


Article

Quantitative Trait Loci Mapping and Association Analysis of Solanesol Content in Tobacco (*Nicotiana tabacum* L.)

Jing Liu ^{1,†}, Dehu Xiang ^{2,†}, Yongmei Du ¹, Zhongfeng Zhang ¹, Hongbo Zhang ¹, Lirui Cheng ¹, Qiujuan Fu ¹, Ning Yan ¹ , Fuzhu Ju ¹, Chaofan Qi ¹, Yunkang Lei ³, Jun Wang ^{4,*} and Yanhua Liu ^{1,*}

- ¹ Plant Functional Ingredient Research Center, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, 11 Keyuanjingsi Road, Qingdao 266101, China; liujing05302021@hotmail.com (J.L.); duyongmei@caas.cn (Y.D.); zhangzhongfeng@caas.cn (Z.Z.); zhanghongbo@caas.cn (H.Z.); chenglirui@caas.cn (L.C.); fuqiujuan@caas.cn (Q.F.); yanning@caas.cn (N.Y.); m13020502056@163.com (C.Q.)
- ² Shimen County Branch of Changde Company of Hunan Tobacco Company, Changde 415300, China; 1559525@163.com
- ³ Sichuan Tobacco Corporation Deyang Branch, Deyang 618400, China; m15902878236@163.com
- ⁴ Tobacco Science Institute of Guangdong Province, Shaoguan 512026, China
- * Correspondence: wangjun4170@126.com (J.W.); liuyanhua@caas.cn (Y.L.)
- † These authors contributed equally to this work.

Abstract: Solanesol, which accumulates predominantly in the leaves of tobacco plants, has medically important bioactive properties. To investigate the genetic basis of solanesol in tobacco (*Nicotiana tabacum*), the solanesol contents of 222 accessions, 206 individuals from an *N. tabacum* Maryland609 (low-solanesol) × K326 (high-solanesol) F2 population and their corresponding F1 self-pollinations, were determined using ultra-performance liquid chromatography. Genome-wide quantitative trait locus (QTL) and association analysis were performed to identify QTLs and markers associated with solanesol content based on simple sequence repeat molecular markers. A total of 12 QTLs underlying solanesol content were mapped to seven linkage groups (LGs), with three of the QTLs (QTL3-1, QTL21-6, and QTL23-3) explaining 5.19–10.05% of the phenotypic variation. Association analysis revealed 38 significant marker-trait associations in at least one environment. The associations confirmed the QTLs located on LG3, LG10, LG14, LG21, and LG23, while new elite makers were located on 11 additional LGs, each explaining, respectively, 5.16–20.07% of the phenotypic variation. The markers LG14-PT54448, LG10-PT60114-2, LG10-PT60510, LG10-PT61061, and LG-21PT20388 may be useful for molecular-assisted selection of solanesol content in tobacco leaves. These results increase our understanding of the inheritance of solanesol-associated genes and will contribute to molecular-assisted breeding and further isolation of regulatory genes involved in solanesol biosynthesis in tobacco leaves.

Keywords: *N. tabacum*; solanesol content; QTL; association analysis



Citation: Liu, J.; Xiang, D.; Du, Y.; Zhang, Z.; Zhang, H.; Cheng, L.; Fu, Q.; Yan, N.; Ju, F.; Qi, C.; et al. Quantitative Trait Loci Mapping and Association Analysis of Solanesol Content in Tobacco (*Nicotiana tabacum* L.). *Agronomy* **2024**, *14*, 1370. <https://doi.org/10.3390/agronomy14071370>

Academic Editor: Mirosław Tyrka

Received: 22 May 2024

Revised: 13 June 2024

Accepted: 19 June 2024

Published: 26 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nicotiana tabacum L., a member of the Solanaceae family, is native to the Americas, Oceania, and the South Pacific but is widely cultivated throughout the world as an important economic crop [1,2]. Tobacco leaves contain an abundance of secondary metabolites such as solanesol, rutin, and chlorogenic acid, which have significant medicinal value. Solanesol in particular has many useful properties, including anti-fungal, anti-viral, anti-inflammatory, and anti-ulcer activities. In addition, solanesol can be used to synthesize coenzyme Q10, vitamin K analogs, and derivatives for the treatment of cardiovascular diseases [3–7]. Despite extensive research, the industrial-scale production of solanesol is limited by the complicated synthesis pathway of the 45-carbon backbone [8]. Currently, the solanesol used by the pharmaceutical industry is mainly extracted from the leaves of tobacco plants [4]. Therefore, it is very important to study the genetic basis of solanesol

in tobacco, which might contribute to enhancing the solanesol exploitation and use of tobacco leaves.

DNA-based molecular markers have been widely used in plant genetic research because they are generally unaffected by environmental and agronomic factors. Recently, large-scale simple sequence repeat (SSR) markers with the advantages of co-dominance, abundant producibility, and high stability and specificity have been successfully developed [9–12]. These SSR markers have been used for quantitative trait locus (QTL) mapping and association mapping of quantitative traits in tobacco. For example, *Abl* and *BMVSE* were identified as being associated with cis-abienol and sucrose ester accumulation following linkage group analysis based on a doubled haploid population [13]. In addition, linkage mapping successfully identified two major QTLs associated with tobacco plant height, with the results confirmed by combined association analysis [14]. These previous studies using SSR markers have provided a theoretical basis and technical support for the breeding of new high-yielding, high-quality tobacco varieties [14–26].

Notably, the genetic basis of solanesol production is less well understood than other tobacco traits, limiting the exploitation of its medicinal value. Levels of solanesol accumulation vary considerably between different tobacco varieties, organs, and developmental stages of tobacco plants. For example, the solanesol content of flue-cured and cigar tobaccos is higher than that of burley and oriental tobaccos. However, the solanesol content of tobacco leaves is higher than that in any other organ of the plant and generally reaches a maximum during vegetative growth [26]. Joint segregation analysis of a major gene plus the polygene mixed genetic model revealed that the solanesol content of tobacco leaves is predominantly controlled by a set of genes with a heritability of 56–65% [27]. Although the genes encoding key enzymes in the solanesol synthesis pathway have been identified, molecular markers associated with these genes have not yet been identified. Such information would be helpful to reduce the detection cost of secondary metabolites and improve the selection efficiency during the early stages of plant growth.

In this study, the solanesol content of mature middle leaves of parental, F₁, and F₂ plants (n = 206), along with 222 accessions, was determined by ultra-performance liquid chromatography (UPLC) analysis. By combining QTL mapping and association analysis, molecular markers putatively associated with solanesol content were identified and used to explain the phenotypic variation. The information obtained in this study will potentially aid in germplasm selection and breeding for specific solanesol content and provide a novel approach to studying the genetic basis of secondary metabolites in tobacco.

2. Materials and Methods

2.1. Plant Materials

Based on the identified solanesol content and diversity data from our previous study [28], we selected the high-solanesol *Nicotiana tabacum* cultivar K326 and the low-solanesol *N. tabacum* cultivar Maryland609 as parental lines. Maryland609 was used as the maternal parent and K326 as the paternal parent. An F₂ population consisting of 206 individuals derived from F₁ self-pollinations was used for QTL mapping. A panel of 222 tobacco accessions selected from tobacco core collections (listed in Table A1 of Appendix A) was used for association analysis. All tobacco germplasm resources, including 86 introduced, 80 breeding, and 56 local germplasms, were provided by the National Infrastructure for Crop Germplasm Resource (Tobacco; Qingdao, China) of the Chinese Academy of Agricultural Sciences.

2.2. Field Trial Design

Field trials for QTL mapping were conducted at the Hubei Burley Experimental Station, Hubei Province, China. P₁, P₂, F₁, and F₂ populations were sown at the experimental station in 2015. A total of 50 individuals were planted for each of the P₁, P₂, and F₁ populations, while 250 individuals were planted for the F₂ population. Plants were grown at a density of 25 plants per row, with a plant spacing of 50 cm and a row spacing of 120 cm.

For association analysis, 222 natural accessions were planted at 4 experimental stations (E1, E2, E3, E4) located in Shandong and Sichuan provinces. The experiments were conducted at Xichang (E1) of Sichuan, Zhucheng (E2) of Shandong in 2014, and Huili (E3) of Sichuan, Jimo (E4) of Shandong in 2015. Sichuan and Shandong represent the southern and northern tobacco-growth regions of China, respectively. The field trials of natural accessions were arranged in a randomized block design and replicated three times. The inter-plant and inter-row spacing were the same as described above.

2.3. Sampling and Treatment

Middle leaves (mature stage; $n = 3$) were collected from each plant 90 days after transplanting. The main vein was removed and the three leaves from each plant were combined, wrapped in foil, and stored in the freezer at $-20\text{ }^{\circ}\text{C}$. After freeze-drying, the lyophilized tobacco leaves were ground using a Q-400B steel grain mill grinder (Shanghai Bingdu Electric Co., Ltd., Shanghai, China). For QTL mapping, solanesol content was determined in three leaf samples from each plant (all P1, P2, F1, and F2 individuals). For association analysis, leaf samples from three representative individuals were used to determine solanesol content for each replication of the 222 accessions.

2.4. UPLC-Based Quantification of Solanesol Content

Samples were prepared as follows. Powdered, freeze-dried leaf samples were passed through a 40-mesh sieve. A 0.1 g aliquot (to the nearest 0.0001 g) of each sample was added to a 20 mL glass centrifuge tube with a polytetrafluoroethylene stopper (spica, Shanghai Jiayi Biotechnology Co., Ltd., Shanghai, China). A 1 mL volume of 0.1 M sodium hydroxide ethanol solution and 5.0 mL of n-hexane were then added sequentially to each sample tube. The tubes were then capped, shaken, and placed in a thermostatically controlled ultrasonic extractor at a frequency of 45 kHz and a temperature of $40\text{--}50\text{ }^{\circ}\text{C}$ for 30 min. After cooling, 8.0 mL of deionized water was added to each sample, then the tubes were capped tightly and centrifuged at $16,000\times g$ for 10 min. A 500 μL aliquot of the top layer of the n-hexane solanesol extract was collected from each tube and mixed with 4.5 mL of acetonitrile (mobile phase). The mixtures were then filtered through 0.2 μM filters (spica, Shanghai Jiayi Biotechnology Co., Ltd., Shanghai, China). Quantification of solanesol content was performed by UPLC (Waters Technologies Ltd., Milford, MA, USA) analysis as described by Pan et al. [29] under the optimized conditions described by Xiang [27]. Comparisons between the P1, P2, F1, and F2 populations and the 222 natural accessions were performed using SPSS 23.0 analysis software (IBM, Corp., Armonk, NY, USA).

2.5. DNA Extraction and Selection of Polymorphic Primers

Genomic DNA was isolated from individual plants of the F2 population and mixed samples of young leaves of 10 plants from the P1, P2, and F1 populations and each accession using the CTAB method [30] optimized by Xiang [28]. The SSR primers used in this study were synthesized according to the sequence published by Bindler and Tong [10–12]. PCR amplification was performed as described by Xiang [28]. A total of 1880 pairs of SSR primers evenly distributed among 24 linkage groups were first screened for the polymorphism using Maryland609, K326, and F1 individuals. In total, 187 pairs of polymorphic primers were selected and used to genotype the parental lines and F1 and F2 individuals. Next, 1381 pairs of primers, consisting of the previously selected polymorphic primers and 1194 SSR primers, were screened for polymorphism using eight relatively genetically distant natural accessions. A total of 143 pairs of polymorphic SSR primers were selected and used to genotype the 222 natural accessions.

2.6. Molecular Data Obtained

Sodium dodecyl sulfate polyacrylamide gel electrophoresis and silver staining were performed as described by Xiang [28]. For individuals in the F2 population, the bands on the 8% non-denaturing polyacrylamide gels were recorded as "A", "B", "H", and "-",

where “A” indicated bands that were the same as in Maryland609, “B” indicated bands that were the same as in K326, and “H” indicated bands that were the same as in F1 individuals. “-” indicated the absence of bands (Figure A1). For the natural accessions, bands were scored as present (1) or absent (0). No amplification was indicated by a “9” at the given position (Figure A2).

2.7. Analysis of Phenotypic Data

Phenotypic data were recorded in an Excel spreadsheet. Basic statistical parameters, and analysis of variance for solanesol content of the P1, P2, F1, F2, and natural populations in 4 environments were performed using SPSS 23.0 (IBM, Corp., Armonk, NY, USA) analysis software.

2.8. Construction of the Genetic Map

The data from the genetic and the natural populations were entered into an Excel spreadsheet and formed into matrices for data analysis, and a genetic map was constructed using Join Map version 4.0 [31] based on the genotyping data from the mapping population.

2.9. QTL Mapping

The genetic map and the solanesol content data for the F2 genetic population were used for QTL mapping using the full QTL model implemented in QTLNetwork2.2 [32] with a step length of 1 cm. $LOD \geq 3$ was defined as an effective locus. The QTLs were named using the format: QTL + linkage group (LG) number + “-” + serial number (if there were multiple QTLs on one LG). The genetic effect of the QTL on the corresponding phenotypic variation was estimated using Markov chain Monte Carlo (MCMC) [33].

2.10. Association Analysis

For association analysis, rare alleles were filtered at the 5% level using DataFormatter2.6.2 [34], and data with values of 0 and 1 were then converted to a data format that met the analysis requirements of the Structure v2.3.4 [35], PowerMarker 3.0 [36], and Tassel 3.0 [37] programs.

The population structure of the 222 accessions was analyzed using STRUCTURE v2.3.4 [35]. The number of populations (K) was set between 1 and 10. Based on an independent allele frequency model, 10 simulation runs were performed after a burn-in period of 100,000 iterations and 100,000 MCMC iterations. The ΔK value was then determined using the method described by Evanno [35]. Based on the $\ln P(D)$ values, the highest value of ΔK was chosen as K minus (the population number).

A total of 588 polymorphic loci were used to perform the association analysis between the SSR markers and the solanesol content data in 4 environments using the general linear method in Tassel 3.0. Polymorphisms with a p value < 0.01 were considered to be significantly associated with the trait [37].

3. Results

3.1. Analysis of Phenotypic Data

The solanesol content of K326, Maryland609, F1, F2 population, and the natural population was determined by UPLC, and the statistical analysis was performed using SPASS 23.0. The basic statistical parameters are presented in Table 1. An F-test showed significant differences in solanesol content between the two parents ($p < 0.01$). The mean solanesol content of the F2 population was 2.292%, with a range of 0.531–3.558%. The differences in the mean solanesol content among the 4 environments were significant ($p < 0.05$), and the range in solanesol content of the 222 accessions was 0.589–4.131%. This indicated that there was a large variation in solanesol content in the F2 population, with a large variation also observed in that of the natural population. The coefficients of kurtosis and skewness were between 0 and 1. The frequency distribution plot of the five populations

(Figure 1) indicated that the distribution of solanesol content fitted a typical quantitative trait inheritance model and was suitable for further analytical study.

Table 1. Basic data characteristics of the solanesol content of the tested materials.

Populations	Environment	Number	Mean ± SD (%)	Range (%)	CV (%)	Kurtosis	Skewness
K326	Hubei	50	2.909 ± 0.201 A	2.868–3.05	10.215		
Maryland609		50	1.126 ± 0.143 C	0.768–1.788	10.484		
F1		50	2.63 ± 0.203 A	2.516–2.823	12.548		
F2 population		206	2.29 ± 0.459 B	0.531–3.558	26.285	0.956	0.386
Natural population	E1	173	1.830 ± 0.438 c	0.695–3.248	33.725	0.293	0.458
	E2	222	2.683 ± 0.447 a	1.535–4.131	26.782	0.404	0.389
	E3	187	1.428 ± 0.402 d	0.589–3.386	34.129	0.303	0.462
	E4	222	2.368 ± 0.426 b	1.690–3.969	28.264	0.412	0.391
	Average	222	2.077 ± 0.428 b	1.127–3.1684	30.725	0.353	0.425

Different letters (A–C) and (a–d) in the same column (mean) indicate extremely significant and significant differences at the level of $p < 0.01$ and $p < 0.05$, respectively.

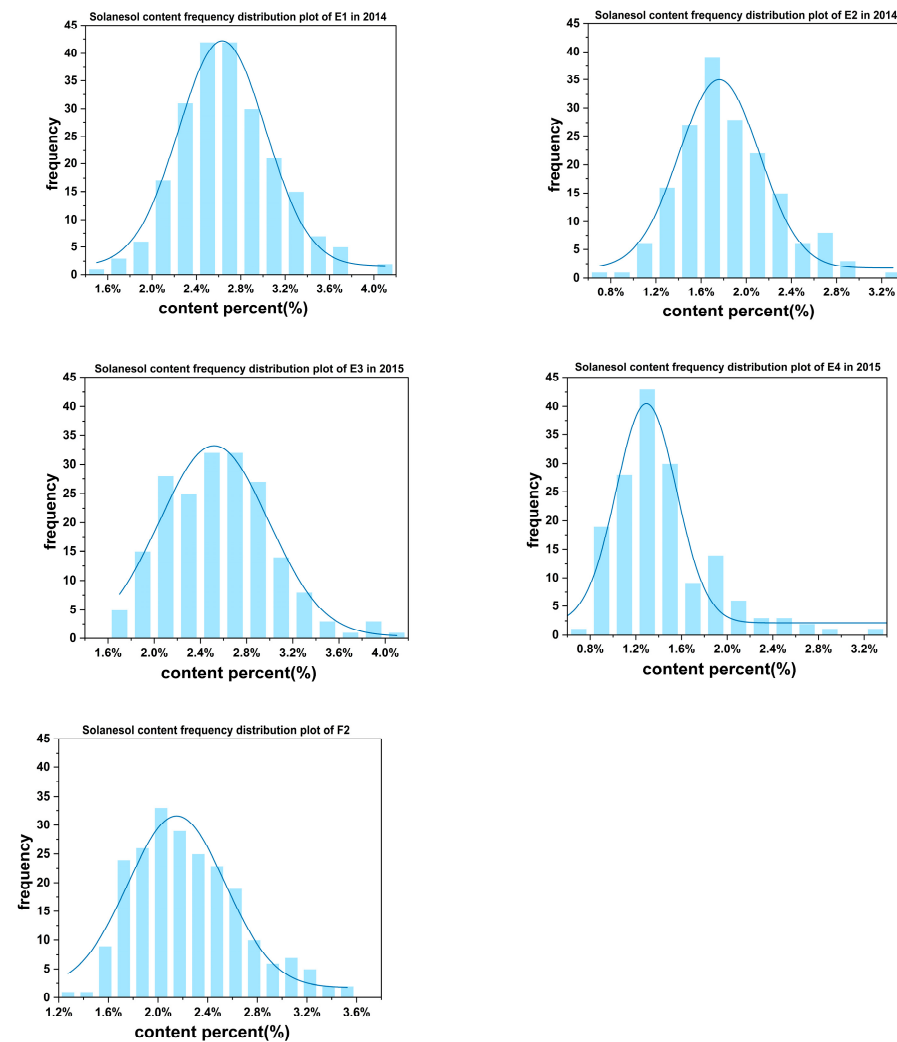


Figure 1. The bar graphs of the frequency distribution of the natural populations of E1, E2, E3, E4, and F2 populations by SPSS 23.0.

3.2. Quantitative Trait Locus Mapping for Solanesol Content

The parental lines Maryland609, K326, and F1 were used for the selection of polymorphic primers. A total of 187 pairs of SSR primers were selected from the 1880 primers screened in this study and were subsequently used to genotype the F2 population. A total of 15 linkage groups consisting of 95 microsatellite markers were then constructed (Figure 2). These linkage groups corresponded to Bindler’s linkage groups LG1, LG3, LG6, LG8, LG10, LG11, LG12, LG13, LG14, LG17, LG20, LG21, LG22, and LG23, and were designated Ch1, Ch3, Ch6, Ch8, Ch10a and Ch10b, Ch11, Ch12, Ch13, Ch14, Ch17, Ch20, Ch21, Ch22, and Ch23, respectively. Ch10a and Ch10b corresponded to two segments of LG10. The total coverage length of the constructed genetic map was 1129 cm. The coverage length of each genetic map ranged from 36.5 to 114.3 cm, and the mean genetic distance between markers was 11.88 cm. The minimum genetic distance between markers was 2.3 cm and the maximum was 32.6 cm. The number of markers in each linkage group ranged from 3 to 12.

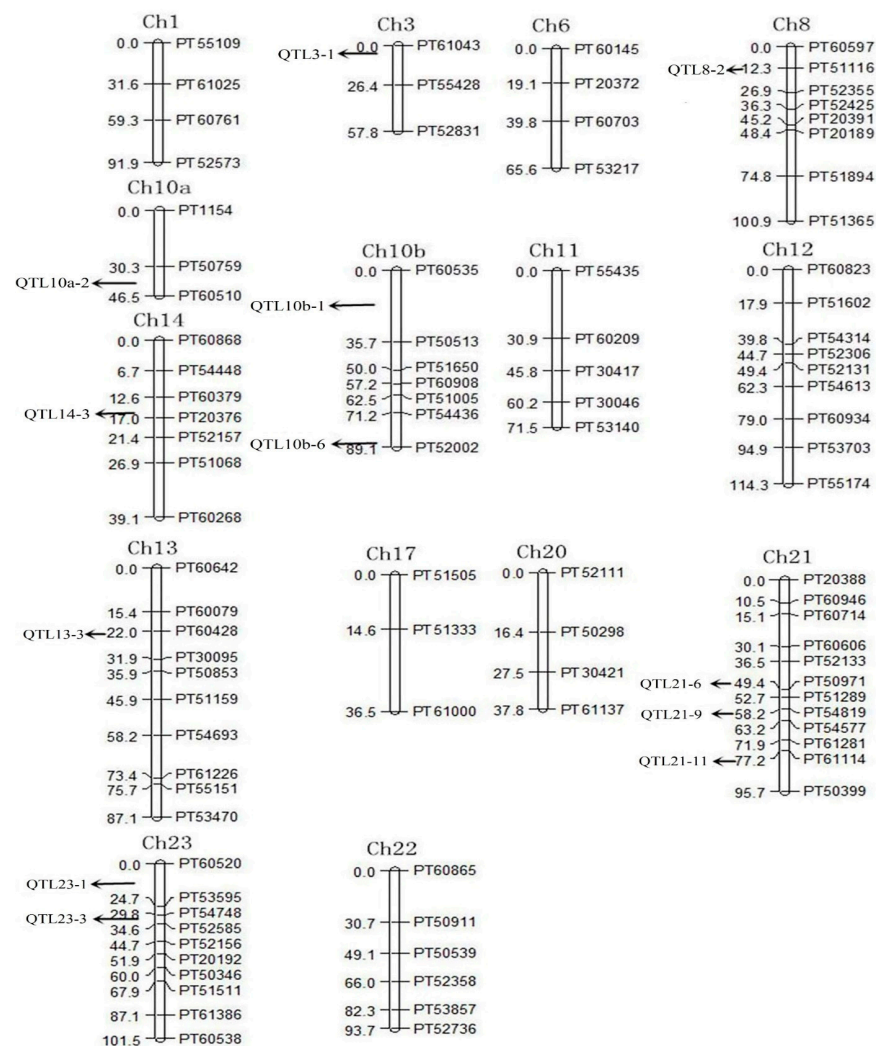


Figure 2. Tobacco solanesol linkage map based on 206 F2 individuals from Maryland609 × k326 cross.

QTL analysis for solanesol content in the tobacco leaves was performed using the full QTL analysis model. A total of 12 QTLs (Figure 2) related to solanesol content were detected and distributed among eight genetic linkage groups. Among the QTLs, QTL8-2, QTL13-3, QTL21-6, QTL21-9, and QTL21-11 showed complete co-segregation with microsatellite markers PT51116, PT60428, PT50971, PT54819, and PT61114, respectively. The genetic distance between the markers associated with the remaining seven QTLs ranged from 0.4

to 24.4 cm. The number of QTLs detected within each linkage group ranged from 1 to 3. Three QTLs associated with solanesol content were detected on linkage group Ch21, linkage groups Ch10b and Ch23 contained two QTLs each, and one QTL each was detected on each of the genetic linkage groups Ch3, Ch8, Ch10a, Ch13, and Ch14.

Among the 12 QTLs, QTL3-1, QTL21-6, and QTL23-3 all had higher phenotypic variation explained, additive, and dominant effects (Table 2). The calculated additive effect values ranged from 0.0685 to 0.1820, while the dominant effect values ranged from 0.0463 to 0.1886. QTL3-1 was located between PT61043 and PT55428 on linkage group LG3, with observed genetic distances between the right and left markers of 2.0 cm and 24.4 cm, respectively. The dominant effect value was -0.1886 , and the genetic effect was associated with the paternal parent, K326. The additive effect value was 0.0685, and the genetic effect was attributed to the maternal parent, Maryland 609. QTL21-6 was located between PT50971 and PT51289 on LG21 and cosegregated with PT50971. The additive and dominant effect values were -0.0866 and -0.2010 , respectively, and the genetic effects were from the paternal parent, K326. QTL3-1 and QTL21-6 had mainly dominant effects and explained 5.19% and 7.59% of the phenotypic variation of solanesol content, respectively. QTL23-3 was located between PT54748 and PT52585 on LG23. The genetic distances to the left and the right markers were 1.0 cm and 3.6 cm, respectively. The additive and dominant effect values were 0.1820 and 0.0463, and the genetic effects were attributed to the material line, Maryland 609. QTL23-3 was mainly an additive effect and explained 10.05% of the phenotypic variation of solanesol content.

Table 2. Estimates of QTL positions, effects, and explained phenotypic variation from the full QTL model.

QTL	Linkage Group	Left Marker	Right Marke	Position (cm)	LOD Value	A [a]	D [b]	PVE [c] (%)
QTL3-1	3	PT61043	PT55428	2.0	4.34	0.0685	-0.1886	5.19
QTL21-6	21	PT50971	PT51289	49.4	5.18	-0.0866	-0.2010	7.59
QTL23-3	23	PT54748	PT52585	30.8	8.6	0.1820	0.0463	10.05

^a Additive effects. ^b Dominant effects. ^c Phenotypic variation explained (%) by the QTL.

3.3. Association Analysis of SSRs and Solanesol Content

Based on 143 pairs of polymorphic primers (Table A2), the population structure of the 222 accessions was analyzed using STRUCTURE v2.3.4. The LnP (D) value increased gradually with increasing K values (hypothetical population number), and there was no inflection point. Therefore, ΔK was used to determine the K value. The largest ΔK value was obtained when $K = 2$ (Figure 3), and so the natural population was divided into two groups. Based on the K-value obtained, the population structure of 222 flue-cured tobacco germplasm was plotted (Figure 4). Group 1 and Group 2 contained 92 and 130 tobacco varieties, respectively. This indicated that the natural population structure was simple and clear. This helped reduce the false positive correlation caused by the complex population structure and to improve the effect of the association analysis. Therefore, the Q matrix with $K = 2$ was used as a covariate for further association analysis.

Association analysis was performed between the markers and the solanesol content of tobacco leaves from 4 different environments using a general linear model in Tassel 3.0. The general linear model threshold was set at $p < 0.01$ for the selection of association markers. A total of 38 significant ($-\log [p \text{ value}] > 3$) marker-trait associations located among 16 linkage groups were detected in at least one environment (Table 3), with the phenotypic variation ranging from 4.22 to 20.07%. Overall, a larger number of associated markers were distributed on LG3, LG10, LG14, LG21, and LG23 compared to the other linkage groups, with LG10 containing the highest number of markers.

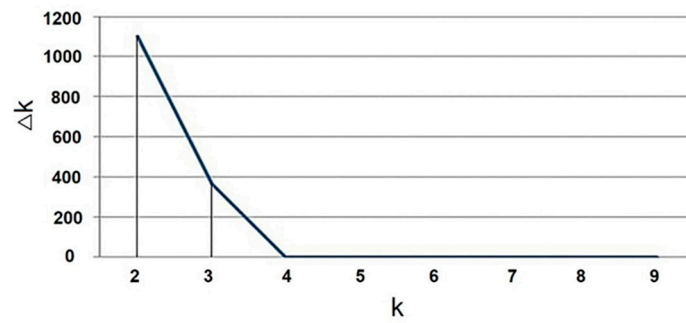


Figure 3. Distribution plot of K and ΔK value by STRUCTURE v2.3.4.

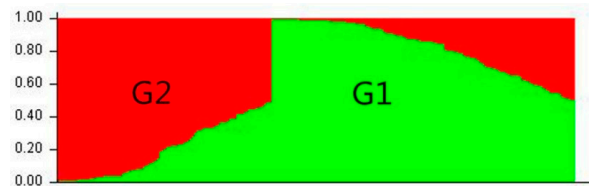


Figure 4. Diagram of population structure of 222 tobacco germplasm resources (K = 2). G1, group 1; G2, group 2 by STRUCTURE v2.3.4.

Table 3. The SSR markers associated with solanesol content and their explained phenotypic variance (%) ($p < 0.001$) based on the GLM procedure.

Marker	LG	Position	E1		E2		E3		E4		Average	
			<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)
PT50457	1	100.72	8.50×10^{-4}	8.31	ns	ns	ns	ns	4.77×10^{-6}	12.47	5.67×10^{-7}	12.69
PT55428	3	114.06	ns	ns	9.43×10^{-5}	7.29	ns	ns	ns	ns	4.67×10^{-5}	10.38
PT55117	3	114.06	ns	ns	1.67×10^{-4}	9.33	ns	ns	ns	ns	ns	ns
PT60524	3	122.13	ns	ns	3.02×10^{-5}	10.30	5.48×10^{-5}	10.38	ns	ns	2.12×10^{-6}	10.56
PT54245	4	38.74	1.48×10^{-4}	8.43	ns	ns	4.55×10^{-4}	7.74	ns	ns	ns	ns
PT61187	5	130.24	8.67×10^{-6}	11.46	ns	ns	1.02×10^{-6}	14.58	ns	ns	1.14×10^{-7}	12.45
PT50923	6	136.38	ns	ns	9.89×10^{-7}	11.47	7.62×10^{-5}	10.20	ns	ns	6.97×10^{-5}	7.73
PT50392	9	48.65	ns	ns	6.52×10^{-5}	8.73	2.04×10^{-4}	10.51	ns	ns	5.45×10^{-6}	10.85
PT60510	10	0.00	3.31×10^{-4}	11.92	ns	ns	8.85×10^{-5}	14.49	1.58×10^{-5}	13.99	3.14×10^{-8}	18.34
PT50759	10	1.65	ns	ns	ns	ns	ns	ns	4.02×10^{-4}	8.63	4.06×10^{-4}	8.62
PT61154	10	4.16	ns	ns	7.88×10^{-5}	7.30	3.98×10^{-4}	7.21	1.01×10^{-4}	7.23	7.68×10^{-8}	13.35
PT61061	10	9.84	3.98×10^{-4}	12.21	ns	ns	ns	ns	7.68×10^{-8}	13.35	3.97×10^{-8}	20.07
PT61339-1	10	51.35	ns	ns	1.81×10^{-4}	6.30	ns	ns	1.28×10^{-6}	11.76	1.54×10^{-8}	13.80
PT51005	10	55.23	ns	ns	ns	ns	ns	ns	2.25×10^{-4}	12.73	1.34×10^{-5}	14.26
PT60114-1	10	57.16	2.63×10^{-4}	13.06	ns	ns	ns	ns	ns	ns	1.46×10^{-6}	12.07
PT60114-2	10	57.16	1.63×10^{-4}	10.10	2.61×10^{-4}	7.59	3.89×10^{-5}	12.58	ns	ns	6.07×10^{-8}	18.42
PT60172	12	89.26	2.26×10^{-4}	9.62	1.33×10^{-4}	8.04	ns	ns	ns	ns	5.15×10^{-6}	10.80
PT60863	14	35.13	ns	ns	3.09×10^{-6}	9.69	4.29×10^{-4}	7.76	ns	ns	2.03×10^{-5}	8.15
PT60868	14	37.34	ns	ns	2.13×10^{-4}	7.57	ns	ns	ns	ns	3.72×10^{-5}	9.05
PT54448	14	42.08	3.03×10^{-4}	7.50	1.96×10^{-4}	6.20	2.39×10^{-6}	13.33	1.51×10^{-4}	7.33	7.59×10^{-8}	12.50
PT52906	15	102.76	ns	ns	2.27×10^{-5}	7.99	ns	ns	1.86×10^{-5}	9.31	3.00×10^{-6}	9.63
PT54811	16	130.98	ns	ns	2.33×10^{-5}	7.87	ns	ns	ns	ns	4.27×10^{-5}	7.38
PT61633	17	75.26	ns	ns	ns	ns	ns	ns	6.83×10^{-4}	6.54	ns	ns
PT52838	21	24.05	ns	ns	0.0231	4.55	ns	ns	ns	ns	0.04700421	4.86
PT61192	21	45.54	0.0020	5.92	ns	ns	ns	ns	ns	ns	0.007355	4.22
PT51054	21	49.70	ns	ns	3.53×10^{-8}	13.33	ns	ns	5.41×10^{-4}	6.25	3.88×10^{-9}	15.06
PT52536	21	49.70	ns	ns	1.48×10^{-4}	6.83	ns	ns	ns	ns	1.96×10^{-5}	8.57
PT61114	21	49.70	ns	ns	7.52×10^{-4}	7.81	ns	ns	ns	ns	4.98×10^{-5}	10.30
PT55472	21	49.70	ns	ns	ns	ns	ns	ns	ns	ns	5.09×10^{-4}	10.67
PT30355	21	52.43	ns	ns	5.50×10^{-4}	12.50	ns	ns	ns	ns	ns	ns
PT20388	21	72.14	ns	ns	1.45×10^{-5}	11.56	ns	ns	5.84×10^{-7}	16.39	2.61×10^{-8}	17.03
PT52760	22	30.24	ns	ns	ns	ns	7.13×10^{-5}	14.60	ns	ns	2.97×10^{-4}	9.32

Table 3. Cont.

Marker	LG	Position	E1		E2		E3		E4		Average	
			<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)	<i>p</i>	R ² (%)
PT60494	22	47.55	ns	ns	2.43 × 10 ⁻⁴	7.23	ns	ns	3.57 × 10 ⁻⁴	5.67	0.03087387	7.52
PT52736	22	68.26	ns	ns	1.31 × 10 ⁻⁴	8.05	ns	ns	ns	ns	1.07 × 10 ⁻⁴	8.22
PT54707	23	23.21	ns	ns	ns	ns	1.82 × 10 ⁻⁴	10.98	ns	ns	ns	ns
PT60520	23	63.69	ns	ns	3.77 × 10 ⁻⁵	8.40	ns	ns	ns	ns	6.16 × 10 ⁻⁵	7.96
PT61584	23	61.71	ns	ns	4.27 × 10 ⁻⁶	9.38	ns	ns	ns	ns	3.11 × 10 ⁻⁵	7.77
PT51170	24	49.25	ns	ns	ns	ns	5.86 × 10 ⁻⁴	7.60	3.03 × 10 ⁻⁵	9.05	2.99 × 10 ⁻⁶	9.94

The *p*-value indicates the significance between the marker and solanesol content; the R² value indicates the percentage of phenotypic variation explained by the marker; ns indicates no significance; average: the mean value of E1, E2, E3, and E4.

Except for PT55117, PT54245, PT61633, PT30355, and PT54707, the other markers were detected by association analysis of markers and means. A total of 15 markers were detected in two environments. Some of them, including PT61339-1, PT52906, PT51054, PT20388, and PT60494, were detected only at two sites in the northern (Shandong Province) tobacco-growing region of China, while PT54245 and PT61187 were detected only at two sites in the southern (Sichuan Province) tobacco growing region of China. PT60510, PT60114-2, and PT11154 were detected in three environments and accounted for 7.30–18.34% of the phenotypic variation. PT54448 was detected in four environments and explained 6.20–13.33% of the phenotypic variation. Overall, PT61061 explained the highest amount of phenotypic variation, with 20.07%. It was located on LG10 and was also detected by association analysis of SSR markers and means.

3.4. The Confirmed Markers Related to Solanesol Content by Two Methods

The 143 pairs and 187 pairs of SSR primers used in the association analysis and QTL mapping, respectively, were from the SSR marker data published by Bindler and Tong [10–12]. Therefore, the results were comparable. Table 4 shows the 14 significant marker-trait associations located near nine QTLs and distributed on LG3, LG10a, LG10b, LG14, LG21, and LG23. Markers PT55428, PT50759 and PT60510, PT61114, and PT60520 and PT53595, which were significantly associated with solanesol content, were detected in the natural population as the right marker of QTL3-1, the left and right markers of QTL10a-2, the left marker of QTL21-11, and the left and right markers of QTL23-1, respectively. Marker PT55117, detected in the natural population, and PT55428, the right marker of QTL3-1, are located at the same site on the linkage map constructed by Bindler et al. [10,11]. Markers PT51054, PT52536, PT61114, and PT55472, the right marker of QTL21-9 (PT51289), and the left marker of QTL21-11 (PT61114) were also located at the same site. Many of the left and right QTL markers were repeatedly detected at the same location using association analysis, confirming the reliability of the results.

Table 4. Location and minimum genetic distance from left or right QTL marker, which was determined by association analysis.

QTL	Linkage Group	Interval	Marker	Location [a]	D [b] (cm)	Molecular Marker Localization
QTL3-1	3	PT61043-PT55428	PT55428	114.06	0.00	138,644,270–138,644,484
QTL10a-2	10a	PT50759-PT60510	PT60510	0.00	0.00	3,640,102–3,640,139
QTL10b-6	10b	PT54436-PT52002	PT60114-1	57.16	2.479	490,772–490,852
			PT60114-2	57.16	2.479	73,974,226–73,974,332
QTL14	14	PT60379-PT20376	PT60868	35.13	−2.48	59,135,301–59,135,403
			PT60863	37.34	3.651	65,675,209–65,675,347

Table 4. Cont.

QTL	Linkage Group	Interval	Marker	Location ^[a]	D ^[b] (cm)	Molecular Marker Localization
QTL21-6	21	PT50971-PT51289	PT51054	49.70	0.00	51,260,251–51,260,374
			PT52536	49.70	0.00	52,765,747–52,765,930
			PT55472	49.70	0.00	54,904,073–54,904,263
QTL21-9	21	PT54577-PT61281	PT54707	23.21	1.49	628,952–629,173
QTL21-11	21	PT61114-PT50399	PT61114	49.70	0.00	108,401,173–108,401,338
			PT20388	49.70	0.00	927,449,98–92,745,182
QTL23-1	23	PT60520-PT53595	PT60520	63.69	0.00	40,922,488–40,922,687
QTL23-3	23	PT54748-PT52585	PT61584	61.71	−0.55	330,872 – 331,056

^a Location of markers on the linkage group constructed by Bindler [11]. ^b Minimum genetic distance of identified associated markers from the left or right QTL marker. The information was obtained in the confirmed interval regions from the Tobacco Genome Database (https://solgenomics.net/organism/Nicotiana_tabacum/genome, accessed on 21 May 2024).

4. Discussion

4.1. Solanesol of Tobacco Leaves

Solanesol is an important natural product due to its anti-cancer, anti-ulcer, anti-aging, neurodegenerative, and immune-enhancing properties, as well as its effect on the quality of flue-cured tobacco leaves [5–7,38]. At present, most of the solanesol used in pharmaceutical applications is derived from tobacco leaves; however, the low solanesol content of the raw materials limits its utilization [3,4,8,9]. Therefore, increasing the solanesol content of tobacco leaves is beneficial for exploiting the medicinal value and discovering new applications of solanesol. To achieve this goal, it is necessary to investigate the genetic basis of solanesol accumulation in tobacco leaves. Previous research has shown that solanesol content is a complex quantitative trait that is significantly influenced by genotype and environment. For example, there are significant differences in solanesol content between tobacco varieties. In this study, solanesol content ranged from 0.589% to 4.131%, with breeding variety 7514 having the highest solanesol concentrations. These phenotypic data provided a solid basis for investigating the genetic architecture of solanesol content.

4.2. QTL Mapping and Association Analysis for Tobacco Traits

QTL mapping is an effective method for studying complex quantitative traits, and the construction of a linkage map is the basis of QTL analysis. Linkage maps based on molecular markers have been widely used in many crops, including maize, rice, wheat, and cotton [39–42]. In tobacco, many linkage maps have been constructed using molecular markers. For example, Lin et al. constructed the first molecular linkage map of tobacco by genotyping 99 F2 individuals derived from a cross between *Nicotiana plumbaginifolia* and *Nicotiana longiflora* using restriction fragment length polymorphism and random amplified polymorphic DNA analyses [43]. The first linkage map included 19 linkage groups and spanned a total genetic distance of 1385.6 cm. In 2007, Bindler et al. constructed a linkage map with 24 linkage groups and a total genetic distance of 1920 cm [11]. Then, in 2011, Bindler et al. developed 5119 pairs of SSR primers based on sequence data from the US Tobacco Genome Sequencing Project, further increasing the density of the linkage map [11]. The high-density linkage map included 24 linkage groups consisting of 2036 pairs of SSR primers and covered 3270 cm. Based on this early work, Tong et al. subsequently developed 4886 pairs of SSR primers to study genetic traits in tobacco [12]. In the current study, we constructed a genetic map with 15 linkage groups using SSR molecular markers. Our primer sequence information was obtained from the studies of Bindler et al. and Tong et al. The molecular markers in the linkage map constructed in the current study were located on the same LGs as those of Bindler et al. In addition, Ch6, Ch10a, Ch10b, Ch17, Ch20, and Ch22 had the same marker order as that reported by Bindler et al. However, the number

of LGs was <24 and had low coverage, which could be caused by the limited genetic base of the bi-parental. A larger population size, representing a broader genetic base, would be helpful for further dissection of the solanesol content in tobacco leaves. In addition, QTL3-1, QTL21-6, and QTL23-3 explained 22.83% of the phenotypic variation in solanesol content. Overall, the phenotypic variation among individual QTLs ranged from 5.19% to 10.05%, which was greater than that of the chemical constituents of tobacco leaves [21].

Compared to linkage mapping, genome-wide association analysis has several advantages for dissecting the genetic basis of quantitative traits. First, association analysis uses natural populations (diverse germplasm) as experimental material in which recombination events have occurred during evolution. Therefore, association analysis can identify a larger number of alleles than those from the two parents. Secondly, there is no need to construct a genetic map, which saves time and reveals environmental effects. The number and resolution of QTLs are mainly determined by genetic diversity, population structure, and linkage disequilibrium, among other factors. To date, association analysis has been used to identify molecular markers associated with agronomic traits, disease resistance, nicotine content, tobacco-specific nitrosamine content, aroma components, and chemical constituents in tobacco. The results have shown that the phenotypic variation associated with agronomic traits and disease resistance is generally greater than that associated with chemical and aroma components [44,45]. With regard to secondary metabolites, previous research has mainly focused on the expression patterns and functions of genes encoding for key biosynthetic enzymes, with very few studies focusing on molecular markers associated with secondary metabolite production. Vontimitta et al. identified several microsatellite markers that co-segregated with *Ab1* and *BMVSE*, which affect the accumulation of cis-abienol and sucrose esters, respectively, in tobacco. The markers have improved our understanding of these two leaf surface components and allowed marker-assisted selection in tobacco [13]. In the current study, 222 core collections with high genetic variation were selected as experimental material. A total of 38 markers were found to be significantly associated ($p < 0.01$) with solanesol content in at least one environment. The 38 markers were distributed among 16 LGs and explained 5.64–20.08% of the phenotypic variation, respectively. The phenotypic variation explained by LG14-PT54448, LG10-PT60114-1, LG10-PT11154, LG10-PT60510, LG10-PT61061, and LG21-PT51054 was relatively high and could be detected stably under different environmental conditions. Therefore, the primers selected in the current study will be useful for future molecular-assisted selection of solanesol content in tobacco. It is difficult to detect polymorphic alleles beyond those inherited from the parental strains using linkage mapping. In addition, association analysis could provide opportunities to identify additional markers associated with a trait due to the increased allelic variation in natural populations resulting from the large number of accessions [46]. However, linkage analysis can detect the additive/dominant effect of QTL and overcome the drawbacks of low efficiency of association analysis for rare alleles detected [47–49]. Thus, the combination of linkage mapping and association analysis can significantly improve the reliability of the located QTL and associated allele. This combination of analysis methods has been widely used for QTL mapping of quantitative traits in crops [50]. In the present study, we identified 12 QTLs for solanesol content located on 8 LGs and 38 significant marker-trait associations located on 16 LGs. The 9 QTLs distributed on LG3, LG10a, LG10b, LG14, LG21, and LG23 were confirmed by association analysis. Through the Tobacco Genome Database (https://solgenomics.net/organism/Nicotiana_tabacum/genome, accessed on 21 May 2024), gene function annotation information was obtained in the confirmed interval regions. Based on the results linked/associated with the solanesol content, many genes were screened in the marker intervals corresponding to the above nine QTLs. Phosphotransferase, tyrosine ferulic transferase, acetyltransferase, and xylosyltransferase were linked to the markers of LG23. Of particular interest were the linked transcription factors *bHLH*, *MYB*, *ERF*, *WRKY*, and *bZIP* distributed on LG3, LG10a, LG10b, LG14, and LG21. Previous studies of plant transcription factors have mainly focused on their developmental and physiological regulatory functions. More recently, the function of regulating secondary

metabolites has been a concern. Here, the present results related to solanesol content could be used as a reference for further studies to improve the accumulation of terpenoids in plants by applying metabolic engineering.

4.3. Analysis of Plant Metabolites

With the development of molecular markers, QTL and association analysis of metabolites have advanced significantly over the past 10 years. For example, Keurentjes et al. investigated metabolites in Arabidopsis leaves using non-targeted liquid chromatography quadrupole time-of-flight mass spectrometry analysis of 160 inbred lines and revealed the genetic pathways involved in aliphatic thioglycoside synthesis based on PCR-based molecular markers [50]. Matsuda et al. determined the primary and secondary metabolites in rice seeds and constructed a metabolic library containing 759 metabolic signals [51]. Glycosyltransferase genes encoding components of the flavone synthesis pathways were also analyzed by linkage analysis [50,52]. To date, there has been a lack of studies on the genetic basis of tobacco metabolites, especially secondary metabolites. The lack of studies may be a consequence of the fact that tobacco has a large, hetero tetraploid genome, narrow genetic background, and low polymorphism, all of which contribute to the high cost of determining the metabolite levels. However, this information gap has severely limited metabolite genetic research and molecular-assisted breeding in tobacco.

5. Conclusions

In the present study, the identified markers were mainly located on LGs 10 and 21, which are referred to as “hot regions”. The markers PT61061, PT54448, PT60510, PT60114-2, and PT20388, which have a high phenotypic variation explained, could be used to select lines with high solanesol content by marker-assisted selection at any stage of tobacco growth. This would dramatically reduce costs and facilitate studies on the genetic basis of solanesol content in tobacco. The identified hot regions contain genes encoding for the major regulatory factors controlling solanesol accumulation in tobacco. Thus, the information obtained in this study will contribute to further fine mapping of genes regulating solanesol content and provide a new avenue for investigating the genetic basis of secondary metabolites in tobacco. In addition, by analyzing the genetic interactions of metabolites, we can further investigate specific metabolic pathways or reconstruct a metabolic network model in the future.

Author Contributions: Conceptualization, Q.F. and N.Y.; Methodology, Y.D., Z.Z. and H.Z.; Validation, F.J. and C.Q.; Formal analysis, J.L., D.X. and L.C.; Investigation, Y.D., H.Z., L.C., N.Y., F.J., Y.L. (Yunkang Lei) and Y.L. (Yanhua Liu); Data curation, Q.F.; Writing—original draft, J.L. and D.X.; Writing—review & editing, Z.Z. and Y.L. (Yanhua Liu); Visualization, Y.L. (Yunkang Lei); Supervision, J.W.; Project administration, J.W.; Funding acquisition, Y.L. (Yanhua Liu). All authors have read and agreed to the published version of the manuscript.

Funding: We are grateful to the Foundation of Agricultural Science and Technology Innovation Programme of Chinese Academy of Agricultural Sciences (ASTIP-TRIC05) for financial support.

Data Availability Statement: The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.12516892>.

Acknowledgments: We thank Min Ren for kindly providing 222 varieties from the National Infrastructure for Crop Germplasm Resource (Tobacco; Qingdao, China) of the Chinese Academy of Agricultural Sciences (CAAS).

Conflicts of Interest: Dehu Xiang was employed by the company Hunan Tobacco Company, and Yunkang Lei was employed by the Sichuan Tobacco Corporation. The remaining 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.

Appendix A

Table A1. List of 222 accessions in natural population tested.

No.	Name	No.	Name	No.	Name	No.	Name
1	NC82	57	86-3002	113	Speight G-41	169	Jintai66-4
2	V2	58	82-3041	114	Virginia 182	170	Wenganpipayae
3	Xiaohuangjin1025	59	7514	115	P3	171	G80B
4	Speight G-28	60	C151	116	Qiannan7	172	Wenganzhongpingkaoyan
5	Dabajin599	61	7411	117	T.I.706	173	Fuquangaojiaohuang
6	Zhongyan90	62	B22	118	Yunduo1	174	Jintai88
7	Zhongyan15	63	CV58	119	Xilouyan	175	Chunlei1
8	9201	64	K149	120	Dabajin2522	176	Bijin1
9	NC567	65	NC 2326	121	Chunlei4	177	Liaoyan11
10	K346	66	Yongding1	122	Changbaziyang	178	Wanye
11	Baihua205	67	Danyu2	123	Baijinhuangmiao	179	JB-200
12	Coker 176	68	Guanghuang54	124	Wuming2	180	Dixie Bright 101
13	Yunyan85	69	Jinxing6007	125	Red Hort	181	Mudan79-2
14	Cuib1	70	Panyuanhuang	126	H68E-1	182	H80A432
15	CF80	71	Puyou1	127	Vesta 30	183	TI1112
16	NC89	72	Yunyan2	128	Zhubo1	184	CNH-NO.7
17	K326	73	CF90NF	129	Kuiyan2487	185	CU263
18	CF965	74	Changbohuang	130	Liaoyan9	186	K358
19	Va116	75	Tailifu1060	131	Anqiumanwuxiang	187	MRS-2
20	Yunyan87	76	Lushanxiaoliuye	132	Changmaohuang	188	NC1108
21	NC55	77	Gexin5	133	Majiangliyan	189	NCTG60
22	RG13	78	Special 401	134	Oxford 4	190	NCTG61
23	CV088	79	Qinyan95	135	Damiaohuang	191	OX2028
24	CV87	80	Qinyan96	136	Gedajinyan	192	OX2101
25	FC8	81	Longjiang851	137	Wengantieganyan	193	RG3414
26	Zhongyan14	82	Longjiang925	138	Hicks 55	194	SPG-169
27	RG8	83	Longjiang935	139	8813	195	SPG-172
28	RG89	84	Yuyan3	140	Daliutiao	196	TI1597
29	Zhongyan102	85	Special 400	141	Anxuan4	197	Va80
30	9111-21	86	Cash	142	Taoliuzi	198	Va411
31	T.T.9	87	Qiaozhuangduoye	143	Wangengzi	199	Chunlei3
32	T.T.11	88	Black Shank Resistant	144	Pingbanliuye	200	Damo
33	Honghuadajinyuan	89	Harrison Special	145	Luodihuang	201	TI 448A
34	Speight G-80	90	Longyan1	146	ETWM 10	202	Fandi3-bing
35	RG11	91	Kutsaga E1	147	NC71	203	84-3117
36	RG17	92	Virginia Gold	148	Criollo c-1-1	204	Enshu
37	Yanyan97	93	SH.86-1	149	Xiaohuangjin0138	205	K730
38	09-53	94	NC-22-NF	150	Y-2	206	OX2007
39	Gexin3	95	TL 106	151	TI1500	207	SPG-168
40	Jingyehuang	96	78-3012	152	Tailifu1011	208	Changgeliuye
41	T.T.8	97	Baisezhong	153	8022	209	Dashuba(Straight)
42	Zhongyan86	98	I-35	154	Changboyan	210	Dashuba2106
43	Zhongyan103	99	Tailifu1061	155	Heimiaoshuba2104	211	Yuyeshuba2109
44	Kang88	100	Longshe	156	77089-12	212	Kaiyangtuanyuye
45	CV91	101	Heiyeyan	157	Va458	213	Huangpingmaoganyan

Table A1. Cont.

No.	Name	No.	Name	No.	Name	No.	Name
46	Speight G-140	102	Yuanyeyan	158	Coker187-Hicks	214	Fuquanzheyuan
47	Beinhart 1000-1	103	Daqingjin	159	Jiyan5	215	Fuquandapipa
48	T.I.245	104	Xiaojianshao	160	Coker 206	216	Huangpingdaliuye
49	Kang66	105	Liuyejian2017	161	Manguangliuyejian	217	Fuquandajiwei
50	Tiebaziyuan	106	Dashuba2101	162	CT709	218	Fuquanchaotianli
51	Manwuxiang	107	Huangmiaoyu2235	163	Xiaohuangyan	219	Lushandawojuye
52	Hicks(Broad Leaf)	108	Fuquanhoujieba	164	Renshenyan	220	K394
53	NC 95	109	Jintai49	165	Coker9	221	Xiaoyehuang
54	Coker 319	110	Pelo De Oro P-1-6	166	Boheyuan	222	Hicks
55	Coker 139	111	Jiulouyan	167	Hycu Ruce		
56	By 4	112	Ky 151	168	Wajiaoyuan		



Figure A1. Amplification results of SSR-PT54023 in K326 Maryland609, F1, and part of F2 population.

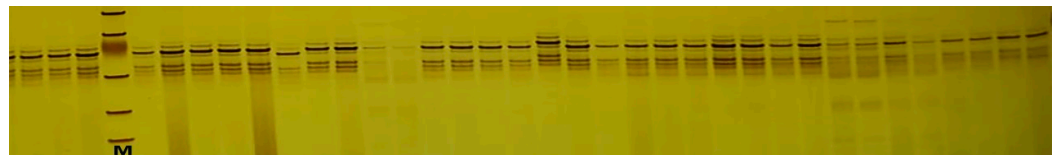


Figure A2. Amplification results of SSR-PT50136 in representative flue-cured tobacco germplasm resources.

Table A2. Genetic diversity for 143 pairs of polymorphic SSR primers.

Marker	Na [a]	PIC [b]	Marker	Na	PIC	Marker	Na	PIC
PT54245	2	0.323	PT53936	3	0.473	PT53595	3	0.558
PT51682	4	0.660	PT53089	2	0.308	PT50136	2	0.355
PT61187	2	0.232	PT51206	3	0.304	PT50727	5	0.546
PT61210	2	0.375	PT60435	3	0.466	PT50472	6	0.725
PT60345	3	0.485	PT50670	2	0.370	PT55266	2	0.310
PT50392	3	0.439	PT51144	6	0.688	PT61428	4	0.547
PT61010	3	0.474	PT53796	2	0.354	PT54887-1	2	0.254
PT52958	2	0.260	PT53362	3	0.359	PT54887-2	2	0.344
PT60369	3	0.345	PT61564	6	0.757	PT54887-3	3	0.484
PT55030	4	0.598	PT61061	4	0.521	PT55296	3	0.272
PT61396	2	0.277	PT51145	3	0.402	PT53026	3	0.436
PT60172	2	0.369	PT60080	6	0.520	PT53915	2	0.245
PT61499	3	0.412	PT60606	8	0.772	PT61339-1	2	0.358
PT60863	2	0.370	PT30355	8	0.788	PT61339-2	3	0.475
PT60868	4	0.505	PT20388	4	0.299	PT60114-1	2	0.284

Table A2. Cont.

Marker	Na ^[a]	PIC ^[b]	Marker	Na	PIC	Marker	Na	PIC
PT53829	2	0.230	PT54644	4	0.412	PT60114-2	5	0.587
PT52906	2	0.366	PT50077	4	0.491	PT30311	3	0.471
PT54448	2	0.348	PT51415	3	0.408	PT60177-1	3	0.514
PT51411	3	0.382	PT55319	5	0.676	PT60177-2	2	0.269
PT54811	2	0.347	PT51085	3	0.554	PT30380	4	0.479
PT53384	3	0.382	PT54778	5	0.711	PT50245	4	0.561
PT61286	3	0.511	PT60427	4	0.648	PT61386	3	0.279
PT60494	2	0.317	PT52838	4	0.623	PT60520	3	0.557
PT50500	2	0.195	PT54819	3	0.589	PT55117	3	0.379
PT60271	3	0.403	PT51951	5	0.693	PT20445	5	0.561
PT51976	2	0.254	PT61192	2	0.151	PT52002	2	0.364
PT61584	2	0.375	PT51054	2	0.370	PT54336	3	0.374
PT54342	2	0.141	PT51289	3	0.412	PT61043	4	0.501
PT52736	3	0.546	PT55472	3	0.554	PT50513	3	0.446
PT52353	3	0.448	PT61114	3	0.460	PT60908	3	0.352
PT52585	2	0.374	PT61319	4	0.512	PT51005	3	0.346
PT54629	2	0.159	PT50352	2	0.333	PT1154	2	0.368
PT54759	2	0.270	PT60946	4	0.588	PT54711	2	0.133
PT54707	3	0.348	PT55428	4	0.681	PT61612	2	0.132
PT51152	4	0.488	PT53466	2	0.364	PT50923	2	0.363
PT50457	4	0.430	PT60524	2	0.328	PT52760	3	0.301
PT60404	3	0.205	PT60510	3	0.471	PT52418	2	0.323
PT61367	2	0.330	PT50759	4	0.555	PT60581	3	0.361
PT60257	2	0.201	PT54061	2	0.305	PT53223	3	0.314
PT60146	2	0.151	PT50062	4	0.513	PT61160	3	0.431
PT60886	5	0.561	PT55453	2	0.271	PT50736	2	0.326
PT60123	5	0.648	PT52718	3	0.480	PT51170	2	0.198
PT52804	6	0.521	PT60861	3	0.547	PT52318	2	0.364
PT52536	2	0.373	PT52509	3	0.463	PT53847	3	0.447
PT51130	10	0.586	PT50346-1	5	0.736	PT50501	3	0.366
PT50434	6	0.616	PT50346-2	2	0.368	PT61633	2	0.365
PT55162	2	0.266	PT51065	3	0.479	PT50488	2	0.191
PT52133	4	0.691	PT52156	2	0.369	Mean	3.133	0.422

^a Number of alleles. ^b Polymorphism information content.

References

- Feng, W.W. Decryption tobacco. *Cenchr Front.* **2014**, *11*, 16–19.
- Jiang, Y.E. *Chinese Tobacco Germplasm Resources*; China Agriculture Press: Beijing, China, 1997.
- Bai, Q.; Yu, J.; Su, M.; Bai, R.; Katumata, G.; Katumata, M.; Chen, X. Antioxidant function of solanesol and its inhibitory effect on tyrosinase. *J. Biomed. Eng.* **2014**, *31*, 833–836. [[CrossRef](#)]
- Taylor, M.A.; Fraser, P.D. Solanesol: Added value from Solanaceous waste. *Phytochemistry* **2011**, *72*, 1323–1327. [[CrossRef](#)] [[PubMed](#)]
- Banožić, M.; Babić, J.; Jokić, S. Recent advances in extraction of bioactive compounds from tobacco industrial waste—A review. *Ind. Crop. Prod.* **2020**, *14*, 112009. [[CrossRef](#)]
- Nidhi, S.; Shubham, U.; Ambika, S.; Rakesh, S.; Nshuman, S.; Bidisha, R.; Sidharth, M. Neuroprotection by solanesol against ethidium bromide-induced multiple sclerosis-like neurobehavioral, molecular, and neurochemical alterations in experimental rats. *Phytomed. Plus* **2021**, *1*, 4. [[CrossRef](#)]
- Withers, S.T.; Keasling, J.D. Biosynthesis and engineering of isoprenoid small molecules. *Appl. Microbiol. Bio.* **2007**, *73*, 980–990. [[CrossRef](#)]
- Lin, F.R.; Zhou, Y.S.; Chen, X.Z. Characterization of solanesol. *Jiangsu Chem. Ind.* **2008**, *36*, 27–29. [[CrossRef](#)]

9. Saygili, I.; Kandemir, N.; Kinay, A.; Aytac, S.; Ayan, A.K. SSR marker-based genetic characterization of Turkish oriental tobaccos. *Mol. Biol. Rep.* **2022**, *49*, 11351–11358. [[CrossRef](#)] [[PubMed](#)]
10. Bindler, G.; Plieske, J.; Bakaher, N.; Gunduz, I.; Ivanov, N.; Van der Hoeven, R.; Ganal, M.; Donini, P. A high density genetic Map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. *Theor. Appl. Genet.* **2011**, *123*, 219–230. [[CrossRef](#)]
11. Bindler, G.; van der Hoeven, R.; Gunduz, I.; Plieske, J.; Ganal, M.; Rossi, L.; Gadani, F.; Donini, P. A microsatellite marker based linkage map of tobacco. *Theor. Appl. Genet.* **2007**, *114*, 341–349. [[CrossRef](#)]
12. Tong, Z.J.; Jiao, F.C.; Wu, X.F.; Wang, F.Q.; Chen, X.J.; Li, X.Y. Mapping of quantitative trait loci underlying six agronomic traits in flue-cured tobacco (*Nicotiana tabacum* L.). *Acta Agron. Sin.* **2012**, *38*, 1407–1415. [[CrossRef](#)]
13. Vontimitta, V.; Danehower, D.A.; Steede, T.; Moon, H.S.; Lewis, R.S. Analysis of a *Nicotiana tabacum* L. Genomic region controlling two leaf surface chemistry traits. *J. Agric. Food Chem.* **2010**, *58*, 294–300. [[CrossRef](#)] [[PubMed](#)]
14. Cheng, L.R.; Yang, A.G.; Jiang, C.H.; Ren, M.; Zhang, Y.; Feng, Q.F. Quantitative Trait Loci Mapping for Plant Height in Tobacco using Linkage and Association Mapping Methods. *Crop Sci.* **2015**, *55*, 641–647. [[CrossRef](#)]
15. Julio, E.; Denoyes-Rothan, B.; Verrier, J.L.; Dorlhac de Borne, F. Detection of QTLs linked to leaf and smoke properties in *Nicotiana tabacum* based on a study of 114 recombinant inbred lines. *Mol. Breed.* **2006**, *18*, 69–91. [[CrossRef](#)]
16. Yang, H.J.; Geng, X.Q.; Zhao, S.M.; Shi, H.Z. Genomic diversity analysis and identification of novel SSR markers in four tobacco varieties by high-throughput resequencing. *Plant Physiol. Biochem.* **2020**, *150*, 80–89. [[CrossRef](#)] [[PubMed](#)]
17. Li, H.L.; Chen, M.X.; Zhou, D.X.; Chen, S.H.; Tao, A.F.; Li, Y.K. QTL analysis of six important traits in tobacco (*Nicotiana tabacum* L.). *Acta Agron. Sin.* **2011**, *37*, 1577–1584. [[CrossRef](#)]
18. Gao, T.T.; Jiang, C.H.; Luo, C.G.; Yang, A.G.; Cheng, L.R.; Dai, S.S. Mapping of quantitative trait loci affecting resistance to brown spot in tobacco line Beinhart1000-1. *Acta Tabacaria Sin.* **2014**, *20*, 104–107. [[CrossRef](#)]
19. Bai, D.; Reeleder, R.; Brandle, J.E. Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. *Theor. Appl. Genet.* **1995**, *91*, 1184–1189. [[CrossRef](#)] [[PubMed](#)]
20. Vontimitta, V.; Lewis, R.S. Mapping of quantitative trait loci affecting resistance to *Phytophthora nicotianae* in tobacco (*Nicotiana tabacum* L.) line Beinhart-1000. *Mol. Breed.* **2012**, *29*, 89–98. [[CrossRef](#)]
21. Xiao, B.G.; Lu, X.P.; Jiao, F.C.; Li, Y.P.; Sun, Y.H.; Guo, Z.K. Preliminary QTL analysis of several chemical components in flue cured tobacco (*Nicotiana tabacum* L.). *Acta Agron. Sin.* **2008**, *34*, 1762–1769. [[CrossRef](#)]
22. Zhang, J.S.; Wang, R.G.; Yang, C.Y.; Wu, C.; Shi, Y.W.; Wang, Z.H.; Wang, Y.; Ren, X.L. Genetic diversity of agronomic traits and association analysis with SRAP markers in flue-cured tobacco (*Nicotiana tabacum*) varieties from China and abroad. *Acta Agron. Sin.* **2012**, *38*, 1029–1041. [[CrossRef](#)]
23. Dadras, A.R.; Sabouri, H.; Nejad, G.M.; Sabouri, A.; Shoai-Deylami, M. Association analysis, genetic diversity and structure analysis of tobacco based on AFLP markers. *Mol. Biol. Rep.* **2014**, *41*, 3317–3329. [[CrossRef](#)] [[PubMed](#)]
24. Xie, X.D.; Cao, P.J.; Wang, Z.; Gao, J.P.; Wu, M.Z.; Li, X.X.; Zhang, J.F.; Wang, Y.F.; Gong, D.P.; Yang, J. Genome-wide characterization and expression profiling of the PDR gene family in tobacco (*Nicotiana tabacum*). *Gene* **2021**, *788*, 145637. [[CrossRef](#)]
25. Wu, C.; Xia, Y.S.; Li, R.H.; Lv, Y.H.; Yu, Y.W.; Zhao, W.C.; Qiu, M.W.; Guo, P.G. Association analysis of tobacco bacterial wilt resistance with molecular markers. *Tob. Sci. Technol.* **2015**, *48*, 1–12. [[CrossRef](#)]
26. Ge, X.L.; Liu, Y.H.; Yao, Z.M.; Du, Y.M.; Yan, N.; Zhang, H.B.; Dai, P.G. Study on the correlation between solanesol accumulation and expression of gene encoding terpenoid synthetic enzymes in tobacco. *Chin. Tob. Sci.* **2017**, *38*, 8–14. [[CrossRef](#)]
27. Xiang, D.H.; Zhao, T.Z.; Du, Y.M.; Zhang, Z.F.; Yang, N.; Huang, W.C.; Wang, A.H.; Fu, Q.J.; Gong, Y.N.; Liu, Y.H. Genetic analysis on solanesol content of tobacco. *Chin. Tob. Sci.* **2015**, *36*, 1–7.
28. Xiang, D.H.; Yao, Z.M.; Liu, Y.H.; Gai, X.L.; Du, Y.M.; Yan, N.; Wang, A.H.; Fu, Q.J. Analysis on Solanesol Content and Genetic Diversity of Chinese Flue-Cured Tobacco (*Nicotiana tabacum* L.). *Crop Sci.* **2017**, *57*, 847–855. [[CrossRef](#)]
29. Pan, W.; Liu, W.J.; Weng, B.Q.; Zhan, W.X. Determining solanesolin in tobacco leaves and its extracts using HPLC standardized method. *Chin. Tob. Sci.* **2013**, *34*, 60–61. [[CrossRef](#)]
30. Tamari, F.; Hinkley, C.S.; Ramprasbad, N.A. Comparison of DNA extraction methods using petunia hybrida tissues. *J. Biomol. Tech.* **2013**, *24*, 113–118. [[CrossRef](#)]
31. Van Ooijen, J.W. *JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations*; Kyazma B.V.: Wageningen, The Netherlands, 2006.
32. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **2003**, *164*, 1567–1587. [[CrossRef](#)]
33. Zhang, F.T.; Zhu, Z.H.; Tong, X.R.; Zhu, Z.X.; Qi, T.; Zhu, J. Mixed Linear Model Approaches of Association Mapping for Complex Traits Based on Omics Variants. *Sci. Rep.* **2015**, *5*, 10298. [[CrossRef](#)] [[PubMed](#)]
34. Su, M.; Zhang, C.; Feng, S. Identification and genetic diversity analysis of hybrid offspring of azalea based on EST-SSR markers. *Sci. Rep.* **2022**, *12*, 15239. [[CrossRef](#)] [[PubMed](#)]
35. Evanno, G.; Regnaut, S.G. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
36. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic Marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)] [[PubMed](#)]

37. Bradbury, P.J.; Zhang, Z.W.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2638. [[CrossRef](#)]
38. Rajkhowa, B.; Mehan, S.; Sethi, P.; Prajapati, A.; Suri, M.; Kumar, S.; Bhalla, S.; Narula, A.S.; Alshammari, A.; Alharbi, M.; et al. Activating SIRT-1 Signalling with the Mitochondrial-CoQ10 Activator Solanesol Improves Neurobehavioral and Neurochemical Defects in Ouabain-Induced Experimental Model of Bipolar Disorder. *Pharmaceuticals* **2022**, *15*, 959. [[CrossRef](#)] [[PubMed](#)]
39. Yang, G.H.; Dong, Y.B.; Li, Y.L.; Wang, Q.L.; Shi, Q.L.; Zhou, Q. Verification of QTL for grain starch content and its genetic correlation with oil content using two connected RIL populations in high-oil maize. *PLoS ONE* **2013**, *8*, e53770. [[CrossRef](#)] [[PubMed](#)]
40. Obara, M.; Tamura, W.; Ebitani, T.; Yano, M.; Sato, T.; Yamaya, T. Fine-mapping of qRL6.1, a major QTL for root length of rice seedlings grown under a wide range of NH₄⁺ Concentrations in hydroponic conditions. *Theor. Appl. Genet.* **2010**, *121*, 535–547. [[CrossRef](#)] [[PubMed](#)]
41. Ma, Z.Q.; Xie, Q.; Li, G.Q.; Jia, H.Y.; Zhou, J.Y.; Kong, Z.X.; Li, N.; Yuan, Y. Germplasm, genetics and genomics for better control of disastrous wheat Fusarium head blight. *Theor. Appl. Genet.* **2020**, *133*, 1541–1568. [[CrossRef](#)]
42. Chai, Q.; Wang, X.; Gao, M.; Zhao, X.; Chen, Y.; Zhang, C.; Jiang, H.; Wang, J.; Wang, Y.; Zheng, M.; et al. A glutathione S-transferase GhTT19 determines flower petal pigmentation via regulating anthocyanin accumulation in cotton. *Plant Biotechnol. J.* **2023**, *21*, 433–448. [[CrossRef](#)]
43. Lin, T.Y.; Kao, Y.Y.; Lin, S.; Lin, R.F.; Chen, C.M.; Huang, C.H.; Wang, C.K.; Lin, Y.Z.; Chen, C.M. A genetic linkage map of *Nicotiana plumbaginifolia*/*Nicotiana longiflora* based on RFLP and RAPD markers. *Theor. Appl. Genet.* **2001**, *103*, 905–911. [[CrossRef](#)]
44. Li, Y.L. Difference Analysis of Aroma Components and Association Analysis with SSR Markers in Flue-Cured Tobacco (*Nicotiana tabacum*) Varieties. Master's Thesis, Chinese Academy of Agricultural Sciences, Beijing, China, 2015.
45. Lai, X.; Yan, L.; Lu, Y.; Schnable, J.C. Largely unlinked gene sets targeted by selection for domestication syndrome phenotypes in maize and sorghum. *Plant J.* **2018**, *93*, 843–855. [[CrossRef](#)] [[PubMed](#)]
46. Lu, Y.; Zhang, S.; Shah, T.; Xie, C.; Hao, Z.; Li, X.; Farkhari, M.; Ribaut, J.M.; Cao, M.; Rong, T.Z.; et al. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19585–19590. [[CrossRef](#)] [[PubMed](#)]
47. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006**, *17*, 155–160. [[CrossRef](#)] [[PubMed](#)]
48. Mackay, I.; Powell, W. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* **2007**, *12*, 57–63. [[CrossRef](#)] [[PubMed](#)]
49. Giraud, H.; Bauland, C.; Falque, M.; Madur, D.; Combes, V.; Jamin, P.; Monteil, C.; Laborde, J.; Palaffre, C.; Gaillard, A.; et al. Linkage Analysis and Association Mapping QTL Detection Models for Hybrids Between Multiparental Populations from Two Heterotic Groups: Application to Biomass Production in Maize (*Zea mays* L.). *G3 Genes Genomes Genet.* **2017**, *7*, 3649–3657. [[CrossRef](#)] [[PubMed](#)]
50. Keurentjes, J.J. Genetical metabolomics: Closing in on phenotypes. *Curr. Opin. Plant Biol.* **2009**, *12*, 223–230. [[CrossRef](#)] [[PubMed](#)]
51. Matsuda, F.; Okazaki, Y.; Oikawa, A.; Kusano, M.; Saito, K. Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **2012**, *70*, 624–636. [[CrossRef](#)]
52. Gao, Z.; Ovchinnikova, O.G.; Huang, B.S.; Liu, F.; Williams, D.E.; Andersen, R.J.; Lowary, T.L.; Whitfield, C.; Withers, S.G. High-Throughput “FP-Tag” Assay for the Identification of Glycosyltransferase Inhibitors. *J. Am. Chem. Soc.* **2019**, *141*, 2201–2204. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.