 두 연관지도 간의 동등성이 이형의 공유 분자표지에 의해 f1 과 k1, f2 와 k3, f3 과 k6, f4 와 k5, f6 과 k2, f7 과 k14 간 각각에 대해 분명하게 확인되었다. 이들 495 개의 AFLP 표지를 143 개의 RAPD 표지와 11 개의 SSR 유전자좌를 단일 data 세트로 통합하여, JoinMap 4.0 에 의해 'One-step method'을 사용하여 차나무에 대한 단일 유전적 연관지도를 작성하였다. 이 연관지도는 총 길이 1,376.9 cM 이며, 인접 분자표지 간 평균 거리는 4.7 cM 으로 15 개의 연관그룹에 대해 균등하게 분포된 295 개의 유전자좌(76 RAPD, 5 SSR 및 214 AFLP 마커)를 포함하고 있으며, 두개의 분자표지가 결합되어 segregation distortion 을 나타낼 수 있는 8 개의 segregation distorted 영역이 존재함을 확인하였다.

또한 high throughput RNA 서열 분석기술을 이용하여 보다 유용성이 높은 분자표지를 선발하여 이전의 포화도가 상당한 연관지도에의 통합이 수행되었다. '후슌'과 '금설'의 꽃과 잎의 전사체 염기 서열을 분석하고 비교한 결과, 1,800 개

의 다형성 SSR 표지가 성공적으로 탐색되었고, 그 중 296 개가 임의 선택되어 차나무 표본집단(두 부모와 6 개의 F1 실생을 포함)에 대해 다형성을 검증한 후 양친에서 재현성 증폭을 나타내는 75 개(25.3%)를 선별하였다. '후슌' 또는 '금설'의 이형 접합 형태를 나타내는 29 개 SSR 표지(38.7%)가 최종 선발되었고, 이전의 유전적 연관지도에의 통합이 수행되었는데, 이 중 11 개의 표지가 성공적으로 유전적 연관지도에 통합되었다. 새로운 통합 유전적 연관지도는 79 개의 RAPDs, 5 개의 공개 SSRs, 214 개의 AFLPs, 새로 선발된 11 개의 SSRs 을 포함하고 있으며, 총 길이는 1,441.6 cM 이고, 두 인접 표지 사이의 평균 거리는 4.7 cM 이었다. 본 연구는 향후 차나무에 있어서 중요한 농업적 특성에 대한 질적 또는 양적 형질의 유전자좌 (QTLs) 분석을 위한 토대를 제공할 것이다.

ACKNOWLEDGEMENTS

Thanks for all the friends who came into my life in the past time, I sincerely wish you good health and everything goes well; simultaneously, I also expect those wonderful persons who will participate into my future and we can together work on some interesting and meaningful things. I wish the world is always peace and human could quickly evolve to be more wise and stronger as most powerful creature on our planet, then better utilize resource, protect our earth, and explore the truth.

Yali Chang