



# QTL Mapping for Leaf Area of Tea Plants (*Camellia sinensis*) Based on a High-Quality Genetic Map Constructed by Whole Genome Resequencing

Yanlin An<sup>1†</sup>, Linbo Chen<sup>2†</sup>, Lingling Tao<sup>1†</sup>, Shengrui Liu<sup>1</sup> and Chaoling Wei<sup>1\*</sup>

<sup>1</sup> State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China, <sup>2</sup> Yunnan Provincial Key Laboratory of Tea Science, Tea Research Institute, Yunnan Academy of Agricultural Sciences, Menghai, China

## OPEN ACCESS

### Edited by:

Hirokazu Tsukaya,  
The University of Tokyo, Japan

### Reviewed by:

Eiji Nitasaka,  
Kyushu University, Japan  
Sean Mayes,  
University of Nottingham,  
United Kingdom

### \*Correspondence:

Chaoling Wei  
weichl@ahau.edu.cn

† These authors have contributed  
equally to this work

### Specialty section:

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

Received: 11 May 2021

Accepted: 07 July 2021

Published: 29 July 2021

### Citation:

An Y, Chen L, Tao L, Liu S and  
Wei C (2021) QTL Mapping for Leaf  
Area of Tea Plants (*Camellia sinensis*)  
Based on a High-Quality Genetic Map  
Constructed by Whole Genome  
Resequencing.  
*Front. Plant Sci.* 12:705285.  
doi: 10.3389/fpls.2021.705285

High-quality genetic maps play important roles in QTL mapping and molecular marker-assisted breeding. Tea leaves are not only important vegetative organs but are also the organ for harvest with important economic value. However, the key genes and genetic mechanism of regulating leaf area have not been clarified. In this study, we performed whole-genome resequencing on “Jinxuan,” “Yuncha 1” and their 96 F1 hybrid offspring. From the 1.84 Tb of original sequencing data, abundant genetic variation loci were identified, including 28,144,625 SNPs and 2,780,380 indels. By integrating the markers of a previously reported genetic map, a high-density genetic map consisting of 15 linkage groups including 8,956 high-quality SNPs was constructed. The total length of the genetic map is 1,490.81 cM, which shows good collinearity with the genome. A total of 25 representative markers (potential QTLs) related to leaf area were identified, and there were genes differentially expressed in large and small leaf samples near these markers. GWAS analysis further verified the reliability of QTL mapping. Thirty-one pairs of newly developed indel markers located near these potential QTLs showed high polymorphism and had good discrimination between large and small leaf tea plant samples. Our research will provide necessary support and new insights for tea plant genetic breeding, quantitative trait mapping and yield improvement.

**Keywords:** *Camellia sinensis*, leaf area, whole-genome resequencing, genetic map, QTL mapping

## INTRODUCTION

The tea plant [*Camellia sinensis* (L.) O. Kuntze] is a perennial, evergreen woody plant originating in southwest China, with a large (approximately 2.84 Gbp) and highly heterozygous genome (Xia et al., 2020a). Many biologically active substances beneficial to the human body, such as catechin, theanine, and caffeine, are abundant in tea, which are important components of tea quality and

**Abbreviations:** QTL, Quantitative trait locus; SNP, Single nucleotide polymorphism; qRT-PCR, Quantitative real-time PCR; indel, Insertion-deletion; LOD, Logarithm of odds.

flavor (Zhao et al., 2018). However, due to the long breeding period and high degree of self-incompatibility, it is difficult to obtain pure line germplasm, explore the formation mechanism of tea quality characteristics, and study the adaptive evolution of different tea populations. In addition, germplasm innovation still faces many difficulties (Meegahakumbura et al., 2016). The construction of a high-density genetic map is of great value for the quantitative trait loci of tea plants, molecular marker-assisted breeding, map-based cloning, and assisted genome assembly.

In the past few decades, based on molecular marker technologies such as simple repeat sequence (SSR) (Hu et al., 2013; Tan et al., 2013; Ma et al., 2014), and amplified fragment length polymorphism (AFLP) (Huang et al., 2005), some tea plant genetic maps have been constructed. However, due to the limitation of the total number of markers and lack of recombinant inbred lines, these genetic maps still have some shortcomings such as poor genome coverage and large gaps, which makes their application in practical research difficult. Fortunately, advances in sequencing technology and reduced costs have made it possible to construct high-density genetic maps. Ma et al. (2015) used specific-locus amplified fragment sequencing technology to obtain 6,448 molecular markers and successfully anchored these markers in 15 linkage groups. Xu et al. (2018) used the F1 population of “Longjing 43” and “Baihaozao” to construct a high-density genetic map with a total length of 1,678.52 cM using RAD sequencing and located 27 QTLs related to catechins or caffeine. However, reduced-representation genome sequencing still has some shortcomings, such as insufficient sequencing depth, uneven distribution of markers, and some pivotal SNP loci related to important traits being undetected. At the same time, different genetic populations will also show different characteristics of segregation. In previous studies, most of the small leaf tea trees were used as parents for hybridization. Due to the insufficient number of harvested F1 generation individuals and the small differences in parental genetic background, the genetic map constructed using these populations is not conducive to accurate QTL mapping for some special traits, such as leaf area, plant type, environmental adaptability, and so on.

The leaves are important nutrient organs and the main harvests of the tea plant, and the delicious tea beverage is processed from the fresh leaves of the tea plant (Tan et al., 2016). The size of tea leaves has an important influence on the yield, especially for black tea (including Pu'er tea), oolong tea, dark tea, and matcha powder. As the two most economically valuable tea types widely planted worldwide, *Camellia sinensis* var. *sinensis* (CSS) and *C. sinensis* var. *assamica* (CSA) have significant differences in leaf area, plant type, cold resistance, secondary metabolic characteristics, and geographical distribution (An et al., 2020). CSA is usually considered to be arborous or semi-arborous with a large leaf area, whereas CSS tea plants are mainly shrubs with small leaf areas. Exploring the key regulatory genes of tea leaf development is of great significance for studying the adaptive evolution of tea trees, breeding new varieties, and increasing the income of tea farmers. In previous studies, Tan et al. (2016) conducted preliminary QTL mapping for the size of mature tea plant leaves, but due to the limited accuracy of the

genetic map, it is still necessary to reconstruct a more refined genetic map for QTL mapping based on resequencing.

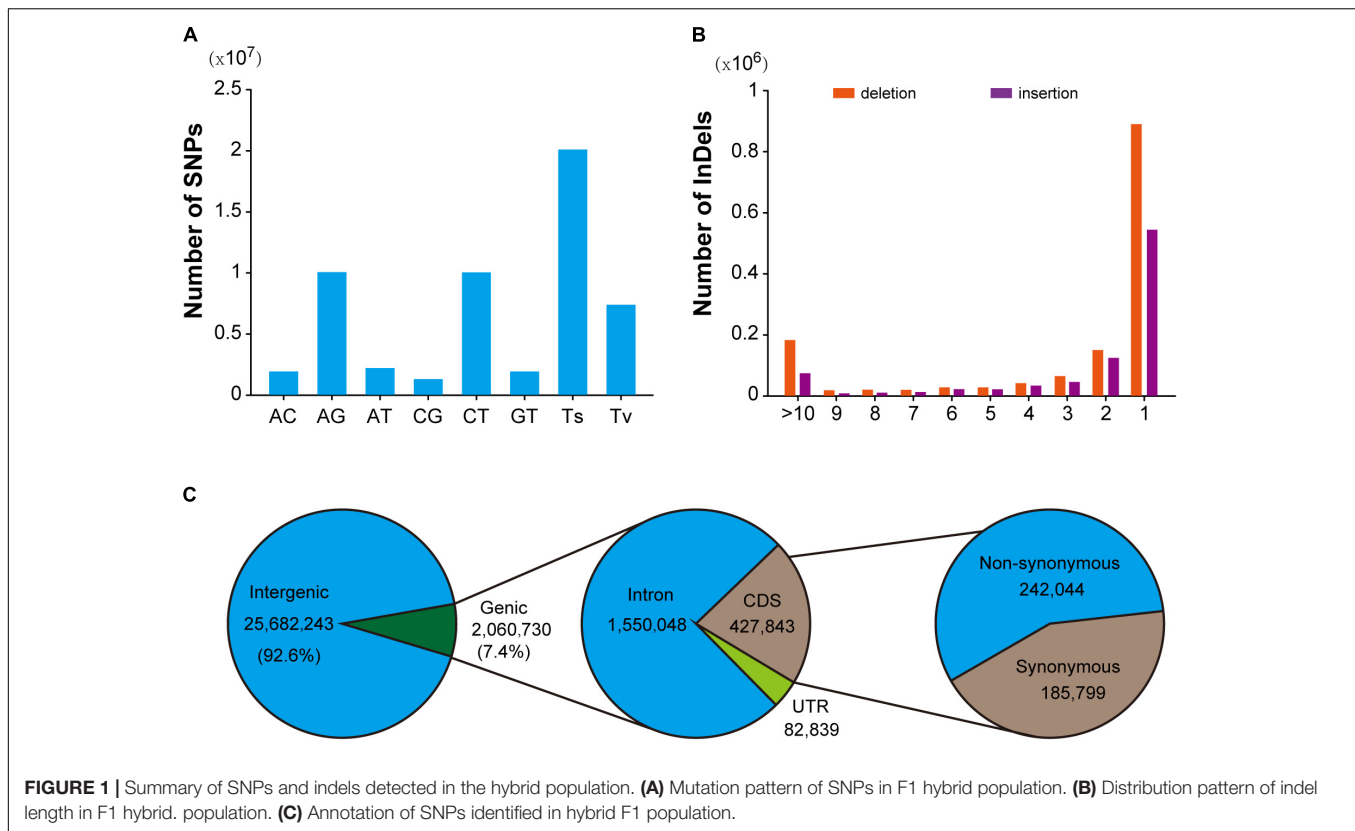
In this study, the CSS tea plant variety “Jinxuan” was used as the female parent and crossed with the CSA tea plant cultivar “Yuncha 1,” and a total of 96 reliable F1 individuals were harvested. Using whole-genome resequencing technology, two parents and 96 offspring were re-sequenced. Sufficient sequencing data was generated for accurate genotyping. A large number of SNP and indel variations were obtained. Finally, a high-quality genetic map consisting of 15 linkage groups and 8,956 markers was obtained, with a total map length of 1,490.81 cM. Using the F1 generation and the newly constructed genetic map, 25 representative markers (potential QTLs) and two SNPs related to leaf area were successfully located in second linkage group. In addition, we also developed 31 pairs of highly polymorphic indel markers. To summarize, our research not only provides a large amount of basic data and polymorphic loci, but also constructs a high-quality genetic map and QTL mapping for leaf area trait. These data and map will provide strong support for further research on tea plant trait separation, functional gene location and marker-assisted breeding.

## RESULTS

### Identification of SNP and Indel Loci by Resequencing

Whole-genome resequencing produced 1.84 Tbp of raw data. Except for the sequencing depth of the parents, which reached 17×, the average sequencing depth of all the offspring was 5×. Mapping clean reads to the reference genome of “Shuchazao” revealed that the properly\_mapped reached 84% (**Supplementary Table 1**). Sufficient sequencing data and a high alignment rate provide the basis for accurate genotyping.

In this population, a total of 28,144,625 SNPs were identified, with an average density of 9.6/Kb. After excluding 648,887 multi-allelic SNPs, it was found that AG and AT mutations were the most abundant types of transformation and transversion mutations, accounting for 35.8% (10,063,296) and 7.9% (2,211,580), respectively. Transformation is the most common mutation type (71.4%), and the ratio of transformation/transversion (ts/tv) is about 2.51 (**Figure 1A**). In addition, the mutation sites at different positions could have various effects on the function of genes. Detailed location and functional annotations provide more comprehensive SNP variation information. As shown in **Figure 1C**, of the 27,742,973 annotated SNPs, 92.6% (25,682,243) were located in the intergenic region, whereas only 7.4% (2,060,730) were located in the genes. Of the SNPs located in the genes, 1,550,048, 427,843, and 82,839 were located in the intron, coding region, and untranslated region, respectively. The annotation results showed that 21,373 and 217,884 loci were predicted to have high and medium effects, respectively, and the abnormal gene functions may be caused by these loci. Furthermore, 242,044 SNPs located in the coding region were annotated as non-synonymous mutations. Moreover, 2,780,380 indel mutations were identified using resequencing, including 432,998 multiple



allelic mutations. Indels with lengths equal to 1 and greater than 10 accounted for 61.6% and 11% of the total biallelic mutation sites, respectively. Detailed indel length information is shown in **Figure 1B**. It should be noted that due to the existence of multiple allelic indel loci, it is impossible to annotate all indel mutation sites accurately. However, similar to SNP, most indel variations are also located in intergenic regions.

## SNP Markers Genetic Typing and Genetic Map Construction

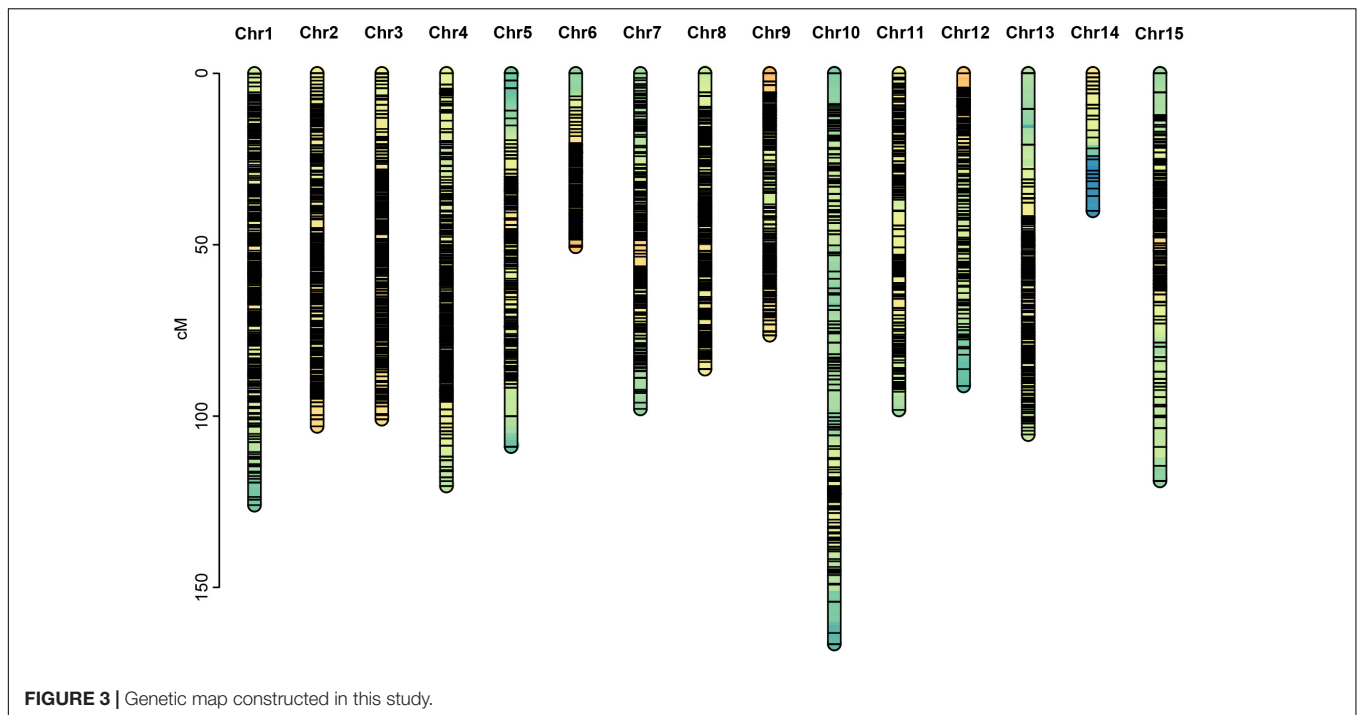
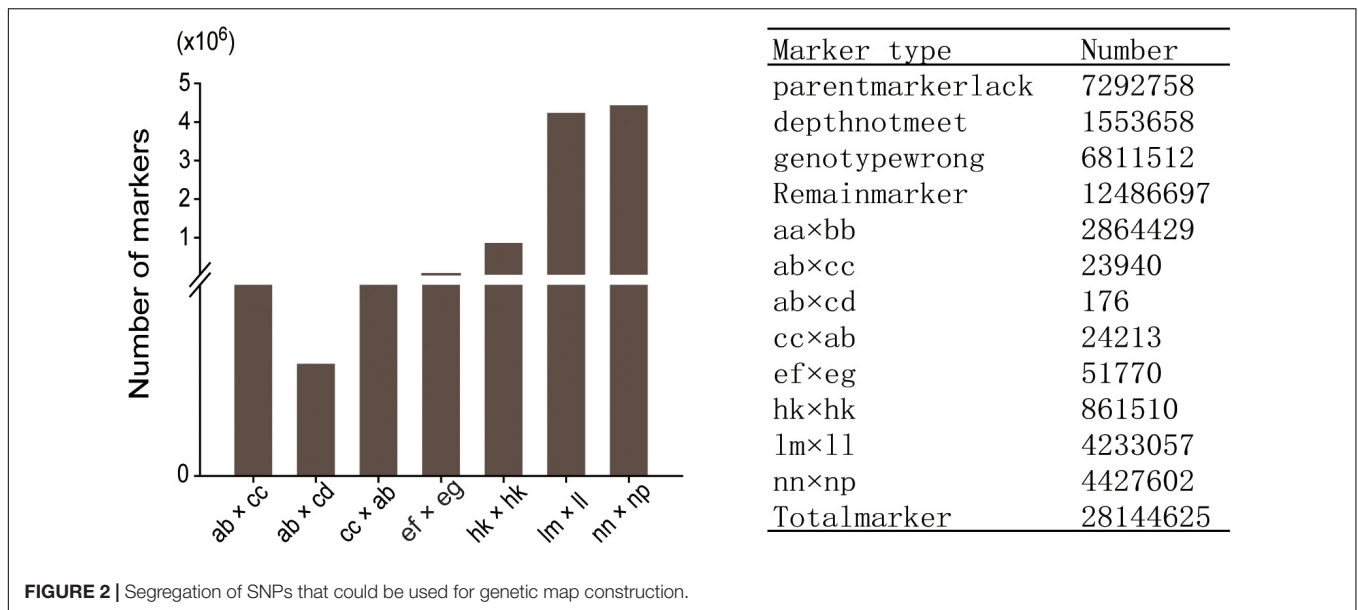
Among the 28,144,625 markers obtained, after filtering the SNPs of missing parent genotypes (7,292,758), insufficient depths (1,553,658), and genotyping errors (6,811,512), 12,486,697 highly integrated SNPs were obtained. Except for 2,864,429 aa × bb type markers that could not be used to construct a genetic map of the pseudo-testcross population, The remaining 9,622,268 SNPs (including ab × cc, ab × cd, cc × ab, ef × eg, hk × hk, lm × ll, nn × np) were further filtered. As shown in **Figure 2**, among the 9,622,268 markers, the nn × np type SNP was the most abundant, accounting for 46% (4,427,602), followed by the lm × ll type, accounting for 44% (4,233,057), and the ab × cd type, which was the least abundant, with only 176 SNPs. After strictly filtering the above-mentioned markers and integrating them with a previously reported genetic map, we constructed a high-quality genetic map containing 8,956 markers, with a total length of 1,490.81 cM (**Figure 3** and **Supplementary Figure 1**).

Among the 15 linkage groups, the longest one was tenth linkage group, with a length of 166.54 cM, whereas the shortest

one was fourteenth linkage group, with a length of only 40.12 cM. Each linkage group contained an average of 597 genetic markers, among which the minimum average map distance and the maximum average map distance appeared in linkage groups 6 and 10, which were 0.1 cM and 0.35 cM, respectively. The maximum gap present in thirteenth linkage group had a length of 10.38 cM, whereas the maximum gap in eight linkage groups was less than 5 cM, indicating that the newly constructed genetic map was high quality. The completeness of the marker data also reflects the quality of the map. In this study, the completeness of all the marker data used to construct the map was over 99%, and there were none missing in the parent genotype. As shown in **Table 1** and **Figure 4**, the closer the spearman coefficient is to 1, the higher the accuracy of the linkage group. Among the 15 linkage groups, the ninth linkage group has the best collinearity with the genome, and the spearman coefficient is as high as 0.993, while the thirteenth linkage group has the lowest collinearity with the spearman coefficient of 0.808. The above results comprehensively indicated that the newly constructed genetic map was high quality and can be used for tea plant genome assembly and QTL mapping research.

## QTL Mapping of Leaf Area and Genome-Wide Association Analysis

Leaves are important organs in plants for photosynthesis, respiration, and transpiration. They affect plant growth and development and resistance to adversity in many ways (Ayala et al., 2015). Tea leaves are not only important nutrient



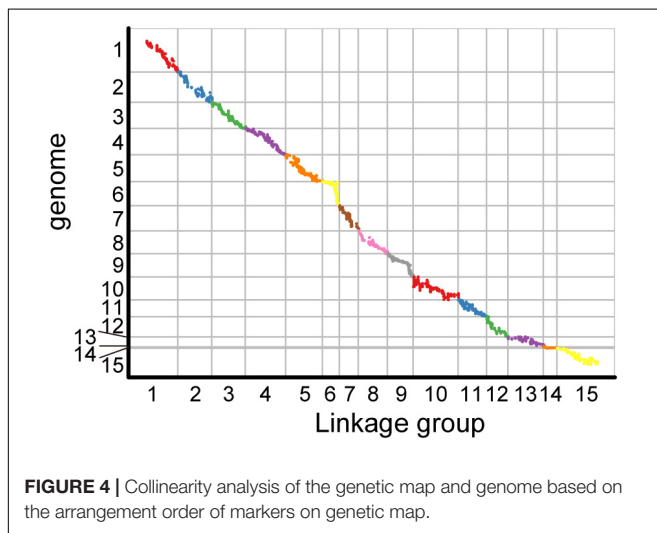
organs, but they are also closely related to the yield and quality of tea. According to the leaf area phenotype data of 96 F1 generations (**Supplementary Table 2**), we used the newly constructed high-density genetic map for QTL mapping. The results are shown in **Table 2**.

Twenty-five representative markers related to leaf area were identified from the second linkage group between 28 and 101 cM. Due to the limitation of QTL mapping accuracy, these markers may be regarded as potential independent QTLs. The LOD threshold ranged from 5.57 to 8.11, and the variation contribution rate ranged from 23.9% (potential QTL:

Marker1749183) to 32.8% (potential QTL: Marker1533238). All the 25 representative markers were regarded as major loci (variation contribution rate >10%), and the average phenotypic contribution rate was 30.32%. The genome-wide association study (GWAS) is one of the most important methods to study functional genes. In this study, 1,000,682 non-linked disequilibrium high-quality SNPs were used for GWAS analysis, and the results showed that the best model was GLM (**Supplementary Figure 2**). Interestingly, similar results were observed in two independent analyses, as shown in **Supplementary Figure 3**, the two leaf area related SNP

**TABLE 1** | Information on the newly constructed genetic map.

Linkage group	Total marker	Total distance (cM)	Average distance	Max gap	Spearman
1	662	126	0.19	4.19	0.992
2	848	103.05	0.12	2.46	0.941
3	929	100.89	0.11	3.22	0.992
4	759	120.45	0.16	3.23	0.989
5	574	108.93	0.19	8.91	0.972
6	506	50.57	0.1	6.68	0.921
7	546	97.93	0.18	3.42	0.944
8	598	86.29	0.14	5.5	0.983
9	612	76.38	0.12	3.31	0.993
10	481	166.54	0.35	9.12	0.91
11	593	98.22	0.17	5.33	0.969
12	482	91.18	0.19	4.87	0.969
13	588	105.39	0.18	10.38	0.808
14	155	40.12	0.26	4.35	0.927
15	623	118.88	0.19	6.68	0.93
Total	8956	1490.81	–	–	–

**FIGURE 4** | Collinearity analysis of the genetic map and genome based on the arrangement order of markers on genetic map.

loci obtained from the GWAS were also located in second linkage group, and the significance test  $q$  values were  $2.9E-8$  (L2\_134740620) and  $5.3E-8$  (L2\_210876632), respectively. There were 80 genes within the 90 kb interval on both sides of the 25 markers (potential QTL), and most of them have been predicted to be related to cell growth and development and plant organ pathology, including Golgi vessel transport (CSS0032416.2), OVATE family protein 14 (CSS0018166.1), LOC (CSS0013181.1), TSO1-like CXC domain (CSS0001030.1), Phosphatidylinositol-4-phosphate 5-kinase (CSS0011322.1), Sec34-like family protein (CSS0016412.1), ARF guanyl-nucleotide exchange factor (CSS0027335.1), and others. Detailed functional annotations of the 80 genes are listed in **Supplementary Table 3**.

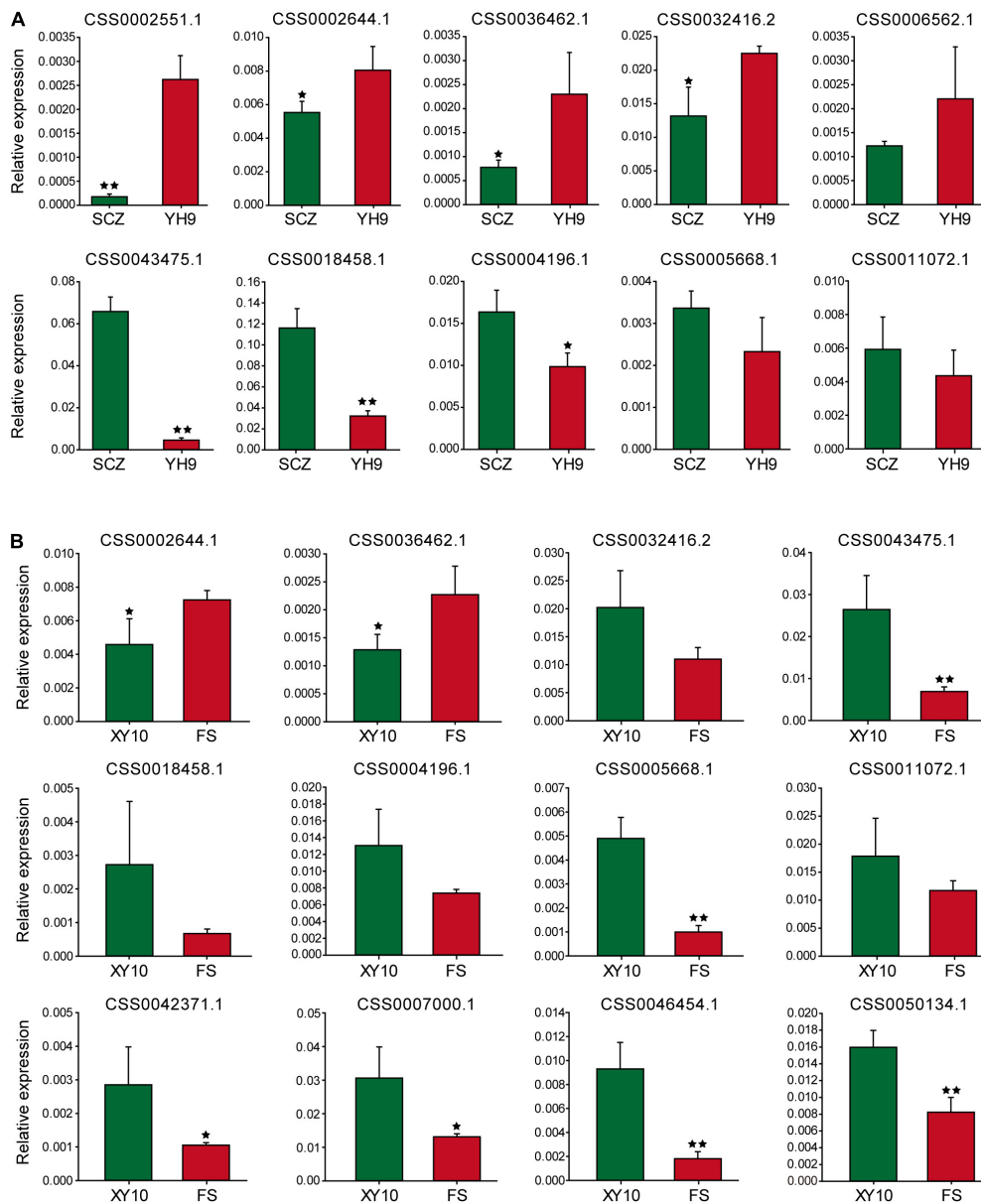
Sixteen genes were selected for qRT-PCR to explore the expression patterns of these genes in large and small leaf samples.

**TABLE 2** | Summary of the information of 25 representative markers (potential QTL).

Linkage group	Position (cM)	Representative marker	LOD	Expl (%)
2	28.329	Marker1533238	8.11	32.8
2	39.826	Marker1713472	7.89	32.1
2	52.559	Marker1890928	7.89	32.1
2	56.223	Marker1980206	7.78	31.7
2	57.436	Marker2005155	8.02	32.5
2	58.648	Marker2011502	7.68	31.4
2	59.138	Marker1755564	7.89	32
2	59.861	Marker1929075	8.02	32.5
2	61.074	Marker1932620	7.78	31.7
2	68.511	Marker1749183	5.57	23.9
2	70.635	Marker1958387	7.48	30.7
2	70.963	Marker1798200	7.69	31.4
2	72.175	Marker1890445	7.69	31.4
2	74.627	Marker1903871	7.87	32
2	75.839	Marker1943728	7.59	31.1
2	77.052	Marker1999608	8.09	32.7
2	84.744	Marker2104738	6.22	26.3
2	87.186	Marker2092844	6.52	27.3
2	91.08	Marker2050694	7.44	30.5
2	92.045	Marker2197408	7.63	31.2
2	93.925	Marker2201972	7.45	30.6
2	94.798	Marker1892873	7.45	30.6
2	97.223	Marker2048068	6.58	27.5
2	99.674	Marker2048163	5.79	24.7
2	100.887	Marker2057762	6.5	27.3

LOD, logarithm of odds; Expl, phenotypic contribution rate.

For “Shuchazao” and “Yinghong 9,” as shown in **Figure 5A**, the expression abundance of 10 genes was successfully detected, and seven of them showed significant differences. Furthermore, two special CSS tea plant varieties, “Xinyang 10” (leaves are narrow and small) and “Foshou” (leaves are round and large) were also



**FIGURE 5 |** The expression of genes near the potential QTL analyzed using qRT-PCR. **(A)** Shuchazao and Yinghong 9, **(B)** Xinyang 10 and Foshou. \* and \*\* represent significant differences when  $p < 0.05$  and  $p < 0.01$ , respectively.

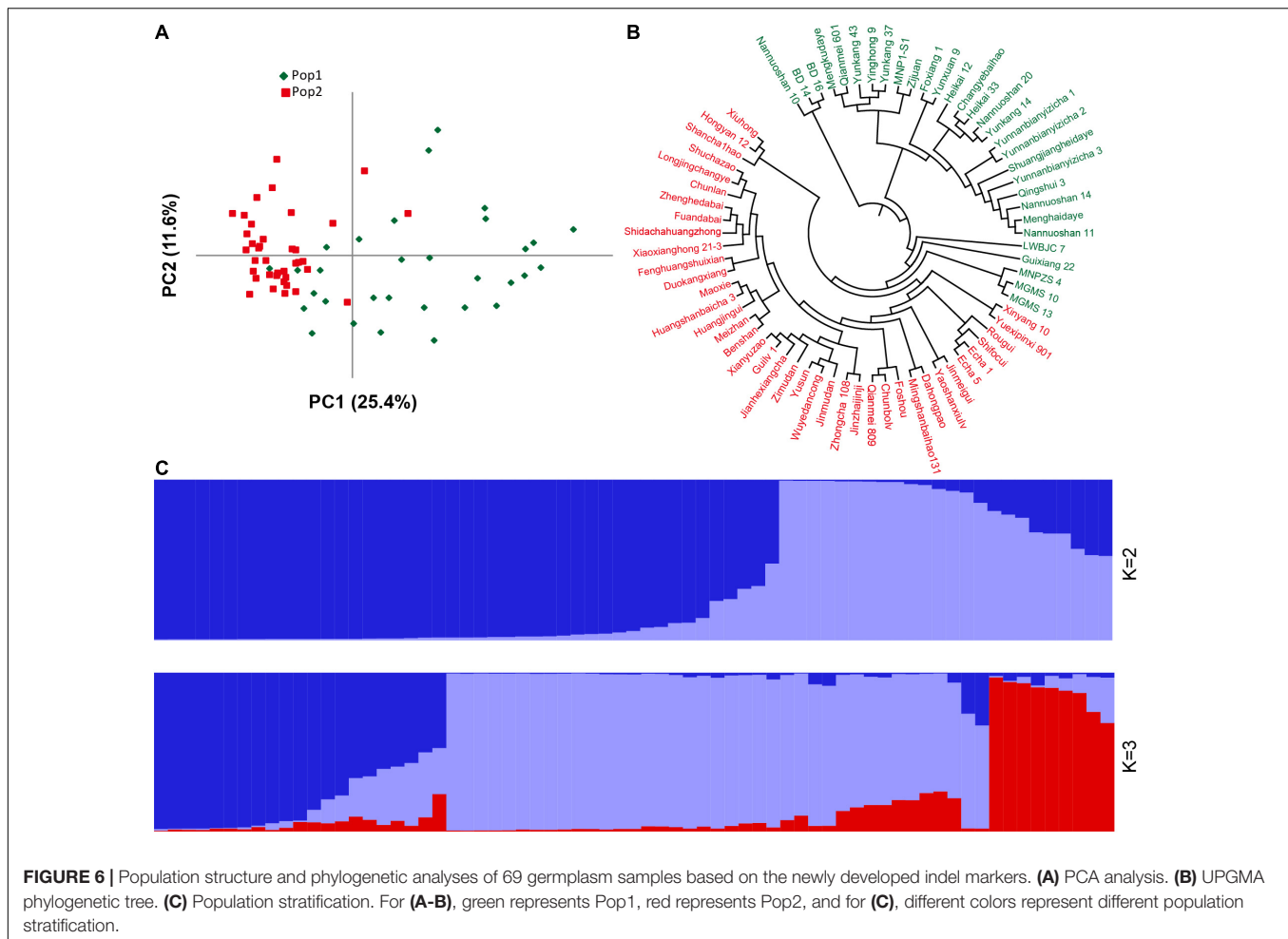
used to verify the expression levels of these genes. The expression levels of 12 out of 16 genes were successfully detected (Figure 5B).

Interestingly, the expression patterns of seven genes were similar to their expression characteristics in “Shuchazao” and “Yinghong 9,” suggesting that these genes may indeed be related to leaf development in tea plants.

## Develop Indel Markers Related to Leaf Area

The mutation loci near potential QTLs are thought to be closely related to the growth and development of tea plants. In order to develop genetic markers related to leaf area, we

randomly selected 60 indel mutations located within 90 kb of the markers associated with potential QTL to design primers. Based on the initial screening results, after excluding eight primer pairs with invalid amplifications, six primer pairs with non-polymorphic amplifications, and 15 primer pairs with ambiguous amplifications, the remaining 31 primer pairs were further applied in 69 germplasms to verify their transferability and polymorphism and to explore their relationship with leaf area. The amplification results are shown in Supplementary Figure 4. Among the newly developed indel markers (Table 3), InDel17 and InDel18 had low polymorphisms, whereas InDel01, InDel03, InDel09, InDel12, and InDel22 had moderate polymorphisms,



and the remaining 24 indel markers were considered to be highly polymorphic (Liu et al., 2019). The genetic diversity of the markers varied from 0.0286 to 0.8354, with an average genetic diversity of 0.622. The PIC and gene diversity were significantly and positively correlated. The observed heterozygosity varied between 0.2903 and 0.7097, with an average of 0.512, whereas the expected heterozygosity ranged from 0.8049 to 0.8604, with an average of 0.837. The marker InDel07 had 12 alleles, whereas InDel17 and InDel18 both had only two alleles. The average number of alleles per marker was 6.48. The MAF varied between 0.2246 and 0.9855, with an average value of 0.5028. Genetic markers with lower major allele frequencies often have higher gene diversity and PIC values.

## Phylogenetic and Population Structure Analysis

There are several of SNP and indel mutations in the vicinity of the potential QTL, and some genomic mutations located in regulatory or coding regions may be related to the development of plant leaves. Therefore, based on the newly developed polymorphic indel markers, we performed population structure and principal component analyses on 69 tea resources. The

results showed that the best model for the 69 tea tree samples was achieved when K was set to 2 (Supplementary Figure 5). Most of the samples were correctly classified together based on different germplasm types when K was 2. As shown in Figure 6, the first category contained the most CSA samples, whereas the second category was mainly composed of the most CSS samples. In particular, when K was set to 3, wild germplasm samples from different regions were clearly classified, implying that our newly developed markers have good discrimination capabilities. Similar clustering results were observed using a PCA (Figure 6A). The variation explanation rates of the first and second principal components were 25.4 and 11.6%, respectively. Furthermore, based on the genetic distance of the newly developed genetic markers, we constructed a phylogenetic tree composed of 69 tea tree samples with the UPGMA model. All the samples were classified into two groups based on a phylogenetic analysis (Figure 6B). Similar to the results of the population structure and principal component analyses, tea varieties with larger leaf areas were assigned to the first group, whereas tea varieties with smaller leaf areas were assigned to the second group (Figures 6A,C). The accurate classification results suggest that the newly developed genetic markers and the genes in which these markers are located may be related to the development of tea leaves.

**TABLE 3** | Sequence and genetic information of the newly developed InDel markers.

Marker	Primer	MAF	Na	Gene diversity	Ho	He	PIC
InDel01	F:AGCGAGGATTCTGTAGAGGAA R:ATGGTCTCTGCTATCATCCCA	0.6957	5	0.4719	0.4516	0.8482	0.4286
InDel02	F:GCCTTTCTGCTTTGGGTTTT R:TGTACGATTTCGATTGAGTGGG	0.5652	4	0.5878	0.7097	0.8392	0.5245
InDel03	F:GAGTTCAAGATTTGTGTCGTCT R:ATCGTCACCTAGGAATTGGGA	0.7319	5	0.4410	0.5161	0.8287	0.4156
InDel04	F:CACTACTCACAGCCACAAGAG R:TGGATAGGTGGGATCAAGTGA	0.3768	11	0.7863	0.4516	0.8408	0.7623
InDel05	F:GATCCGGTGTATCATTACCC R:AACCCCTGAACAACCTTTTCCC	0.3768	6	0.7418	0.6774	0.8361	0.6998
InDel06	F:ATGCATGCCCTTGAACATAGTT R:GACAAGTCACAAGCAGCAAAG	0.3551	10	0.7770	0.5484	0.8604	0.7482
InDel07	F:ACCAATGATTGTTCTGCACCT R:ACAAACCCCACTACAAAAGCT	0.5000	12	0.6839	0.4516	0.8186	0.6497
InDel08	F:CCAGTGATCGATCGAATTGGA R:GCATGAAAAAGCGCGCAATAA	0.4783	9	0.6596	0.6129	0.8456	0.6057
InDel09	F:TGAGAAGGACCAGAGTGAGAA R:TGACTGCCCAAAATTTAGAGCT	0.6739	4	0.4933	0.5161	0.8371	0.4443
InDel10	F:CACATTGGTAAGATAGCCCTT R:AGGCCCTTTTGTAGTTGCA	0.4058	5	0.6577	0.5806	0.8318	0.5932
InDel11	F:CTAAGGCCACAAACATGCTTG R:ACAATATTGGCAGCTCAGGTT	0.5145	4	0.6142	0.5806	0.8281	0.5442
InDel12	F:GTACGCTCAATTTGGTGTTCG R:CAATTCAAAGTCAAAGCCAGGC	0.6014	6	0.5564	0.5161	0.8519	0.4962
InDel13	F:AATGGGCAAAAACAATGGAGC R:CCAAAACAAAACCCAAAGCAG	0.4420	6	0.6971	0.6452	0.8525	0.6503
InDel14	F:TTTGGCCACTCGAAAATGTTT R:GAGGAGTTCTTTGTGTGGGTT	0.4565	4	0.6612	0.5161	0.8498	0.6015
InDel15	F:GAAGGCAAAAGCAAGATGATGG R:ATGCAGCTTCTACTTCCCAAG	0.3043	9	0.7990	0.5161	0.817	0.7712
InDel16	F:GTAAAGTATGCAGGCTCCGAA R:TTGGACCCAGCTACTTCAATG	0.4783	5	0.6796	0.5806	0.8361	0.6329
InDel17	F:GTTTGGCGTTAGATGTGCATT R:ACTTGGATCAGCTTGGAAATGG	0.9855	2	0.0286	0.5161	0.8049	0.0282
InDel18	F:ACATTGGTCATTTCCCTTCA R:AGGCTTTGATGGTTTGGTCAT	0.9783	2	0.0425	0.4194	0.8186	0.0416
InDel19	F:CAACAGCAACACCACTAACAC R:GCACTTCACACCTCAACAATG	0.5652	8	0.6133	0.4516	0.8149	0.5681
InDel20	F:ACTAGTTGCTTTGCTTCGCTA R:TAAGCATCATGAAACACGGGT	0.3623	4	0.6706	0.4194	0.8546	0.6003
InDel21	F:AATATTCGGCAGTGTGTTTCC R:TCAGGTCCGAACAATAGAACT	0.4058	10	0.7438	0.5806	0.8361	0.7093
InDel22	F:CTGATGCTTCCGACCATATGT R:AGAGGAATGAGCCAACCTGTC	0.4783	3	0.5843	0.3871	0.8451	0.4952
InDel23	F:ATTTGGTTGCAGGAAGAGGTG R:GTGGTATGGGATTTGGTAA	0.5870	6	0.5735	0.5161	0.8408	0.5161
InDel24	F:TGCAATTTAGGGGCCACATTA R:GTCCCCTTTCATGAACCACTC	0.5797	7	0.6067	0.4516	0.8197	0.5672
InDel25	F:CACCACAAACAACCCCAAA R:AGGTCTCCTCCATCCAAAGAA	0.2246	9	0.8354	0.4839	0.8228	0.8145
InDel26	F:CGACCTATCCCTTTTCCCA R:GGTCCACGCAGATTTAC	0.4275	8	0.7049	0.4194	0.8176	0.6579
InDel27	F:TTATGGCGAGTTTCGATCTG R:TCTGTAGGAGTCTCTGTGGT	0.5942	4	0.5752	0.4839	0.8435	0.5222
InDel28	F:ACCGTTGGAATGCTCTTAGG R:CATGCGCATAAACGGCTTAAT	0.3188	10	0.8240	0.5484	0.8451	0.8046
InDel29	F:TCCTTTGATCTGGCATGTTT R:TAGCCAGCCTATAGAACGACA	0.2391	11	0.8043	0.4839	0.8599	0.7758

(Continued)



TABLE 3 | Continued

Marker	Primer	MAF	Na	Gene diversity	Ho	He	PIC
InDel30	F:TTCCATAGTATTGCCGCTTG R:GGATCTTTGTGCCTCTCACCG	0.4928	5	0.6332	0.2903	0.8599	0.5666
InDel31	F:AATCTCTCCGCCAGCAATACC R:GGACGAAAACCAATAACTCA	0.3913	7	0.7476	0.5484	0.853	0.7109
Mean	–	0.5028	6.48	0.6224	0.512	0.837	0.5789

Na, number of alleles; MAF, major allele frequency; Ho, observed heterozygosity; He, expected heterozygosity; PIC, polymorphism information content.

## DISCUSSION

Genetic maps are widely used for QTL mapping of key genes of many important agronomic traits (Toojinda et al., 2003; Liu et al., 2015; Ren et al., 2018). Most of the early genetic maps were constructed using randomly amplified polymorphic DNA (RAPD) (Ota, 1999), simple sequence repeats (SSR) (Yamamoto et al., 2005), inter-simple sequence repeats (ISSR) (Kojima et al., 1998), AFLP (amplified fragment length polymorphism) (Marsan et al., 2001), and other molecular markers (Jarvis et al., 1994). Low marker density and large genetic distances make it difficult to apply these genetic maps in practical research. In recent years, with the advancements in technology and the reduction in sequencing costs, large-scale resequencing has been made possible. A large number of genetic markers can be mined using extensive resequencing, which provides strong data-based support for the construction of high-density genetic maps. It was not until 2018 that the genome of the first small-leaf tea plant “Shuchazao” was released (Wei et al., 2018), which greatly promoted basic research on tea trees. However, the underlying mechanisms of many important agronomic traits in tea trees have not yet been revealed (Xia et al., 2020b). The construction of a high-density genetic map is still of great significance for anchoring these markers in tea plant functional genes and breeding of excellent varieties. In recent years, some high-density genetic maps of tea plants have been reported. For example, Xu et al. (2018) constructed a genetic map using 2b-RAD sequencing with the hybrid population of Longjing 43 (“LJ43”) and Baihaozao (“BHZ”); based on specific-locus amplified fragment sequencing, Ma et al. (2015) constructed a genetic map with Yingshuang (“YS”) and Beiyue Danzhu (“BD”) as parents. Although these genetic maps have high marker density because most of the mapping parents belong to CSS varieties, there are still some restrictions on QTL mapping and map-based cloning.

In this study, the whole genome of the CSS tea variety “Jinxuan,” CSA tea variety “Yuncha 1” and their 96 hybrid F1 generations were re-sequenced, and a total of 1.84 Tb of raw sequencing data were obtained. The sequencing depths of the parents and offspring reached 17× and 5×, respectively. Of the clean reads, 84% could be well-aligned to the genome. Compared with previous studies (Su et al., 2017; Rehman et al., 2020), genome-wide resequencing can provide more abundant raw data than ever before. All clean data were submitted to GATK-HaplotypeCaller for mutation detection. In this population, 30,925,005 variations, including

28,144,625 SNPs and 2,780,380 indels, were identified. The average SNP density was 9.6/Kb, which is much higher than that previously reported (An et al., 2020). Similar to other crops (Shen et al., 2017; Thakur and Randhawa, 2018; Muñoz-Espinoza et al., 2020), transitions had an absolute advantage in SNPs, and the ts/tv value was approximately 2.51, which was higher than Cucurbita (Xanthopoulou et al., 2019) but lower than lotus (Huang et al., 2018). Moreover, 242,044 and 185,799 SNPs were annotated as non-synonymous mutations and synonymous mutations, respectively. The ratio of the non-synonymous\_variant to the synonymous\_variant was approximately 1.3, which was higher than those for watermelons (Guo et al., 2019) and sweet potatoes (Zhang et al., 2020), but equivalent to that for rice (Yano et al., 2016).

After a series of filtering of 28,144,625 SNPs, 12,486,697 highly integrated SNPs were obtained. Since 2,864,429 aa × bb type markers could not be used for the construction of the pseudo-testcross population genetic map, the remaining 9,622,268 SNPs were used for further genetic analysis and genotyping. Among the 9,622,268 SNPs, the most abundant types were nn × np and lm × ll, accounting for 46 and 44% of the total, respectively. Finally, a high-quality genetic map containing 8,956 markers with a total length of 1,490.81 cM was successfully constructed. These markers were relatively and evenly distributed across 15 linkage groups, and each linkage group contained an average of 597 genetic markers. The genetic map evaluation revealed that eight of the 15 linkage groups had a maximum gap of less than 5 cM. The integrity of all the markers used in the genetic map exceeded 99% and the parental genotype was complete. A collinearity analysis of the genetic map and genome is also commonly used to evaluate the quality of each linkage group. It was found that the newly constructed genetic map and genome had superior collinearity, with an average spearman coefficient of 0.95. These results indicated that the genetic map constructed in this study was high quality and could be applied to QTL mapping and marker-assisted breeding in tea plants. Compared with the other previously reported genetic maps of tea plants, the total map length here was moderate and contained most of the genetic markers (Ma et al., 2014, 2015; Xu et al., 2018). To the best of our knowledge, this the first high-quality genetic map based on the hybrid population of CSS and CSA constructed using whole-genome resequencing technology.

Leaves are important vegetative organs in plants and provide energy for the growth and development of plants through photosynthesis. Fresh tea leaves have important economic value

because they can be made into a popular beverage with unique flavor (Tai et al., 2015; Xu et al., 2017). CSS and CSA are the two most widely distributed and used tea plant varieties (Wang et al., 2020). However, due to the impacts of many factors such as geographical distribution, environmental adaptation, and artificial selection, they are different in many aspects such as morphological characteristics and content. Most notably, CSA variants tend to have a larger leaf area compared to CSS, and this difference has not been clearly explained. We identified 25 representative markers and two SNPs related to leaf area via QTL mapping and GWAS analysis, which distributed in the second linkage group between 28 and 101 cM. Notably, 25 representative markers may indicate some potential QTLs that have not been identified considering of the small scale of mapping population and elongated QTL shape. All the 25 representative markers (potential QTL) were regarded as major loci (variation contribution rate >10%) with LOD values ranging from 5.57 to 8.11. A significant SNP (L2\_210876632,  $q$  value 5.3E-8) obtained using GWAS was only 1.3 Mb away from potential QTL Marker2197408. Similar results that were obtained using different methods suggested that the QTL mapping was reliable. Many genes that may be related to tea leaf development were identified near the potential QTL. For example, Wang et al. (2018) showed that the CPP-like gene family in *Arabidopsis* (CSS0001030.1) can expand cells and differentiate buds, and Zhou et al. (2021) found that OVATE family protein genes (CSS0018166.1) can regulate peach fruit traits and leaf shape. In order to further explore the expression patterns of these genes in large-leaf and small-leaf tea plants, 16 genes were selected to verify their expression levels using qRT-PCR. The results showed that there were some genes that were significantly and differentially expressed in the large leaf varieties and small-leaf varieties, and the expression trend was related to leaf area and not to whether it was CSS or CSA. Interestingly, the expression level of these genes was very low, which may indicate that the development of tea leaves is regulated more by minor genes.

There are many genetic mutations in the plant genome that play an important role in plant growth and development. Zhou et al. (2021) identified a 1.7 Mb chromosomal inversion that is responsible for regulating the shape of peach fruit; Cheng et al. (2019) found that a single nucleotide mutation in *GID1c* can dwarf peach trees; Wang et al. (2016) found that genetic variations in *ZmVPP1* contributes to drought tolerance in maize seedlings by GWAS. It is these mutations that enhance genetic diversity within species. In this study, based on sufficient genetic variations provided by extensive resequencing, we developed 31 pairs of indel markers related to leaf area located near QTL. Most of these markers located in genes, which may alter gene function. The average gene diversity and PIC were 0.6224 and 0.5789, respectively. Twenty-four of the 31 markers had high polymorphism and good mobility. Based on these markers, we analyzed the population structure and heredity of various tea germplasm samples. The results showed that 69 tea plant resources with different genetic backgrounds were mainly classified into two groups. When  $K = 3$ , obvious genetic structure was observed, and wild tea plant germplasm samples

and cultivated tea plants had significantly different lineages. In the phylogenetic tree, 29 CSA varieties and one CSS species “Guixiang 22” were divided into the first category, whereas 37 CSS varieties and two CSA varieties “Xiuhong” and “Qianmei 809” were divided into the second group. Considering that one parent of “Guixiang 22” belongs to the *C. sinensis* var. *pubilimba* variety and that “Xiuhong” and “Qianmei 809” each have a parent belonging to the CSS class of tea plants, this classification may be reasonable. This result is also supported by the population analysis results.

The growth and development of plant leaves are controlled by many factors, such as light, temperature, disease, nutritional status, and transcriptional regulatory genes (Franklin, 2009; Vega et al., 2019; García-Gómez et al., 2020; Li et al., 2020). For example, Sun et al. (2020) found that OVATE family protein 6 can regulate the angle of rice leaves and Michael (Muszynski et al., 2020) found that cytokinin can change the development pattern of maize leaves. Long-term hybridization means that most cultivated tea trees have complex genetic backgrounds, which brings many challenges for studying the genetic domestication of tea trees and locating key genes that control important agronomic traits. One of them is that the developmental mechanism of tea leaves has not been clearly revealed. In our study, the 31 pairs of polymorphic markers that were developed based on QTL mapping results could divide most tea plants according to their leaf area. However, some special tea varieties, such as “Foshou” had even larger leaf areas than some CSA varieties, but they were clustered together in the second main branch of the phylogenetic tree. Since quantitative traits are often regulated by multiple genes, this result suggested that there might be more unidentified genetic regions that could regulate the development of tea leaves.

## CONCLUSION

In this study, we performed whole-genome resequencing on “Jinxuan,” “Yuncha 1,” and their 96 hybrid F1 samples and identified a large number of SNPs and indel mutation loci. A high-quality and high-density genetic map of tea plants with good collinearity with the reference genome was successfully constructed. Using QTL mapping and GWAS, 25 representative markers (potential QTL) were identified, which distributed in the second linkage group with an average phenotypic contribution rate more than 30% and containing two SNP mutations related to leaf area. A total of 31 polymorphic indel primers near the potential QTL were developed, and PCR was performed in 69 diverse tea germplasm samples to detect the mobility and discrimination capabilities of these markers. Population structure and phylogenetic analyses showed that the newly developed markers can better distinguish large-leaf species from small-leaf species. However, due to their complex genetic background, there may be more unknown genetic regions in the genome that may play an important role in regulating the development of tea leaves. To conclude, in this study, besides preliminary QTL mapping of genes that regulate tea leaves, the newly constructed genetic map and developed molecular markers will provide basic support for tea breeding and quantitative trait research.

## MATERIALS AND METHODS

### Plant Materials

Ninety-six hybrid F1 generations and their parents “Jinxuan” and “Yuncha 1” were cultivated in the national germplasm repository of large-leaf tea in Menghai, China. The young buds and leaves were picked and quickly frozen and subsequently stored in a  $-80^{\circ}\text{C}$  freezer until they were sequenced. Another 69 germplasms from different countries and regions with different leaf shapes were used to evaluate the newly developed genetic markers. Detailed information on these samples was listed in **Supplementary Table 4**.

### DNA Extraction and Whole-Genome Resequencing

The DNA of all the samples was extracted from young leaves using a modified CTAB method, and the extracts were electrophoresed on a 0.8% agarose gel to check their quality and integrity. Qualified DNA samples were broken into short fragments of 200–500 bp using ultrasound, and then paired-end sequencing was performed on F1 generations using the Illumina HiSeq 2500 sequencing platform. Using the Trimmomatic software, the original data were filtered according to the following standards: (1) remove adapters, (2) remove paired reads if the ratio of N in the sequenced fragments is greater than 10%, and (3) remove reads if more than 50% of the bases had a base quality value  $q < 10$  in the sequenced fragments and ensure that the bases with  $Q > 20$  are above 90%.

### Detection and Filtering of Genetic Variations and Construction of Genetic Maps

The filtered reads were aligned to the reference genome “Shuchazao” through the built-in MEM module of the BWA (Burrows-Wheeler Aligner) program. The Picard script was used to remove PCR duplications in the BAM files (Liu et al., 2019). These files were then submitted to GATK-HaplotypeCaller for mutation detection (Niu et al., 2019). The raw mutation site files of each sample were merged and filtered according to the following criteria: (1) two SNP sites within 5 bp, (2) SNP sites within 10 bp near indel, (3) two indel with a distance less than 10 bp, (4) for diploid species, one SNP marker locus can contain at most four genotypes, so SNP loci with more than four alleles were discarded subsequently. Only SNPs with two to four alleles were identified as polymorphic and considered potential markers. All polymorphism SNPs loci were genotyped with consistency in the parental and offspring SNP loci. In order to facilitate analysis, all the SNP locus genotypes were coded following the two-allele coding rules commonly used in genetics. Among them, the  $aa \times bb$  genotype locus was directly removed because it was not suitable for constructing a genetic map in this project. The remaining SNPs were filtered according to the following criteria to ensure the quality of the map: (1) the sequencing depth of the parents is less than  $10\times$ , (2) the sequencing depth of the offspring is less than  $2\times$ , (3) the completeness of the genotype is less than 90%, (4) the  $p$ -values of segregation distortion are

less than 0.05, and (4) the SNPs have MLOD values lower than 3. With the help of the HighMap software (Xu et al., 2015), the linear arrangement of the markers in the linkage group was obtained, and the genetic distance between the adjacent markers was estimated. At the same time, the previous map (Xu et al., 2018) were integrated using the BioMercator (Sosnowski et al., 2012) software, and finally, the high-quality genetic map was drawn. Besides, vcftools was used to perform statistical analysis on the final variations. Based on the index file of the “Shuchazao” genome constructed using SnpEff, all the variation sites were annotated and classified.

### QTL Mapping and GWAS

The length and width of the six mature leaves of each F1 individual were accurately measured, and the leaf area of each leaf was calculated according to the formula: leaf length  $\times$  leaf width  $\times 0.7$  (NY/T 1312-2007), and then the average of the six leaf areas was taken as the leaf area of the offspring. The built-in interval mapping method of the MapQTL software was used to locate the QTL of the leaf area, and all the default parameters of the software were used. With the PLINK software, all the SNP loci were filtered under the conditions of  $-maf 0.05$ ,  $-geno 0.1$ , and  $-indep-pairwise 50 10 0.2$ . Based on the 1,000,682 high-quality SNP loci of non-linkage disequilibrium obtained after filtering, the three models (including GLM, MLM, and CMLM) provided by the TASSEL (Bradbury et al., 2007) software was used for the GWAS. Similarly, the kinship and principal component analysis (PCA) files of all the samples were calculated under the default parameters of the TASSEL software. The Manhattan map and linkage group were drawn using the CMplot and LinkageMapView packages in R.

### RNA Extraction and Quantitative Real-Time PCR (qRT-PCR)

The young second leaves of “Shuchazao” and “Yinghong 9” were collected from the Tea Tree Resource Nursery of the Anhui Agricultural University, and their total RNAs were extracted using the Total RNA Purification Kit (Norgen Biotek Corporation, Canada) following the protocol of the manufacturer. In addition, a CSS variety called “Foshou” which has a very large-leaf area, and a CSS variety called “Xinyang 10” which has a small leaf area, were also collected from the Guangdong province and their RNAs were extracted as described above. Subsequently, first-strand cDNA was synthesized based on the qualified RNA using a PrimeScript RT Reagent Kit (cat 6110A, Takara, Japan) for qRT-PCR. Detailed PCR amplification and thermal cycling conditions were performed referring to our previous research. The relevant primer sequences are listed in **Supplementary Table 5**.

### Development of Indel Markers

In the 2 Mb region before and after potential QTL, indels with a length  $\geq 4$  were used as candidate sites. The 300 bp double-flanking sequence of the indel was extracted as the primer design region. The primers were designed using Primer3 based on the following criteria: amplicon lengths should range from

150 to 350 bp; primer lengths, 20–22 bp, with the optimum length being 21 bp; T<sub>m</sub> 55–60°C, with 58°C being the optimum, and the maximum difference between the annealing temperatures of the primer pairs should not exceed 5°C; GC content 40–60%, with 55% being the optimum. A total of 60 indels located near the potential QTL that met the above conditions were randomly selected, and the primers were synthesized. Eight germplasms, including “Hekai10,” “Hekai38,” “Nannuoshan 9,” “Nannuoshan 13,” “Yaoshanxiulv,” “Shidacha 1,” “Xiaoxianghong 21-3,” and “Echa 1” were used to screen all the primers preliminarily using the Fragment Analyzer<sup>TM</sup> 96 system (Advanced Analytical Technologies, Inc., Ames, IA, United States). More detailed amplification conditions and primer screening can be found in our previous study.

## Genetic Diversity, Population Structure and Phylogenetic Analysis

The selected primers were further applied to 69 diverse tea plant samples to verify their mobility and polymorphism. The amplification results were displayed graphically using prosize2.0 built in the Fragment Analyzer<sup>TM</sup> 96 system, and the amplified fragment length was identified and converted into letter codes. Subsequently, the expected heterozygosity (He), observed heterozygosity (Ho), number of alleles (Na), major allele frequency (MAF), and polymorphism information content (PIC) were also calculated using PowerMarker3.25 and PopGene32. Concurrently, the phylogenetic tree of 69 diverse tea plant samples was also constructed using the unweighted pair-group method with the arithmetic means (UPGMA) algorithm using PowerMarker at the default setting of 1,000 bootstraps. The population structure analysis was performed with reference to previous reports (Liu et al., 2019). A web-based online software, Structure Harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>), was used to evaluate the best population *k* value. GenALEX 6.5 (Liu et al., 2017) was used to perform the PCA analysis.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI database, BioProject ID is PRJNA727668.

## AUTHOR CONTRIBUTIONS

YA performed the data analysis and manuscript drafting. CW and SL conceived the project and designed the research. LC and LT performed the hybridization, experiments, interpreted

the results, and analyzed the data. All authors have read and approved the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (U20A2045), the Project of Major Science and Technology in Anhui Province (202003a06020021), the National Key Research and Development Program of China (2019YFD1001601), the Base of Introducing Talents for Tea plant Biology and Quality Chemistry (D20026), the Special Funds for the Tea Germplasm Resource Garden (2060502 and 201834040003), and the Project for Construction Technology Innovation Center and Achievements Industrialization for Large Leaf of Tea Plant in Yunnan Province.

## ACKNOWLEDGMENTS

The authors are greatly thankful to Yiping Tian and Chunlin Chen from the Tea Research Institute, Yunnan Academy of Agricultural Sciences and our team member Enhua Xia for their help with this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.705285/full#supplementary-material>

**Supplementary Figure 1** | Location information of markers in the genetic map.

**Supplementary Figure 2** | Manhattan and QQ plots of GLM, MLM, and CMLM models.

**Supplementary Figure 3** | The position of SNP loci on the second linkage group obtained by GLM model.

**Supplementary Figure 4** | Amplification polymorphism of 31 pairs of indel markers among the 69 germplasm samples.

**Supplementary Figure 5** | Estimation of the optimal group number through  $\Delta K$ . The value of *K* was set from 2 to 8.

**Supplementary Table 1** | Summary of the sequencing data of the F1 generation and their parents.

**Supplementary Table 2** | Average leaf area information for each offspring.

**Supplementary Table 3** | Annotation information of genes within 90 kb of both sides of the 25 representative markers (potential QTL).

**Supplementary Table 4** | Tea plant samples information used in indel marker development.

**Supplementary Table 5** | Primer information used for qRT-PCR.

## REFERENCES

An, Y. L., Mi, X. Z., Zhao, S. Q., Guo, R., Xia, X. B., Liu, S. R., et al. (2020). Revealing Distinctions in Genetic Diversity and Adaptive Evolution Between

Two Varieties of *Camellia sinensis* by Whole-Genome Resequencing. *Front. Plant Sci.* 11:603819. doi: 10.3389/fpls.2020.603819  
Bradbury, P. J., Zhang, Z. W., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. T. A. S. E. L. (2007). software for association mapping of

- complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Cheng, J., Zhang, M. M., Tan, B., Jiang, Y. J., Zheng, X. B., Ye, X., et al. (2019). A single nucleotide mutation in *GID1c* disrupts its interaction with *DELLA1* and causes a GA-insensitive dwarf phenotype in peach. *Plant Biotechnol. J.* 17, 1723–1735. doi: 10.1111/pbi.13094
- Franklin, K. A. (2009). Light and temperature signal crosstalk in plant development. *Curr. Opin. Plant Biol.* 12, 63–68. doi: 10.1016/j.pbi.2008.09.007
- García-Gómez, M. L., Castillo-Jiménez, A., Martínez-García, J. C., and Álvarez-Buylla, E. R. (2020). Multi-level gene regulatory network models to understand complex mechanisms underlying plant development. *Curr. Opin. Plant Biol.* 57, 171–179. doi: 10.1016/j.pbi.2020.09.004
- Guo, S. G., Zhao, S. J., Sun, H. H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623. doi: 10.1038/s41588-019-0518-4
- Hu, C. Y., Lee, T. C., Tsai, H. T., Tsai, Y. Z., and Lin, S. F. (2013). Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). *Euphytica* 191, 141–152. doi: 10.1007/s10681-013-0908-0
- Huang, L. Y., Yang, M., Li, L., Li, H., Yang, D., Shi, T., et al. (2018). Whole genome re-sequencing reveals evolutionary patterns of sacred lotus (*Nelumbo nucifera*). *J. Integrat. Plant Biol.* 60, 2–15. doi: 10.1111/jipb.12606
- Jarvis, P., Lister, C., Szabo, V., and Dean, C. (1994). Integration of CAPS markers into the RFLP map generated using recombinant inbred lines of *Arabidopsis thaliana*. *Plant Mol. Biol.* 24, 685–687. doi: 10.1007/bf00023565
- Huang, J. A., Li, J. X., Huang, Y. H., Luo, J. W., Gong, Z. H., and Liu, Z. H. (2005). Construction of AFLP molecular markers linkage map in tea plant. *J. Tea Sci.* 25, 7–15.
- Kojima, T., Nagaoka, T., Noda, K., and Ogihara, Y. (1998). Genetic linkage map of ISSR and RAPD markers in Einkorn wheat in relation to that of RFLP markers. *Theoret. Appl. Genet.* 96, 37–45. doi: 10.1007/s001220050706
- Ayala, V. R., Kunapuli, S., and Manguashca, J. (2015). Development of yield prediction models in the maize crop using spectral data for precision agriculture applications. *Ecuador Es Calidad* 2:v211.5.
- Li, Y. Y., Mi, X. Z., Zhao, S. Q., Zhu, J. Y., Guo, R., Xia, X. B., et al. (2020). Comprehensive profiling of alternative splicing landscape during cold acclimation in tea plant. *BMC Genom.* 21:65. doi: 10.1186/s12864-020-6491-6
- Liu, H. J., Niu, Y. C., Gonzalez-Portilla, P. J., Zhou, H. K., Wang, L. Y., Zuo, T., et al. (2015). An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. *BMC Genom.* 16:1078. doi: 10.1186/s12864-015-2242-5
- Liu, S. R., An, Y. L., Tong, W., Qin, X. J., Samarina, L., Guo, R., et al. (2019). Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. *BMC Genomics* 20:935. doi: 10.1186/s12864-019-6347-0
- Liu, S. R., Liu, H. W., Wu, A. L., Hou, Y., An, Y. L., and Wei, C. L. (2017). Construction of fingerprinting for tea plant (*Camellia sinensis*) accessions using new genomic SSR markers. *Mol. Breed.* 37, 1–14.
- Ma, J. Q., Huang, L., Ma, C. L., Jin, J. Q., Li, C. F., Wang, R. K., et al. (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:e0128798. doi: 10.1371/journal.pone.0128798
- Ma, J. Q., Yao, M. Z., Ma, C. L., Wang, X. C., Jin, J. Q., Wang, X. M., et al. (2014). Construction of a SSR-based genetic map and identification of QTLs for catechins content in tea plant (*Camellia sinensis*). *PLoS One* 9:e93131. doi: 10.1371/journal.pone.0093131
- Meegahakumbura, M. K., Wambulwa, M. C., Thapa, K. K., Li, M. M., Möller, M., Xu, J. C., et al. (2016). Indications for three independent domestication events for the tea plant (*Camellia sinensis* (L.) O. Kuntze) and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatellites. *PLoS One* 11:e0155369. doi: 10.1371/journal.pone.0155369
- Muñoz-Espinoza, C., Di Genova, A., Sánchez, A., Corraet, J., and Hinrichsen, P. (2020). Identification of SNPs and InDels associated with berry size in table grapes integrating genetic and transcriptomic approaches. *BMC Plant Biol.* 20:365. doi: 10.1186/s12870-020-02564-4
- Muszynski, M. G., Moss-Taylor, L., Chudalayandi, S., Cahill, J., Valleechevarria, A., Alvarezcastro, I., et al. (2020). The maize hairy sheath frayed1 (*Hsf1*) mutation alters leaf patterning through increased cytokinin signaling. *Plant Cell* 32, 1501–1518. doi: 10.1105/tpc.19.00677
- Niu, S. Z., Song, Q. F., Koiwa, H., Qiao, D. H., Zhao, D. G., Chen, Z. W., et al. (2019). Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biol.* 19:328. doi: 10.1186/s12870-019-1917-5
- Ota, S. R. A. P. D. - (1999). based linkage mapping using *F1* segregating populations derived from crossings between tea cultivar 'Sayamakaori' and strain 'Kana-Ck17'. *Breed. Res.* 1:16.
- Marsan, P. A., Gorni, C., Chitto, A., Redaelli, R., Vijk, R., Stam, P., et al. (2001). Identification of QTLs for grain yield and grain-related traits of maize (*Zea mays* L.) using an AFLP map, different testers, and cofactor analysis. *Theoret. Appl. Genet.* 102, 230–243. doi: 10.1007/s001220051640
- Rehman, F., Gong, H., Li, Z., Zeng, S., and Wang, Y. (2020). Identification of fruit size associated quantitative trait loci featuring SLAF based high-density linkage map of goji berry (*Lycium spp.*). *BMC Plant Biol.* 20:474. doi: 10.1186/s12870-020-02567-1
- Ren, Y., Guo, S. G., Zhang, J., He, H. J., Sun, H. H., Tian, S. W., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* 176, 836–850. doi: 10.1104/pp.17.01290
- Shen, C., Xin, J., Zhu, D., and Lin, Z. X. (2017). Uncovering SNP and indel variations of tetraploid cottons by SLAF-seq. *BMC Genomics* 18:247. doi: 10.1186/s12864-017-3643-4
- Sosnowski, O., Charcosset, A., and Joets, J. (2012). BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* 28, 2082–2083.
- Su, C. F., Wang, W., Gong, S. L., Zuo, J. H., Li, S. J., and Xu, S. Z. (2017). High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Front. Plant Sci.* 8:706. doi: 10.3389/fpls.2017.00706
- Sun, X. X., Ma, Y. M., Yang, C., and Li, J. X. (2020). Rice *OVATE* family protein 6 regulates leaf angle by modulating secondary cell wall biosynthesis. *Plant Mol. Biol.* 104, 249–261.
- Tai, Y. L., Wei, C. L., Yang, H., Zhang, L., Chen, Q., Deng, W. W., et al. (2015). Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*). *BMC Plant Biol.* 15:190. doi: 10.1186/s12870-015-0574-6
- Tan, L. Q., Wang, L. Y., Xu, L. Y., Wu, L. Y., Peng, M., Zhang, C. C., et al. (2016). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (*Camellia sinensis*). *Tree Genet. Genomes* 12:52.
- Tan, L. Q., Wang, L. Y., Wei, K., Zhang, C. C., Wu, L. Y., Qi, G. N., et al. (2013). Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS One* 8:e81611. doi: 10.1371/journal.pone.0081611
- Thakur, O., and Randhawa, G. S. (2018). Identification and characterization of SSR, SNP and InDel molecular markers from RNA-Seq data of guar (*Cyamopsis tetragonoloba*, L. Taub.) roots. *BMC Genomics* 19:951. doi: 10.1186/s12864-018-5205-9
- Toojinda, T., Siangliw, M., Tragoonrun, S., and Vanavichit, A. (2003). Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. *Ann. Bot.* 91, 243–253.
- Vega, A., O'Brien, J. A., and Gutiérrez, R. A. (2019). Nitrate and hormonal signaling crosstalk for plant growth and development. *Curr. Opin. Plant Biol.* 52, 155–163.
- Wang, W. P., Sijacic, P., Xu, P. B., Lian, H. L., and Liu, Z. C. (2018). *Arabidopsis* *TSO1* and *MYB3R1* form a regulatory module to coordinate cell proliferation with differentiation in shoot and root. *Proc. Natl. Acad. Sci.* 115, E3045–E3054.
- Wang, X. C., Feng, H., Chang, Y. X., Ma, C. L., Wang, L. Y., Hao, X. Y., et al. (2020). Population sequencing enhances understanding of tea plant evolution. *Nat. Commun.* 11, 1–10.
- Wang, X. L., Wang, H. W., Liu, S. X., Ferjani, A., Li, J. S., Yan, J. B., et al. (2016). Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48, 1233–1241.

- Wei, C. L., Yang, H., Wang, S. B., Zhao, J., Liu, C., Gao, L. P., et al. (2018). Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* 115, E4151–E4158.
- Xanthopoulou, A., Montero-Pau, J., Mellidou, I., Kissoudis, C., Blanca, J., Picó, B., et al. (2019). Whole-genome resequencing of *Cucurbita pepo* morphotypes to discover genomic variants associated with morphology and horticulturally valuable traits. *Horticult. Res.* 6, 1–17.
- Xia, E. H., Tong, W., Hou, Y., An, Y. L., Chen, L. B., Wu, Q., et al. (2020a). The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* 13, 1013–1026.
- Xia, E. H., Tong, W., Wu, Q., Wei, S., Zhao, J., Zhang, Z. Z., et al. (2020b). Tea plant genomics: achievements, challenges and perspectives. *Horticult. Res.* 7, 1–19.
- Xu, L. Y., Wang, L. Y., Wei, K., Tan, L. Q., Su, J. J., and Cheng, H. (2018). High-density SNP linkage map construction and QTL mapping for flavonoid-related traits in a tea plant (*Camellia sinensis*) using 2b-RAD sequencing. *BMC Genomics* 19:955. doi: 10.1186/s12864-018-5291-8
- Xu, Q. S., Zhu, J. Y., Zhao, S. Q., Hou, Y., Li, F. D., Tai, Y. L., et al. (2017). Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front. Plant Sci.* 8:1205. doi: 10.3389/fpls.2017.01205
- Xu, X. W., Xu, R. X., Zhu, B. Y., Yu, T., Qu, W. Q., Lu, L., et al. (2015). A high-density genetic map of cucumber derived from Specific Length Amplified Fragment sequencing (SLAF-seq). *Front. Plant Sci.* 5:768. doi: 10.3389/fpls.2014.00768
- Yamamoto, T., Yamaguchi, M., and Hayashi, T. (2005). An integrated genetic linkage map of peach by SSR, STS, AFLP and RAPD. *Engei Gakkai Zasshi* 74, 204–213.
- Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48:927.
- Zhang, L., Yu, Y. C., Shi, T. Y., Kou, M., Sun, J., Xu, T., et al. (2020). Genome-wide analysis of expression quantitative trait loci (eQTLs) reveals the regulatory architecture of gene expression variation in the storage roots of sweet potato. *Horticult. Res.* 7, 1–12.
- Zhao, S. Q., Wang, X. W., Yan, X. M., Guo, L. X., Mi, X. Z., Xu, Q. S., et al. (2018). Revealing of microRNA involved regulatory gene networks on terpenoid biosynthesis in *Camellia sinensis* in different growing time points. *J. Agricult. Food Chem.* 66, 12604–12616.
- Zhou, H., Ma, R. J., Gao, L., Zhang, J. Y., Zhang, A. D., Zhang, X. J., et al. (2021). A 1.7-Mb chromosomal inversion downstream of a PpOFP1 gene is responsible for flat fruit shape in peach. *Plant Biotechnol. J.* 19:192.

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 An, Chen, Tao, Liu and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.