Construction of high-density genetic map and QTL mapping in Nicotiana tabacum backcrossing BC4F3 population using whole-genome sequencing

Tong, Zhijun; Jiang, Sanjie; He, Weiming; Chen, Xuejun; Yin, Lixin; Fang, Dunhuang; Hu, Yafei; Jiao, Fangchan; Zhang, Chi; Zeng, Jianmin

Total number of authors:
14

Published in:
Czech Journal of Genetics and Plant Breeding

Link to article, DOI:
10.17221/8/2021-CJGPB

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

https://doi.org/10.17221/8/2021-CJGPB

Construction of high-density genetic map and QTL mapping in *Nicotiana tabacum* backcrossing BC4F3 population using whole-genome sequencing

Zhijun Tong[1], Sanjie Jiang[2], Weiming He[2], Xuejun Chen[1], Lixin Yin[2], Dunhuang Fang[1], Yafei Hu[2], Fangchan Jiao[1], Chi Zhang[2], Jianmin Zeng[1], Xinfu Wu[1], Shancen Zhao[3], Jianbo Jian[2,4], Bingguang Xiao[1]**

[1]*Key Laboratory of Tobacco Biotechnological Breeding, National Tobacco Genetic Engineering Research Center, Yunnan Academy of Tobacco Agricultural Sciences, Kunming, Yunnan, P.R. China*
[2]*BGI Genomics, BGI-Shenzhen, Shenzhen, Guangdong, P.R. China*
[3]*BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen, Guangdong, P.R. China*
[4]*Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark*

**Corresponding authors: xiaobgsubmission@126.com; jianjianbo@genomics.cn*
Zhijun Tong, Sanjie Jiang and Weiming He contributed equally to this work.

Citation: Tong Z., Jiang S., He W., Chen X., Yin L., Fang D., Hu Y., Jiao F., Zhang Ch., Zeng J., Wu X., Zhao S., Jian J., Xiao B. (2021): Construction of high-density genetic map and QTL mapping in *Nicotiana tabacum* backcrossing BC4F3 population using whole-genome sequencing. Czech J. Genet. Plant Breed., 57: 102–112.

Abstract: Backcrossing is a powerful tool for plant breeding. The improved marker-assisted backcrossing intends to transfer targeted genes or quantitative trait loci (QTLs) of interest from a donor parent into a recurrent parent. In this study, a tobacco BC4F3 population was generated using Y3 and K326 as hybrid parents and YF1-1 as F_1 parents. High-throughput sequencing data of 381 pedigree populations were used to construct high-density genetic maps containing 24 142 high-quality single nucleotide polymorphism (SNP) markers with an average genetic distance of 0.59 cM. A genome module analysis was then performed for all the offspring. A total of forty-three candidate QTLs for six agronomics traits were identified. This study provides original biomarkers for tobacco breeding and offers clues for prospective backcrossing applications in other plants.

Keywords: crossing; single nucleotide polymorphism (SNP); whole genome sequencing (WGS)

Nicotiana, commonly referred to as tobacco, is a genus of annual or limited perennial herbs belonging to the nightshade family. The genus consists of many species, all of which exhibit glandular hairs on their surface. The most common species of *Nicotiana* are *Nicotiana tabacum* and, to a lesser extent *Nicotiana rustica* (Saini et al. 2013). Tobacco originated from the Americas, Oceania, modern Africa, and some South Pacific islands (Swanson 1956; Baldwin 1992). It is one of the most extensively grown non-food crops with a high economic value in commercial agriculture. Currently, the crop is cultivated in more than 120 countries with trillions of values (Nawaz et al. 2019). Tobacco dry leaves are used to manufacture

Supported by the Fundamental Research Program of Yunnan Province (2018FB064), China National Tobacco Company (110201901014 (JY-01), No. 110201801029(JY-06)) and Yunnan Tobacco Company (No. 2020530000241007, 2018530000241005).

https://doi.org/10.17221/8/2021-CJGPB

nicotine products, such as cigars, cigarettes, snuff, tobacco, chewing tobacco, and flavoured shisha. Recently, the potential application of tobacco as a biomass in biopharmaceutical and biofuel applications has generated renewed interest from researchers. Thus, improving tobacco genomic resources is necessary (Tusé et al. 2014; Vanhercke et al. 2014; Ma et al. 2015).

Backcrossing is a powerful tool in plant breeding and can be employed to introduce a specific trait, such as disease resistance, from one line to another (Priyadarshan & Priyadarshan 2019). Marker-assisted backcrossing is an advancement of conventional backcrossing processes, which enable the precise selection and transfer of targeted genes or quantitative trait loci (QTLs) of interest from a donor parent into a recurrent parent (Pandey et al. 2019). Marker-assisted backcrossing has been used to transfer diverse traits of interest, such as those that confer disease and pest resistance or abiotic stress tolerance into many crops (Hasan et al. 2016; Ofner et al. 2016; Kimurto et al. 2017; Xu et al. 2018; Pandey et al. 2019). Specifically, marker-assisted backcrossing has been applied to improve virus resistance, thus increasing the yield for commercial release (Amankwa et al. 2019; Zhao et al. 2019). Molecular markers are essential tools of QTL research. At present, the most widely used molecular markers in tobacco breeding are simple-sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) (Julio et al. 2006; Tong et al. 2019; Agacka-Mołdoch et al. 2021). For example, a recent study identified an important QTL trait for improving brown spot in tobacco through marker-assisted backcrossing (Sun et al. 2018). The number of detected SNPs in the QTL analysis and genome-wide association studies is key to ensuring analytical precision. In one study, SNP markers were identified based on a reference genome without reference to their backcross population. The markers were 4 138 and 2 162 in number, with a total length of 1 944.74 and 2 000.9 cM in the genetic maps, respectively (Xiao et al. 2015).

Nicotiana tabacum is an allotetraploid species ($2n = 4x = 48$) with a massive and complex genome of approximately 4.5 Gb, which is 1.5 times bigger than the human genome. This complex genome arose through intraspecific hybridisation between *Nicotiana Sylvestris* ($2n = 24$, maternal donor with S genome) and *Nicotiana tomentosiformis* ($2n = 24$, paternal donor with T genome) (Gazdová et al. 1995; Leitch et al. 2008). The Tobacco Genome Initiative has been carried out in many countries to obtain more information on the tobacco genome. Till now, more than 300 000 expressed sequence tags have been identified. These sequence tags are published on the RIKEN Center for Biosystems Dynamics Research (http://www.brc.riken.jp/lab/epd/catalog/n_tabacum.html) and the NCBI Gen Bank (http://www.ncbi.nlm.nih.gov/bioproject). At present, only a few studies have focused on the genome-wide SNP identification and genetic map construction in tobacco (Matsuoka et al. 2004; Edwards et al. 2010; Xiao et al. 2015; Tong et al. 2020).

In this study, the elite flue-cured tobacco varieties *Nicotiana tabacum* Y3 and K326 were used as parents. Y3 has excellent elite agronomic traits, whereas K326 has moderate agronomic traits, but a high disease resistance level; it is one of the most famous commercial varieties grown in the USA. The KYF1 as F_1 parents was used to generate the tobacco BC4F3 population, and the whole-genome resequencing of 381 pedigree populations and the three parents were performed. High-quality SNP markers were acquired and used for the genome module analysis and construction of a high-quality and density genetic map. The genetic information obtained in this study will assist in the molecular marker discovery, which is crucial in genetic breeding to improve specific crop phenotypes.

MATERIAL AND METHODS

Construction of tobacco BC4F3 population and trait evaluation. The inbred tobacco line Y3 (recurrent parent) was crossed with another inbred tobacco line K326 (donor parent) to create F_1 crosses. KYF1 is the F_1 parent and was backcrossed with Y3 until BC4F1 was generated. BC4F1 was used to generate BC4F2 through inbreeding; this process was repeated to obtain the BC4F3 population. Then, it was crossed with Y3 to generate the offspring. After four backcrosses and two self-crosses, the BC4F3 mapping population was created, from which a high-density genetic map was constructed. Crop management, including disease and insect pest control, was performed as recommended locally. Before the plants blossomed, the stamens and pistils were bagged and self-pollinated to obtain homozygous seeds. The construction process is shown in Figure 1. For the agronomic trait evaluation, six agronomic traits, including the plant height, nature leaf, topping height, efficient leaf, pitch, and middle leaf length,

https://doi.org/10.17221/8/2021-CJGPB

were measured at the Yanhe and Shilin stations for two consecutive years (2018 and 2019). Nature leaf refers to the total number of leaves in a plant and was recorded after the normal flowering of the tobacco plant. Effective leaf is the number of leaves left on the plant after the plant is cut from the top to about 1/5 of the total height after flowering (topping). The natural plant height is the height of the entire plant (from the ground to the position of the flowers) when the tobacco plant blooms. The topping plant height is the height of the plant after topping. These agronomic traits were measured in the field 65 days after planting. The phenotypic distribution characteristics are shown in Figure S1 in the Electronic Supplementary Material (ESM).

DNA extraction and Sequencing strategy. The genomic DNA of BC4F3 was extracted from the leaves of the 7-week-old tobacco seedlings using a Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) per the manufacturer's instructions. The high-throughput sequencing of the tobacco DNA samples was performed on the MGIseq-500 platform using PE100 (MGI, Shenzhen). The DNA molecular anchor and fluorescence probe were polymerised on the nanospheres, and then the optical signal was collected using a high-resolution imaging system. After the digital processing of the collected optical signal, a valid sequence was obtained. A DNA nano ball (DNB) reduces the error rate of a single copy by enhancing the signal through linear amplification. Moreover, the size of the DNB matches the size of the active sites on the chip, and each site combines with a DNA nanosphere. This improves the utilisation efficiency of the sequencing chip while ensuring the accuracy of the sequencing.

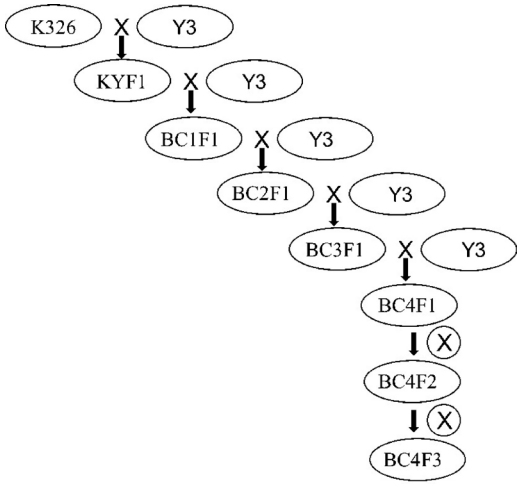

Figure 1. The process of constructing the tobacco BC4F3 population

Y3 and K326 species are hybrid parents; KYF1 is the F_1 parent and backcrossed with Y3 until BC4F1 was generated; BC4F1 generated BC4F2 by inbreeding; this process was repeated to obtain the BC4F3 population

Data filtering. The raw data were filtered and processed to remove adaptor contamination, and low-quality reads. The specific filtering process was performed using SOAPnuke (Chen et al. 2018) developed by the Beijing Genomics Institute (BGI), as follows: (1) remove the reads containing the adaptor; (2) remove the low-quality reads (the bases with a quality value ≤ 20 that account for 40% of the bases); (3) the reads with an N base ratio above 5% were removed.

A tetraploid genome K326 reported in a previous study was used as the reference genome (ftp://ftp.sgn.cornell.edu/genomes/Nicotiana_tabacum/edwards_et_al_2017/assembly/Nitabv4.5_genome_Scf_Edwards2017.fasta.gz) (Edwards et al. 2017). This reference genome had 1 084 432 scaffolds. The genome size was 4 694 948 798 bp with an effective genome size of 4 049 118 047 bp. All scaffolds less than 500 bp in length were filtered, and the remains were used for the analysis. The remaining 942 190 scaffolds (with a total length of 4 651 541 558 bp) were connected to 40 pseudo-groups for analysis (150 N Scaffold for connection between each scaffold). The "mem" algorithm of the short sequence alignment software BWA (Ver. 0.7.12-r1039) was used to align the clean reads to the reference genome (alignment parameters: -t 8 -k 19 -M -R). Two tools SortSam.jar and MarkDuplicates.jar of Picard (Ver. 1.117) were used to sort and eliminate the redundancy from the alignment result. The alignment files from different libraries/lanes of the same sample were merged using the MergeSamFiles.jar tool. Given that the reference genome was of a tetraploid species, only reads with a mapQ value above 10 and properly paired were selected for the subsequent mutation detection and analysis (filtration parameter: samtools view -f 2-q 10).

SNPs and Indels detection. We used GATK (GATK3.3.0) (McKenna et al. 2010) to detect the SNP and indel variation in the tobacco population. To improve the accuracy of the SNP sites around the indel, the RealignerTargetCreator and IndelRealigner tools of GATK were used to re-align and correct the sequences near the indel. Bam files after treatment were used for the mutation detection. The HaplotypeCaller in the GATK was adopted to conduct the individual SNP mutation detection on the alignment

https://doi.org/10.17221/8/2021-CJGPB

files after filtering and generated the GVCF files of each chromosome for each individual. The GenotypeGVCFs were used to integrate the information on each chromosome variation of all the individuals in the population. A VCF format file of the population mutation results was generated, including the SNP and indel mutation. The SNP and indel mutation were extracted and filtered through the SelectVariants and VariantFiltration tools in GATK to acquire highly credible population variation information.

Construction of genetic map. The original SNP loci were filtered to remove the SNPs with a missing rate above 80% using the VCFtools software (Ver. 4.2), and only the second class was retained. Due to the high missing rate, Beagle software (Ver. 4.1) was applied to impute the missing loci of the genotype. The completed population SNPs were filtered under the following rules: (1) only SNPs in which both parents had polymorphic loci and where F_1 was heterozygous were selected; (2) the loci according to a chi-square test of the population separation by a significant level; (3) only one marker was reserved within a physical distance of 50 k (theoretically, the closer the physical distance between markers, the lower the probability of recombination and the higher the probability of marker redundancy). The filtered population SNPs were analysed using LepMap3 software (Rastas 2017). Firstly, the ParentCall2 function was applied to obtain markers for the subsequent analysis. Then, the JoinSingles2All function was performed to regroup the single marker into the linkage group. The SeparateChromosomes2 function was used to divide the linkage group. Finally, the genetic distance between each linkage group was calculated using the OrderMarkers2 function.

The analysis of the offspring genome module was also performed herein. A chromosomal level genome mounted on 274 inbred tobacco lines was used for the module analysis. The markers in the BC4F3 population were targeted on the G1 genome based on the marker location on the original scaffold. A binmap module diagram was constructed utilising the 24 142 markers located on the chromosome. Of note, the chromosome was regarded as a unit during the process. A specific number of continuous SNP markers on the chromosome were used as the unit to estimate the chromosome recombination time. This assisted in predicting the probability that each bin of the individual offspring came from the male and female parent.

QTL mapping. MapQTL 6.0 (Van Ooijen 2009) was adopted for the QTL analysis of the phenotypes with the related traits. The phenotypic data, including the trait, linkage group, and sample genotype, were sorted into MapQTL in the required format. The PERMUTATION test was utilised to obtain the genome-wide limit of detection (LOD) thresholds. A total of 10 000 permutations were used for each trait to derive the empirical LOD threshold. Next, the interval mapping model with a regression approach was introduced to obtain the LOD values of markers associated with the candidate traits. Finally, the multiple-QTL mapping (MQM) analysis was performed to connect the identified markers to the significant QTLs in each group. The markers from the candidate regions with LOD values greater than the threshold were annotated.

RESULTS

Sequencing data evaluation. A total of 3.51 Tb of raw data were obtained from 381 offspring samples. Subsequently, 3.24 Tb of clean data were retained after filtering out the adapter-containing reads, duplicates arising from the polymerase chain reaction (PCR) amplification during the library construction, and low-quality reads. The average sequencing data per sample was around 8.5 Gb with an average sequencing depth of about 1.8 X. The clean data of 381 offspring samples were mapped to the pseudochromosome of the reference genome, and the statistics of each sample on the mapping rate, coverage rate, and depth were collected. The mapping rate of each sample was above 95%, with a pair mapping rate of over 90% (Table S1 in the ESM). The coverage rate was around 55% with a float of 5%, whereas the sequencing depth of every sample was more than 1X (Figure S2 in the ESM).

SNPs and Indels calling. The population SNP and indel variation were filtered according to the requirements of the map population type. In total, 10 915 091 original population SNPs were obtained from 384 samples, including two hybridisation parents and one F_1 parent. In addition, 1 257 782 original population indels were obtained. All the SNPs and indels were distributed across the 40 pseudochromosomes. To obtain the density of the SNPs and indels, the original population of the SNPs and indels were anchored to the chromosomes based on the physical location between the chromosomes and scaffolds. As a result, 5 844 611 SNPs and 570 728 indels were restored to the chromosomes, and the density was calculated with a 200 kb window (Figure 2A, B). A transition and transversion analysis of the original

https://doi.org/10.17221/8/2021-CJGPB

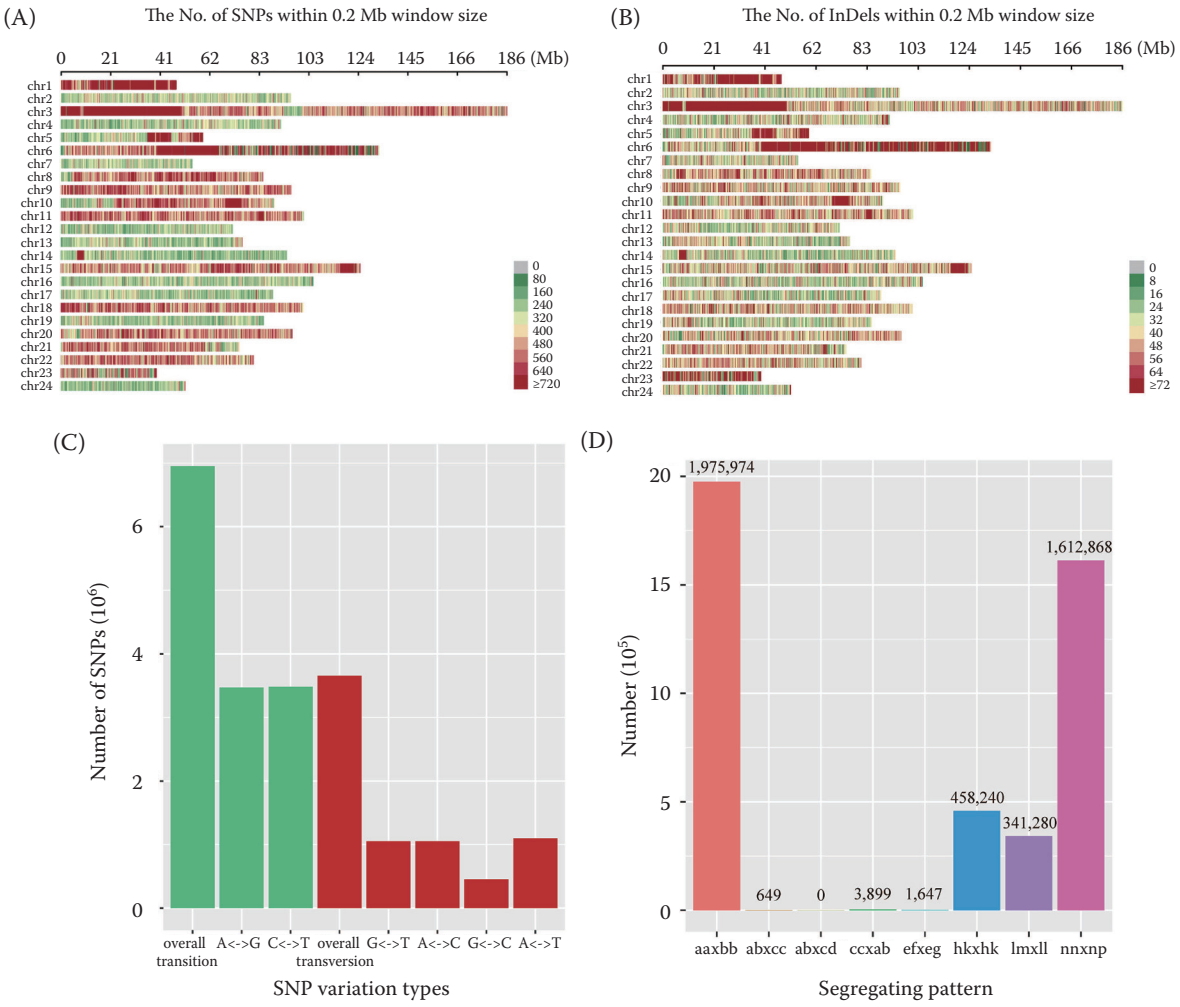

Figure 2. Statistics of the SNPs and indels: the density of the indels within a 200 kb window size (A), density of the SNPs within a 200 kb window size (B), variation types (transition and transversion) of the SNPs (C), statistics of all the polymorphic marker isolation types between the parents (D); the abscissa represents the type of markers between the parents, and the ordinate represents the number of markers

population of SNPs indicated that the nn X np form occupied the highest number of markers in various segregation patterns (Figure 2C, D).

Genetic map. The originally acquired population SNP loci were selected using VCFtools, and 10 610 359 loci were retained. After imputation filtering and the chi-square test (P = 0.01; value < 9.2103), 815 986 candidate loci were obtained. Notably, the number of our candidate loci (815 986) was above the upper limit of the major composition analysis software. Therefore, a filtering process was performed before constructing a genetic map. Notably, one marker was reserved within a 50 k physical distance on one scaffold. After the filtering step, we found 24 647 markers suitable for the genetic map construction. These markers were distributed over 15 669 scaffolds. A total of 24 632 information markers were obtained after the software screening. By setting the LOD value to 20, all the markers were divided into 24 linkage groups. Finally, a high-density genetic map containing 24 142 SNP markers was successfully constructed. The map included 4 895 bin markers with a genetic distance of 2 885.36 cM and an average genetic distance of 0.59 cM (Figure 3 and Table 1, Table S2 in the ESM). In addition, the analysis of the offspring genome module was performed using the same 24 142 markers of the BC4F3 population. Based on the marker location on the original scaffold, 17 970 markers were located on the G1 genome. A binmap was constructed based on the genotype of these located markers using the chromosome as a unit (Figure 4).

https://doi.org/10.17221/8/2021-CJGPB

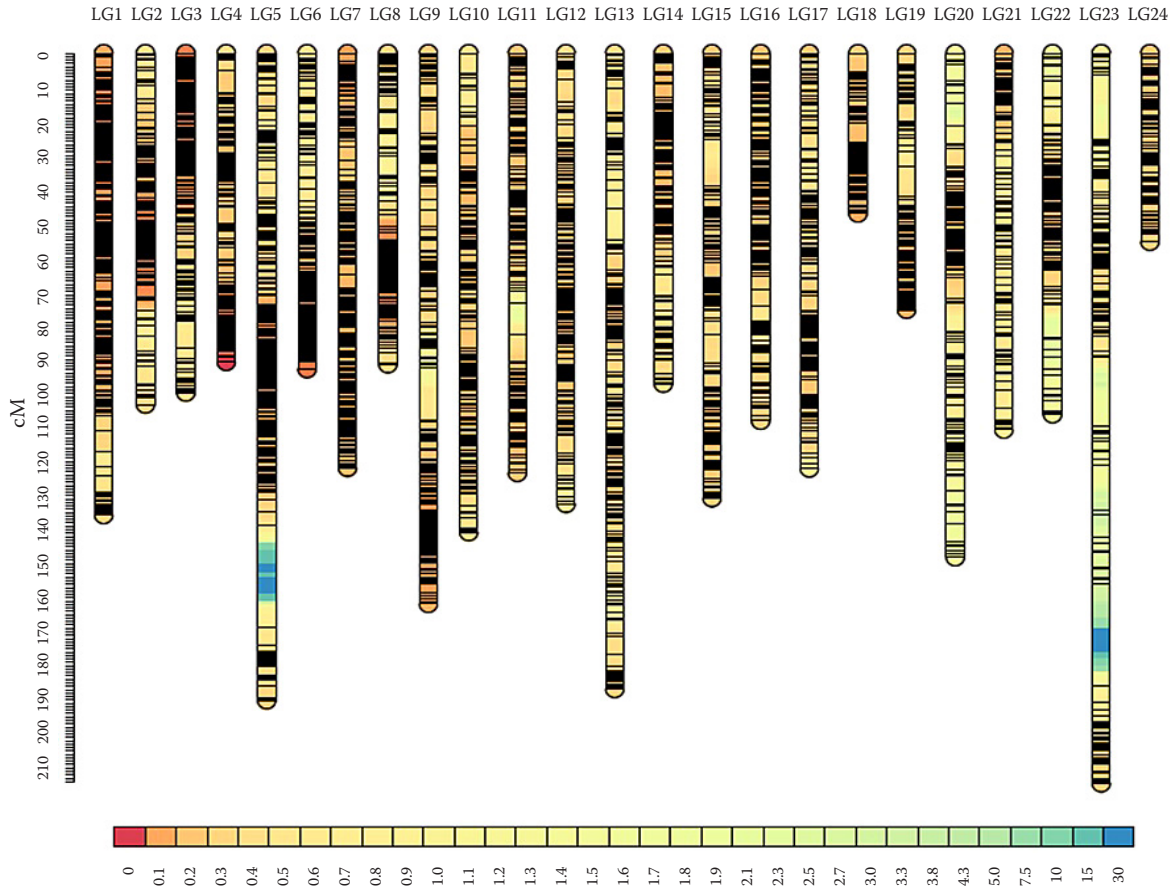

Figure 3. Genetic marker distribution profile

The background colour of the linkage group, from red to blue, represents the marker density from large to small; the 'LinkageMapView' R package was used to graph the distribution of the genetic markers

Table 1. Statistics of the genetic map

Linkage map	SNP marker	Bin marker	Distance (cM)	Average distance (cM)	Linkage map	SNP marker	Bin marker	Distance (cM)	Average distance (cM)
LG1	3 118	339	135.7	0.4	LG14	665	172	97.11	0.565
LG2	2 139	179	103.2	0.577	LG15	632	205	130.71	0.638
LG3	2 032	234	99.72	0.426	LG16	612	213	107.99	0.507
LG4	1 851	195	90.67	0.465	LG17	540	204	122.02	0.598
LG5	1 503	303	190.14	0.628	LG18	483	107	47.1	0.44
LG6	1 442	213	92.84	0.436	LG19	442	139	75.41	0.543
LG7	1 397	285	121.8	0.427	LG20	386	169	148.07	0.876
LG8	1 303	155	91.4	0.59	LG21	377	149	110.59	0.742
LG9	1 262	269	161.81	0.602	LG22	358	137	106.11	0.775
LG10	858	224	140.71	0.628	LG23	319	178	214.46	1.205
LG11	759	215	123.26	0.573	LG24	285	117	55.36	0.473
LG12	706	243	132.38	0.545	Total	24 142	4 895	2 885.36	0.59
LG13	673	251	186.8	0.744					

SNP markers were separated into 24 linkage group; LOD value equal to 20; LG – linkage group; cM – centimorgan

https://doi.org/10.17221/8/2021-CJGPB

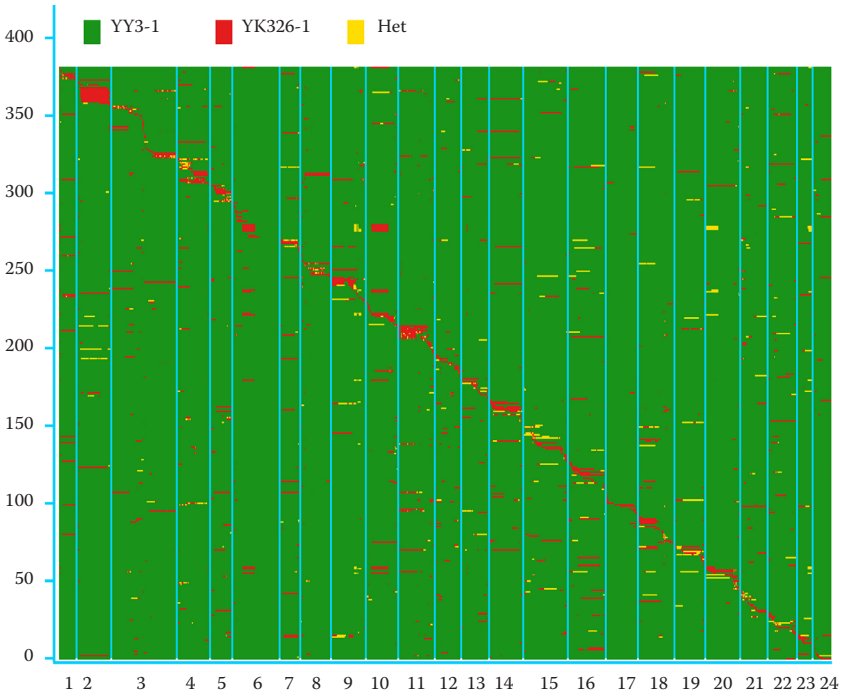

Figure 4. Binmap module profile; the genome for this module analysis is a chromosomal level genome based on 3 tobacco inbred genetic maps; for the convenience of the presentation, the genome is defined as G1
YY3-1 and YK326-1 are the individual parents used for the initial hybridisation; Het stands for KYF1 as F_1 parents were used to generate the tobacco BC4F3 population

QTL mapping. The quantitative trait loci were located based on the genetic map and provided the phenotype information. The selected LOD value was 3.0. In total, forty-three candidate loci were identified in six agronomic phenotypes, including the plant height, nature leaf, topping height, efficient leaf, pitch, and middle leaf length. Among these phenotypes, fifteen candidate loci were obtained in the nature leaf phenotype; however, only two loci were found in the efficient leaf loci (Table 2).

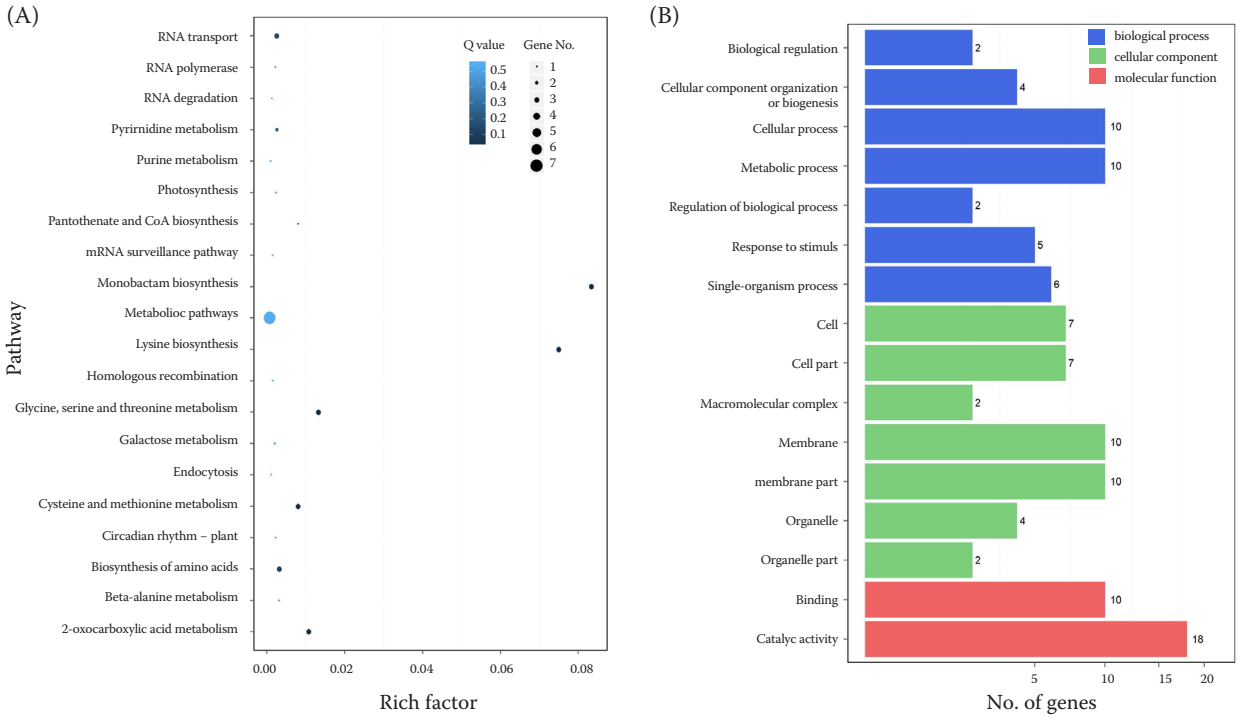

Figure 5. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of the QTLs: the metabolic pathway and catalytic activity were the most enriched pathways by the KEGG (A) and gene ontology (B) analysis in the nature leaf QTL mapping

https://doi.org/10.17221/8/2021-CJGPB

Table 2. Statistics of the QTLs

Phenotype	LG	Distance (cM)	Marker ID	LOD value	Contribution (%expl)
Plant heights	17	1.29	Nitab4.5_0122456_697	3.7	11.2
	17	14.95	Nitab4.5_0000198_371816	3.7	11.2
	17	101.54	Nitab4.5_0001520_266999	3.09	9.5
	19	3.88	Nitab4.5_0279350_588	3.7	11.2
	21	7.93	Nitab4.5_0002849_129028	3.7	11.2
	21	8.86	Nitab4.5_0000823_523053	3.7	11.2
	21	9.04	Nitab4.5_0023910_506	3.7	11.2
	21	9.41	Nitab4.5_0002017_328447	3.7	11.2
	21	9.04	Nitab4.5_0000823_397377	3.03	9.3
Nature leaf	15	98.6	Nitab4.5_0267005_237	4.1	12.4
	15	40	Nitab4.5_0000313_734881	3.59	10.9
	15	129.24	Nitab4.5_0004179_68394	3.59	10.9
	15	40.18	Nitab4.5_0000313_796187	3.4	10.4
	15	40.37	Nitab4.5_0000313_1028406	3.4	10.4
	15	40.37	Nitab4.5_0000313_856745	3.4	10.4
	15	40.37	Nitab4.5_0000313_924575	3.4	10.4
	15	40.37	Nitab4.5_0000313_608579	3.4	10.4
	15	40.37	Nitab4.5_0000313_974890	3.4	10.4
	15	98.6	Nitab4.5_0028180_3778	3.3	10.1
	17	1.29	Nitab4.5_0122456_697	4.1	12.4
	17	14.95	Nitab4.5_0000198_371816	3.59	10.9
	17	14.95	Nitab4.5_0000198_543642	3.4	10.4
	18	13.86	Nitab4.5_0179109_695	3.4	10.4
	18	16.81	Nitab4.5_0007820_23752	3.4	10.4
Topping heights	2	59.61	Nitab4.5_0002134_311833	3.71	11.3
	2	58.14	Nitab4.5_0002908_16685	3.15	9.6
	4	85.5	Nitab4.5_0352858_493	3.94	11.9
	6	66.27	Nitab4.5_0228918_1293	3.33	10.2
	7	6.46	Nitab4.5_0409488_602	3.17	9.7
	8	55.42	Nitab4.5_0085135_247	3.57	10.8
	17	76.24	Nitab4.5_0002102_222217	4.07	12.3
Efficient leaf	3	34.87	Nitab4.5_0910012_297	3.07	9.4
	16	61.08	Nitab4.5_0014263_14206	3.04	9.3
Pitch	3	22.69	Nitab4.5_0248435_973	3.48	10.6
	8	64.09	Nitab4.5_0000016_1193474	3.12	9.6
	12	95.26	Nitab4.5_0762992_363	3.22	9.8
	15	46.27	Nitab4.5_0000249_992265	3.57	10.9
	15	112.07	Nitab4.5_0270462_111	3.57	10.9
	17	101.54	Nitab4.5_0001520_266999	3.56	10.8
	17	13.29	Nitab4.5_0000009_454740	3.13	9.6
Middle leaf lengths	3	48.53	Nitab4.5_0000289_510228	3.31	10.1
	6	83.24	Nitab4.5_0028993_2042	3.8	11.5
	11	53.15	Nitab4.5_0000030_2269959	3.19	9.8

A total of 43 candidate loci in 6 agronomic phenotypes were separated into 14 linkage groups (LG), LOD value equal to 3.0

https://doi.org/10.17221/8/2021-CJGPB

All the genes in the loci were annotated (Table S3 in ESM). The loci were separated into 14 linkage groups (LG). We found nine markers in LG17 across four kinds of phenotypes, including the plant height, nature leaf, topping height, and pitch. Ten markers in LG15 were obtained in the nature leaf, whereas two markers were discovered in the pitch phenotype. In addition, seven loci were observed in the topping height trait. These seven loci could be grouped into six different LGs. The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and Gene Ontology (GO) terms indicated that these loci were enriched in the metabolic pathway and catalytic activity in the nature leaf QTL mapping (Figure 5). A gene *pbsr* (N0000313g0040), which is a photosystem II subunit R, was identified in the QTLs. *Pbsr* was closely related to the photosynthesis, which might be a critical factor for the nature leaf trait. Another gene *Root phototropism protein 3* (N0000313g0110), a light-dependent dynamic modulator of phototropin, was also identified in the QTLs relating to the nature leaf trait. The interaction between the key genes and the traits underlies their important role in tobacco development and breeding.

DISCUSSION

Backcrossing is one of the most basic methods of germplasm improvement. The approach can be used to develop new tobacco varieties with high quality, disease resistance, and high yield. Marker-assisted backcrossing is the simplest form of marker-assisted selection (MAS), which plays a crucial role in DNA marker identification (Xiao et al. 2015). The present study provides an excellent example for constructing a high-density linkage map and downstream genetic investigations, such as QTL mapping. Herein, 5 844 611 SNPs and 570 728 indels were restored to the chromosomes and used for the genetic map construction and QTL mapping.

Developing molecular markers for the genetic map construction is the basis of QTL mapping and MAS breeding. The first tobacco genetic map was established using 99 F_2 plants, obtained by hybridisation of *N. plumbaginifolia* × *N. longiflora* (Lin et al. 2001). The genetic similarity of cultivated tobacco is considered high, which was difficult in the polymorphism detection (Ren & Timko 2001; Qi et al. 2006; Xiao & Yang 2007). Another widely acknowledged tobacco SSR genetic map was published in 2011. The genetic map had 2 363 loci distributed on 24 chromosomes with a total distance of 3 270 cM and an average distance of 1.5 cM (Bindler et al. 2011). In the present study, high-throughput sequencing was used to construct high-quality and high-density genetic maps. Furthermore, QTL mapping was performed, and a total genetic distance of 2 885.36 cM and an average distance of 0.59 cM were obtained. These data were in line with the requirement for the QTL mapping analysis. Moreover, we identified 43 trait loci of six agronomic traits through the QTL mapping.

A few studies have conducted a QTL mapping analysis of tobacco and attempted to identify the QTLs of different traits, including the agronomic traits, chemical components, and quality (Julio et al. 2006; Li et al. 2011; Cai et al. 2014). However, the average number of QTLs identified for one trait was low due to various reasons. One of the reasons could be that the quality of the genetic map used for the QTL scanning analysis of the whole genome was poor. Notably, a poor-quality genetic map can lead to a low QTL detection ability of the targeted phenotype. The genetic map constructed in this study is of high quality and saturation, enabling a more robust QTL mapping analysis of the whole tobacco genome. This study provides more evidence for the large-scale identification of QTLs of important traits in tobacco. Thus, the findings of this study will be useful in the marker-assisted selection breeding of tobacco in the future.

REFERENCES

Agacka-Mołdoch M., Rehman Arif M.A., Lohwasser U., Doroszewska T., Lewis R.S., Börner A. (2021): QTL analysis of seed germination traits in tobacco (*Nicotiana tabacum* L.). Journal of Applied Genetics, Advance Access published March 5, 2021, doi:10.1007/s13353-021-00623-6.

Amankwa G.A., Mishra S., Shearer A.D., Saude C., Van Hooren D.L., Richmond M.D. (2019): CTH144 flue-cured tobacco F_1 hybrid. Canadian Journal of Plant Science, 99: 966–968.

Baldwin B.G. (1992): Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. Molecular Phylogenetics and Evolution, 1: 3–16.

Bindler G., Plieske J., Bakaher N., Gunduz I., Ivanov N., Van der Hoeven R., Ganal M., Donini P. (2011): A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theoretical and Applied Genetics, 123: 219–230.

https://doi.org/10.17221/8/2021-CJGPB

Cai C., Deng H., Feng J., Cheng L., Yu J., Wang X. (2014): Preliminary genetic analysis of relevant traits in fluecured tobacco at growth and development period. Acta Tabacaria Sinica, 20: 70–77, 84.

Chen Y., Chen Y., Shi C., Huang Z., Zhang Y., Li S., Li Y., Ye J., Yu C., Li Z., Zhang X., Wang J., Yang H., Fang L., Chen Q. (2018): SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience, 7: 1–6.

Edwards K.D., Bombarely A., Story G.W., Allen F., Mueller L.A., Coates S.A., Jones L. (2010): TobEA: An atlas of tobacco gene expression from seed to senescence. BMC Genomics, 11: 142.

Edwards K.D., Fernandez-Pozo N., Drake-Stowe K., Humphry M., Evans A.D., Bombarely A., Allen F., Hurst R., White B., Kernodle S.P., Bromley J.R., Sanchez-Tamburrino J.P., Lewis R.S., Mueller L.A. (2017): A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genomics, 18: 448.

Gazdová B., Široký J., Fajkus J., Brzobohatý B., Kenton A., Parokonny A., Heslop-Harrison J.S., Palme K., Bezděk M. (1995): Characterization of a new family of tobacco highly repetitive DNA, GRS, specific for the *Nicotiana tomentosiformis* genomic component. Chromosome Research, 3: 245–254.

Hasan M.M., Rafii M.Y., Ismail M.R., Mahmood M., Alam M.A., Abdul Rahim H., Malek M.A., Latif M.A. (2016): Introgression of blast resistance genes into the elite rice variety MR263 through marker-assisted backcrossing. Journal of the Science of Food and Agriculture, 96: 1297–1305.

Julio E., Denoyes-Rothan B., Verrier J.L., Dorlhac De Borne F. (2006a): Detection of QTLs linked to leaf and smoke properties in *Nicotiana tabacum* based on a study of 114 recombinant inbred lines. Molecular Breeding, 18: 69–91.

Kimurto P., Oyier M., Mulwa R., Songok S., Towett B., Cheruiyot E., Varshney R.K., Gaur P.M., Mahender T., Ganga Rao N.V.P.R., Silim S., Siambi M. (2017): Performance of marker assisted backcross breeding (MABC) elite chickpea lines under drought conditions in Kenya. In: InterDrought-V, Hyderabad, Feb 21–25, 2017.

Leitch I.J., Hanson L., Lim K.Y., Kovarik A., Chase M.W., Clarkson J.J., Leitch A.R. (2008): The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Annals of Botany, 101: 805–814.

Li H.-L., Chen M.-X., Zhou D.-X., Chen S.-H., Tao A.-F., Li Y.-K., Ma H.-B., Qi J.-M., Guo Y.-C. (2011): QTL analysis of six important traits in tobacco (*Nicotiana tabacum* L.). Acta Agronomica Sinica, 37: 1577–1584.

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.C. (2001): A genetic linkage map of *Nicotiana plumbaginifolia*/*Nicotiana longiflora* based on RFLP and RAPD markers. Theoretical and Applied Genetics, 103: 905–911.

Ma J.K.C., Drossard J., Lewis D., Altmann F., Boyle J., Christou P., Cole T., Dale P., van Dolleweerd C.J., Isitt V., Katinger D., Lobedan M., Mertens H., Paul M.J., Rademacher T., Sack M., Hundleby P.A.C., Stiegler G., Stoger E., Twyman R.M., Vcelar B., Fischer R. (2015): Regulatory approval and a first-in-human phase I clinical trial of a monoclonal antibody produced in transgenic tobacco plants. Plant Biotechnology Journal, 13: 1106–1120.

Matsuoka K., Demura T., Galis I., Horiguchi T., Sasaki M., Tashiro G., Fukuda H. (2004): A comprehensive gene expression analysis towards the understanding of growth and differentiation of tobacco BY-2 cells. Plant & Cell Physiology, 45: 1280–1289.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. (2010): The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20: 1297–1303.

Nawaz Z., Kakar K.U., Ullah R., Yu S., Zhang J., Shu Q.Y., Ren X.-L. (2019): Genome-wide identification, evolution and expression analysis of cyclic nucleotide-gated channels in tobacco (*Nicotiana tabacum* L.). Genomics, 111: 142–158.

Ofner I., Lashbrooke J., Pleban T., Aharoni A., Zamir D. (2016): Solanum pennellii backcross inbred lines (BILs) link small genomic bins with tomato traits. Plant Journal, 87: 151–160.

Pandey K., Dangi R., Prajapati U., Kumar S., Maurya N.K., Singh A.V., Pandey A.K., Singh J., Rajan R. (2019): Advance breeding and biotechnological approaches for crop improvement: A review. International Journal of Chemical Studies, 7: 837–841.

Priyadarshan P.M., Priyadarshan P.M. (2019): Introduction to plant breeding. In: Plant Breeding: Classical to Modern, Singapore, Springer: 3–33.

Qi J.M., Wang T., Chen Sh., Fance P.P., Tao A.F., Liang J.X., Wu W.R. (2006): Genetic diversity and genetic relatives analysis of tobacco germplasm based on inter-simple sequence repeat (ISSR). Acta Agronomica Sinica, 32: 373–378.

Rastas P. (2017): Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics, 33: 3726–3732.

Ren N., Timko M.P. (2001): AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome, 44: 559–571.

Saini A.K., Chauhan P.K., Singh V., Sharma P. (2013): Phytochemical, antioxidant and in vitro antibacterial activity

https://doi.org/10.17221/8/2021-CJGPB

of aqueous and ethanolic fruit extracts of *Kigelia africana*. Indian Journal of Pharmaceutical and Biological Research, 1: 46–52.

Sun M., Cheng L., Jiang C., Zhu C., Ren M., Zhang Y., Zhang Y., Liu D., Zhao Q., Geng R., Hu X., Yang A., Wang Y. (2018): Identification of a major QTL affecting resistance to brown spot in tobacco (*Nicotiana tabacum* L.) via linkage and association mapping methods. Euphytica, 214: 1–14.

Swanson C.P. (1956): The Genus Nicotiana. Origins, Relationships and Evolution of Its Species in the Light of Their Distribution, Morphology and Cytogenetics. Parts I-V: Distribution, Morphology, Cytology of Species, Cytology of F1 Interspecific Hybrids, Phylesis. Thomas Harper Goodspeed The Genus Nicotiana. Origins, Relationships and Evolution of Its Species in the Light of Their Distribution, Morphology and Cytogenetics. Part VI: Taxonomy. Goodspeed T.H., Wheeler H.M., Hutchinson P.C. The Quarterly Review of Biology, 31: 47–48.

Tong Z.J., Zhang Y.H., Chen X.J., Zeng J.M., Fang D.H., Xiao B.G. (2019): Mapping of quantitative trait loci conferring resistance to brown spot in cigar tobacco cultivar Beinhart1000-1. Acta Agronomica Sinica, 45: 477–482.

Tong Z., Zhou J., Xiu Z., Jiao F., Hu Y., Zheng F., Chen X., Li Y., Fang D., Li S., Wu X., Zeng J., Zhao S., Jian J., Xiao B. (2020): Construction of a high-density genetic map with whole genome sequencing in *Nicotiana tabacum* L. Genomics, 112: 2028–2033.

Tusé D., Tu T., McDonald K.A. (2014): Manufacturing economics of plant-made biologics: Case studies in therapeutic and industrial enzymes. BioMed Research International, 2014: 256135.

Van Ooijen J.W. (2009): MapQTL 6, Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species. Northampton,Advance Access.

Vanhercke T., El Tahchy A., Liu Q., Zhou X.R., Shrestha P., Divi U.K., Ral J.P., Mansour M.P., Nichols P.D., James C.N., Horn P.J., Chapman K.D., Beaudoin F., Riuz-Lopez N., Larkin P.J., de Feyter R.C., Singh S.P., Petrie J.R. (2014): Metabolic engineering of biomass for high energy density: Oilseed-like triacylglycerol yields from plant leaves. Plant Biotechnology Journal, 12: 231–239.

Xiao B.-G., Yang B.-C. (2007): Assessment of genetic diversity among tobacco germplasms by ISSR markers. Scientia Agricultura Sinica, 40: 2153–2161.

Xiao B., Tan Y., Long N., Chen X., Tong Z., Dong Y., Li Y. (2015): SNP-based genetic linkage map of tobacco (*Nicotiana tabacum* L.) using next-generation RAD sequencing. Journal of Biological Research-Thessaloniki, 22: 11.

Xu Y., Zhang X.Q., Harasymow S., Westcott S., Zhang W., Li C. (2018): Molecular marker-assisted backcrossing breeding: an example to transfer a thermostable β-amylase gene from wild barley. Molecular Breeding, 38: 1–9.

Zhao L., Li W., Wang B., Gao Y., Sui X., LiuY., Chen X., Yao X., Jiao F., Song Z. (2019): Development of a PVY resistant flue-cured tobacco line via EMS mutagenesis of eIF4E. Agronomy, 10: 36.

Received: February 1, 2021
Accepted: May 24, 2021
Published online: June 8, 2021