THE CROP JOURNAL 1 (2013) 151–159

Available online at www.sciencedirect.com

ScienceDirect

Mapping epistasis and environment × QTX interaction based on four -omics genotypes for the detected QTX loci controlling complex traits in tobacco

Liyuan Zhou[a,1], *Ruiyuan Li*[b,1], *Longjiang Fan*[a], *Yuewei Shi*[b], *Zhihong Wang*[b], *Shengdong Xie*[b], *Yijie Gui*[a], *Xueliang Ren*[b,*], *Jun Zhu*[a,*]

[a]*Key Laboratory of Crop Germplasm Resource of Zhejiang Province, Department of Agronomy, Zhejiang University, Hangzhou, Zhejiang 310058, China*

[b]*Key Laboratory of Molecular Biology, Guizhou Academy of Tobacco Sciences, Guiyang, Guizhou 550013, China*

ARTICLE INFO

Article history:
Received 16 April 2013
Received in revised form
2 August 2013
Accepted 2 September 2013
Available online 11 September 2013

Keywords:
Association mapping study
Complex trait analysis
Epistasis effects
Environment × treatment interaction
Plant architecture control
QTX locus mapping

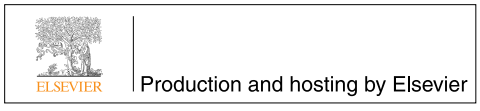
ABSTRACT

Using newly developed methods and software, association mapping was conducted for chromium content and total sugar in tobacco leaf, based on four -omics datasets. Our objective was to collect data on genotype and phenotype for 60 leaf samples at four developmental stages, from three plant architectural positions and for three cultivars that were grown in two locations. Association mapping was conducted to detect genetic variants at quantitative trait SNP (QTS) loci, quantitative trait transcript (QTT) differences, quantitative trait protein (QTP) variability, and quantitative trait metabolite (QTM) changes, which can be summarized as QTX locus variation. The total heritabilities of the four -omics loci for both traits tested were 23.60% for epistasis and 15.26% for treatment interaction. Epistasis and environment × treatment interaction had important impacts on complex traits at all -omics levels. For decreasing chromium content and increasing total sugar in tobacco leaf, six methylated loci can be directly used for marker-assisted selection, and expression of ten QTTs, seven QTPs and six QTMs can be modified by selection or cultivation.

1. Introduction

Most important agronomic traits are complex [1]. Decoding the genetic constitution of complex traits and obtaining information on phenotypic variation are some of the most important challenges of genetic analysis. In contrast to Mendelian traits controlled by individual major genes, the phenotypic variations of complex traits are due to segregation of multiple loci with small effects which are sensitive to environmental factors. Using gel-based or next generation

* *Corresponding authors.*
E-mail addresses: renxuel@126.com (X. Ren), jzhu@zju.edu.cn (J. Zhu).
[1] These authors contributed equally to this work.
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.



2214-5141/$ – see front matter © 2013 Production and hosting by Elsevier B.V. on behalf of Crop Science Society of China and Institute of Crop Science, CAAS.
http://dx.doi.org/10.1016/j.cj.2013.09.001

sequencing and molecular marker analysis technology, genetic linkage analysis of quantitative trait locus (QTL) has become one of the most commonly used techniques in complex trait analysis [2,3]. QTL analysis can also be combined with available transcript, protein and metabolite profiles for a mapping or association population generally resulting in regression analysis between markers and endogenous phenotypes (e.g. gene expression levels, protein modification, or levels of a particular secondary metabolite). By using such molecular, protein or biochemical variants as trait phenotypes, the linkage or association QTL mapping is known as expression-QTL (eQTL), protein-QTL (pQTL) and metabolite-QTL (mQTL), respectively.

These full pathway molecular phenotypes, from transcript to translated protein to metabolic product, help elucidate genotypic variation that underlies morphological and physiological traits [4]. However, due to the limited recombination events in the mapping population derived from bi-parental crosses, regardless of the choice of either molecular variants or complex phenotypic traits, the QTLs detected via linkage analysis can only be mapped to large genomic regions [5]. Recently, the increasing use of high-throughput molecular techniques from the -omics sciences (genomics, transcriptomics, proteomics and metabolomics) has created a huge amount of -omics data, which can be applied to traditional genetic or agronomic experiments [6].

Recent genotyping methods (e.g. for SNPs) produce much higher density genome coverage than traditional gel-based molecular markers, and thus are much more informative. This leads to the potential of association mapping for complex trait analyses [7]. Compared with linkage mapping, association mapping is a high-resolution method based on linkage disequilibrium (LD), and has recently been applied to plant populations [8–10]. Here we propose that instead of just using SNPs as variants in LD analysis for the detection of QTL, the molecular variants of four -omics datasets can also be used as generalized genotypes in association mapping for complex traits. This multi-omics approach would be crucial for the identification of what we term quantitative trait SNPs (QTS), quantitative trait transcripts (QTT) [5], quantitative trait proteins (QTP), and quantitative trait metabolites (QTM). The association mapping based on the four -omics datasets can in compendium or in conjunction be called QTX mapping, a more general term we suggest for use in this type of research.

In addition to the detection of QTX themselves, G × G interaction (epistasis) and G × E interaction can also be detected by QTX mapping. These interaction effects may explain a considerable proportion of the missing heritability associated with QTL based on individual molecular marker loci [11]. In general, the size of datasets involved in QTX mapping will be an order of magnitude larger than the size of datasets for typical QTL detection. This has presented a challenge that has hardly been matched by contemporary hardware, making QTX analysis difficult to perform efficiently until recently with the application of GPU (Graphic Processor Unit) parallel computation which has significantly increased the ability to solve computationally intensive biological problems [12,13]. GPU parallel computation addresses the ever-increasing demand for higher computational speed and has paved the way for the analysis of -omics data from large scale or multiple layer experiments.

Tobacco (*Nicotiana tabacum* L.) is one of the most important model plants in genetic analysis. The quality of tobacco leaves is determined by the composition and quantity of metabolites [14], which are quantitative traits controlled by multiple genes and environmental factors. Previous studies on genetic architecture and regulated network of such complex traits were unable to comprehensively dissect the mechanism of catabolism, anabolism or accumulation of these metabolites [15–22]. Implementation of QTX mapping by using various types of -omics datasets in tobacco was predicted as a useful opportunity to illustrate the regulated networks involved in genetic control of these complex traits. Therefore, for this study, we conducted QTX mapping to reveal the genetic architecture of two complex traits in tobacco leaves. Four -omics datasets were used as genotypes: genome methylation data for QTS, transcription and miRNA levels for QTT, enzyme and protein amounts for QTP, and metabolite levels for QTM. The genotypic and phenotypic data were collected from 60 tobacco leaf samples of three cultivars, from four development stages and three positions in the plants and grown in two cultivation environments. The potential application of the QTX results in breeding practice was also discussed.

2. Materials and methods

2.1. Plant growth and sample preparation

In 2010, three tobacco cultivars (K326, Hongda and Zunyan 6) were grown in farm fields at Guiding (26.58° N, 107.23° E) and Xingyi (25.08° N, 104.90° E) under normal condition for crop production in Guizhou province, China. The plants of three cultivars were planted in 10 rows with 30 plants per row and in three blocks. The plant-to-plant spaces between and within rows were 100 cm and 50 cm, respectively. For each of the three cultivars, leaves of 25 plants from 5 points were pooled for 1) the combination of upper or middle leaves from two locations within the plant at four developmental time points with sampling every 12 days, and 2) the lower leaves from two locations within the plant at two developmental time points, thus resulting in a total of 60 samples. The pooled leaves were immediately frozen in liquid nitrogen and stored at −80 °C for further use.

2.2. Methylation profiling and clone sequencing

A methylation DArT chip for tobacco was developed by Diversity Array Technology Ltd. (Canberra, Australia) as described at http://www.diversityarrays.com/dnamethylation.html. Total DNA was extracted and hybridization followed the DArT methylation profiling protocol as described by Lu et al. [23]. The program DArT Soft was used to determine whether the fragments in the arrays tested for each sample were methylated or not.

2.3. mRNA array design, hybridization and normalization

A custom-designed microarray platform was used for the analysis of total RNA extracted from the tobacco leaf samples. The microarray was comprised of three 60-mer probes for each of 44,873 unigenes derived from public Expressed Sequence Tags (ESTs) of tobacco and was made following a protocol provided by Roche Co. (http://www.nimblegen.com/).

The 60-mer probes were chosen from a group of six to seven non-overlapping probes designed against different parts of each gene model. The probes with E-values most similar to the average of the six to seven non-overlapping experimental probes were assumed to be the most reliable for transcript level estimation. Total RNA was extracted with the RNeasy Mini Kit (Qiagen Corp, Valencia, CA, United States) and DNase treated in-column with the RNase-Free DNase set (also from Qiagen). Double-stranded cDNA was synthesized using the SuperScript Double Strand cDNA Synthesis Kit (Invitrogen Inc., Carlsbad, CA, United States) with oligo (dT) primers following the manufacturer's protocol. Cy-3 and Cy-5 labeling and hybridization steps were performed by NimbleGen using standard procedures (http://www.nimblegen.com/). Expression values were generated by Roche NimbleGen proprietary software using quantile normalization [24] and the Robust Multichip Average algorithm [25].

2.4. *miRNA array design, hybridization and normalization*

A total of 1000 candidate tobacco miRNA genes were obtained from Guo et al. [26], which were identified by sequencing and advanced bioinformatics analysis of small fragment RNAs. These miRNAs were used to design the miRNA array based on Agilent miRNA chip technology. Total RNA was extracted using mirVanamiRNA Isolation Kit (Applied Biosystems/Ambion, Austin, TX, United States), and RNA concentrations were determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States). Following this, a total of 120 ng of total RNA was fluorescently labeled with Cyanine 3-pCp, and hybridized onto the arrays for 18–20 h at 55 °C. Slides were scanned by an Agilent microarray scanner G2565BA and the images obtained were processed with Feature Extraction Software 9.5.3.1 (also from Agilent). Intensity values were processed using Cluster 3.0 software whereby data were normalized, log transformed, and median centered [27]. Only normalized miRNAs with less than 20% missing values across the samples were included in the subsequent analyses.

2.5. *Metabolite quantified*

Content of Threonine, Lysine, Serine and Phenylalanine was quantified by HPLC (Waters 2695, Waters Alliance). Briefly, 1.0 g dry leaf powder was placed in 50 mL Erlenmeyer flask after sifting with a 40 mm mesh sieve. Totals of 200 μL of 0.1 mg mL^{-1} internal standard solution and 50 mL of ultrapure water were added, and then ultrasonic vibration was conducted for 60 min at room temperature. The resulting suspension was filtered through a 0.45 μm membrane filter. Subsequently, 50 μL of the filtrate was added to a hydrolysis tube, where it was combined with 70 μL AccQ-1 derivatization buffer solution. A shock treatment of 10 s of vigorous stirring using a vortex followed while 20 μL AccQ-2A amino acid derivatization reagent was added. An additional 10 s of shaking was needed after the first vortexing was finished. The extract was then placed in an oven for the full derivatization reaction at 55 °C for 10 min. The solution was then used for HPLC analysis.

Total sugar and fructose content was quantified spectrophotometrically with a Dionex ICS-2000+ED40. The fresh sample was ground in liquid nitrogen. An aliquot of 0.5 g of ground powder for each sample was then placed into 100-mL volumetric flasks each with 70 mL of deionized water added. Extraction by ultrasound was used for 1 h. The volume was set to the 100-mL mark and separated for 15 min under centrifugation at 9000 r min^{-1}. The supernatant was filtered using a membrane of 0.45 μm pore size (Tianjin Jinteng Experiment Equipment Co., Tianjin, China) to remove impurities, and then passed over a RP pre-treatment column to remove pigments and macromolecules. Finally 0.20 mL of the filtered liquid was taken, diluted to 10.0 mL, and passed through a second membrane of 0.22 μm pore size (Tianjin Jinteng Experiment Equipment Co., Tianjin, China), which the resulting effluent was analyzed. Peak area was quantified by software accompanied with the equipment.

Linolenic acid and linoleic acid were quantified with a gas chromatography (GC) method using an HP-6890 analyzer (Agilent Corp., Santa Clara, CA, United States) following the protocol provided by the vendor. Briefly, 1000 g mixed sample was taken in 100 mL conical flask, then 500 μL of an adipic acid methanol internal standard solution was added along with 25 mL of 10% H_2SO_4–CH_3OH solution. This mixture was shaken by mechanically oscillated overnight at low-speed for the derivatization reaction. The solution was then transferred into 250 mL pyriform separatory funnels with 50 mL distilled water added. The solution was extracted three times by gently shaking with 15 mL CH_2Cl_2, followed by collecting and placing the extract in a 100 mL conical flask with grinding stopper. The appropriate amount of anhydrous sodium sulfate was then added to remove trace water, and the clear and transparent extract was used for analysis.

Chlorogenic acid was quantified by high-pressure liquid chromatography (HPLC) using the LC-2010AHT from Shimadzu Corp. (Shimadzu Corp., Nakagyo-ku, Kyoto, Japan) and default protocol. Briefly, fresh leaf sample was ground in liquid nitrogen and a 0.5 g milled sample taken to a 5 mL centrifuge tube, where 1.5 mL of a 50% aqueous methanol solution was added before treating with ultrasound for 20 min at 56 kHz. The extract was then filtered with liquid membranes (0.22 μm) and stored in a bottle for further analysis.

Chromium content was quantified using microwave digestion and inductively coupled plasma optical emission spectrometry (ICP-OES). A 0.5 g sample was placed in the inner digestion tank of poly-tetra-fluoro-ethylene, which was itself put into an outer tank to which 4 mL of nitrate, 1 mL of hydrogen peroxide, and 0.5 mL of hydrofluoric were added. The sample was sequentially digested by the following procedures in the microwave workstation: digestion at 100 °C for 10 min, at 180 °C for 10 min, and at 220 °C for 20 min. When the digestion was completed, the tank was cooled down to room temperature, and the pressure was reduced to lower than 0.1 MPa. Then the digestion mixture was transferred into a 25 mL volumetric flask after adding 5 mL of boric acid solution, and the inner tank was washed with a small amount of ultrapure water several times, during which the cleaning liquid was merged into the digestion mixture until the final volume was topped up to the original volume. A blank test was performed simultaneously. The parameters of ICP-OES analysis were set as: RF generator transmission power of 1.2 kW; plasma gas flow of 15 L min^{-1}; auxiliary gas flow of 1.5 L min^{-1}; nebulizer pressure of 240 kPa;

and cleaning time of 20 s. Measurements were conducted 3 times at intervals of 10 s each. Meanwhile, the peristaltic pump speed was 15 r min^{-1} and a Fitted Model was used to correct for background.

2.6. Mixed linear model

We used the four -omics datasets to conduct QTX mapping. In order to predict genetic effects of QTXs along with epistasis and QTX × environment interaction, a mixed linear model approach was used by including treatments (three varieties in two locations) as fixed effects (*E*), compared to QTX (Q and QQ) and QTX × environment interaction (QE and QQE) as random effects, namely:

$$y = \mu + E + Q + QQ + QQE + \varepsilon \tag{1}$$

where, y = phenotypic value; μ = population mean; E = total treatment effect; Q = total individual loci effect; QQ = total epistasis effect between loci; QE and QQE = the total Q × environment interaction effect and total QQ × environment interaction effect, respectively; ε = residual effect.

The genetic model for the phenotypic value of the k-th genotypes in the h-th treatment (y_{hk}) can be expressed by the following mixed linear model,

$$\begin{aligned} y_{hk} = \mu + e_h + \sum_i q_i u_{ik} + \sum_{i<j} qq_{ij} u_{ik} u_{jk} + \sum_i qe_{hi} u_{hik} \\ + \sum_{i<j} qqe_{hij} u_{hik} u_{hjk} + \varepsilon_{hk} \end{aligned} \tag{2}$$

where, μ = the population mean; e_h = the fixed effect of the h-th treatment; q_i = the i-th locus effect with coefficient u_{ik}; qq_{ij} = the epistasis effect of the i-th locus by j-th locus with coefficient $u_{ik}u_{jk}$; qe_{hi} = the locus × treatment interaction effect of the i-th locus in the h-th treatment with coefficient u_{hik}; qqe_{hji} = the epistasis × treatment interaction effect of the i-th locus and j-th locus in the h-th treatment with coefficient $u_{hik}u_{hjk}$; and ε_{hk} = the random residual effect of the k-th breeding line in the h-th treatment.

The mixed linear model can be presented in matrix notation,

$$\begin{aligned} \mathbf{y} &= \mathbf{Xb} + \mathbf{U}_Q\mathbf{e}_Q + \mathbf{U}_{QQ}\mathbf{e}_{QQ} + \mathbf{U}_{QE}\mathbf{e}_{QE} + \mathbf{U}_{QQE}\mathbf{e}_{QQE} + \mathbf{e}_\varepsilon \\ &= \mathbf{Xb} + \sum_{v=1}^{4} \mathbf{U}_v\mathbf{e}_v + \mathbf{e}_\varepsilon \\ &\sim MVN\left(\mathbf{Xb}, \sum_{v=1}^{4} \sigma_v^2 \mathbf{U}_v\mathbf{U}_v^T + \mathbf{I}\sigma_\varepsilon^2\right) \end{aligned} \tag{3}$$

where $\mathbf{y}$ is an $n \times 1$ column vector of phenotypic values and n is the sample size of observations; $\mathbf{b}$ is a column vector of μ, treatments in the experiment; $\mathbf{X}$ is the known incidence matrix relating to the fixed effects; $\mathbf{U}_v$ is the known coefficient matrix relating to the v-th random vector $\mathbf{e}_v$; $\mathbf{e}_\varepsilon \sim MVN(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$ is an $n \times 1$ column vector of residual effects.

2.7. Prediction of genetic effects

The estimation of fixed effects (*e*) and prediction of random effects (*q*, *qq*, *qe* and *qqe*) were obtained using QTXNetwork software based on GPU parallel computation (http://ibi.zju.edu.cn/software/QTXNetwork/). By using mixed linear model approaches described in QTLNetwork 2.0 [28], association was conducted for complex traits against a panel of genetic markers for the QTS dataset, or quantitative expression of transcripts/proteins/metabolites for the QTT/P/M datasets, respectively.

2.8. Estimation of heritability

The total phenotypic variance was considered as the sum of genotype variance ($V_G = V_Q + V_{QQ}$), genotype × treatment interaction variance ($V_{GE} = V_{QE} + V_{QQE}$), and residual variance (V_ε):

$$\begin{aligned} V_P &= V_G + V_{GE} + V_\varepsilon \\ &= (V_Q + V_{QQ}) + (V_{QE} + V_{QQE}) + V_\varepsilon \\ &= \left(\frac{1}{df_Q}\sum_i q_i^2 + \frac{1}{df_{QQ}}\sum_{i<j} qq_{ij}^2\right) \\ &\quad + \left(\frac{1}{df_{QE}}\sum_i qe_{hi}^2 + \frac{1}{df_{QQE}}\sum_{i<j} qqe_{hij}^2\right) + V_\varepsilon. \end{aligned} \tag{4}$$

Individual heritability of specific genetic effects was estimated as $h_g^2 = g^2/V_P$. Total heritability was defined as the relative contribution of genetic variance to phenotypic variance with the following estimation model:

$$\begin{aligned} h_{G+GE}^2 &= h_Q^2 + h_{QQ}^2 + h_{QE}^2 + h_{QQE}^2 \\ &= \sum h_q^2 + \sum h_{qq}^2 + \sum h_{qe}^2 + \sum h_{qqe}^2 \end{aligned} \tag{5}$$

where h_{G+GE}^2 = total heritability; h_Q^2 = heritability contributed by sum of individual QTXs, h_{QQ}^2 = heritability contributed by sum of pair-wise epistasis (*qq*), h_{QE}^2 = treatment-specific heritability contributed by sum of individual treatment interaction (*qe*), h_{QQE}^2 = treatment-specific epistasis heritability contributed by sum of pair-wise epistasis by treatment interaction (*qqe*); h_q^2 = heritability of individual locus effect, h_{qq}^2 = heritability of individual epistasis effect, h_{qe}^2 = heritability of treatment-specific effect, h_{qqe}^2 = heritability of treatment-specific epistasis effect.

3. Results

3.1. QTS and QTT/P/M mapping

There were 567 methylated genome loci used for mapping QTSs of two tobacco leaf traits (chromium content and total sugar content) with 60 phenotypes obtained from harvested leaves at three positions and different time points for three varieties grown in two locations. The QTS module in QTXNetwork was applied for mapping significant QTSs by setting two varieties and three locations as six treatments for detecting treatment-specific genetic effects.

The QTT/P/M module in QTXNetwork was applied to screen for significant RNA transcripts and to predict their genetic effects. A total of 2894 mRNA transcripts and 802 miRNA transcripts were used for QTT mapping. Similarly, QTP and QTM were applied to search significant proteins and metabolites. The concentration of 14 amino acids was measured for QTP mapping and 35 metabolites for QTM mapping accordingly. For QTS, QTP and QTM search, a total of 60 observations in six treatments were collected. The raw data of expression of RNAs, proteins and metabolites were transformed by standardization $(y - \mu) / \sigma$ for association mapping.

3.2. Association mapping for chromium content based on -omics data

There were a total of nine QTX loci (three for QTSs, four for QTTs, one for QTP, and one for QTM) that were detected as controlling chromium content in tobacco leaves (Table 1, Figs. 1 and 2). Three treatment-specific epistatic effects were identified between two QTSs with no individual effects. Large treatment-specific additive effects were found for QTSs, QTTs and QTP.

In the case of QTS, there were three methylated SNP loci (DArT markers) detected with significant additive (q), additive × treatment interaction (qe), and epistasis × treatment interaction effect (qqe) (Table 1). Phm1376 had a very significant additive effect ($-\log_{10}P = 10.05$) and high heritability ($h_q^2 = 20.29\%$). The total additive × environment interaction had higher heritability ($h_{qqe}^2 = 30.09\%$). For the three varieties tested, treatment interaction effects of *Phm1376* were negative in Guiding, but positive in Xingyi. It was suggested that *Phm1376* could decrease chromium content in Guiding for all three varieties. *Phm1053* and *Phm1471* had epistasis × treatment interaction in the Xingyi with negative qqe effect only for cultivar Zunyan 6. It was indicated that the loci could have opposite impact on chromium content in tobacco leaves of a set of cultivars in different environments or various cultivars in the same environment.

For the analysis of QTT, there were two transcripts or unigenes and two miRNA detected for controlling chromium content of tobacco leaf (Table 1, Figs. 1 and 2). Unigene mRNA2713 had positive main effect (q) stable across environments and varieties. There was one miRNA (miRNA644) negatively associated with chromium content for two varieties in three environments. This QTT could also significantly decrease chromium content for K326 and Zunyan 6 in Xingyi. The data suggested that decreasing expression of mRNA2713 and increasing expression of miRNA644 could reduce chromium content in different varieties and environments. An epistasis between miQNA445 and mRNA1119 was detected for increasing chromium content stably across environments in three varieties. These results indicated that suppression of epistasis could significantly reduce chromium content in tobacco leaf.

In QTP mapping, only one amino acid (Threonine) was detected with significantly negative main effect (q) and treatment-specific effect (qe) in Xingyi for Zunyan 6 (Table 1, Figs. 1 and 2). It was suggested that threonine might actively participate in chromium degradation or be a byproduct of such activity when certain other factors presenting. Meanwhile, there was only one significant ($-\log_{10}P = 5.22$) QTM (Triacontanol) which was identified with a positive main effect (q) on chromium content and high heritability ($h_q^2 = 21.94\%$) (Table 1, Figs. 1 and 2). This suggested that decreasing triacontanol content could reduce chromium content in tobacco.

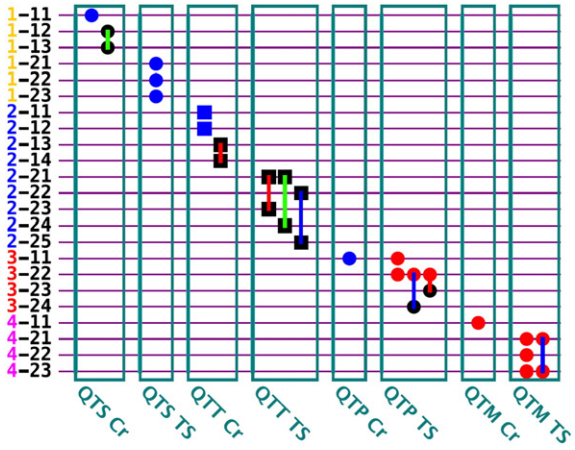


Fig. 1 – Network plot of detected QTXs for chromium content (Cr) and total sugar (TS) in tobacco leaves. Circle: QTX individual effect; Line between two QTXs: epistasis effects; Red: general effects for six environments; Green: environment-specific effects; Blue: general and environment-specific effects; Black: epistasis effects but without detected individual effects.

Table 1 – Predicted QTX effects, QTX × environment interaction with significance and heritability for chromium content in tobacco leaf of three cultivars in two locations.

Type	QTX[a]	Effect[b]	Predict	$-\log_{10}P$	h^2 (%)
QTS	Phm1376	q	0.43	10.05	20.29
		qe_1v_1	−0.43	2.18	30.09
		qe_1v_2	−0.37	1.72	
		qe_1v_3	−0.41	2.00	
		qe_2v_1	0.45	2.38	
		qe_2v_2	0.60	3.78	
		qe_2v_3	0.78	6.09	
	Phm1053 × Phm1471	qqe_2v_1	0.41	2.09	11.24
		qqe_2v_2	0.47	2.55	
		qqe_2v_3	−0.48	2.65	
QTT	mRNA2713	q	0.17	3.17	3.87
		qe_2v_1	0.37	5.11	24.62
		qe_2v_2	0.96	14.77	
	miRNA644	q	−0.25	6.35	8.53
		qe_2v_1	−0.47	5.79	16.29
		qe_2v_3	−0.69	6.57	
	mRNA1119 × miRNA445	qq	0.43	17.75	26.25
QTP	Threonine	q	−0.25	1.88	6.09
		qe_2v_3	−0.63	2.69	6.21
QTM	Triacontanol	q	0.51	5.22	21.94

[a] mRNA2713: GenBank accession number FH405569; mRNA1119: GenBank accession number FH569168; miRNA644: sequence 5′-GCTTATCCATATTTGACCCGT TTTT-3′; miRNA445: sequence 5′-GAGCACGTACCCTGCTTCTCCA-3′.

[b] e_1: Guiding; e_2: Xingyi; v_1: K326; v_2: Hongda; v_3: Zunyan 6.

3.3. Association mapping for total sugar content based on -omics data

Twelve loci were identified with the QTX module for association analysis between total sugar content (TS) and the four -omics datasets (three QTSs for methylated genes, five QTTs for mRNAs or miRNAs, four QTPs for amino acids, and three QTMs for metabolite elements (Table 2, Figs. 1 and 2)).

Three methylated genome loci with significant additive (q) effects and additive × treatment interaction (qe) were detected (Table 2, Figs. 1 and 2). The locus Phm1132 had highly

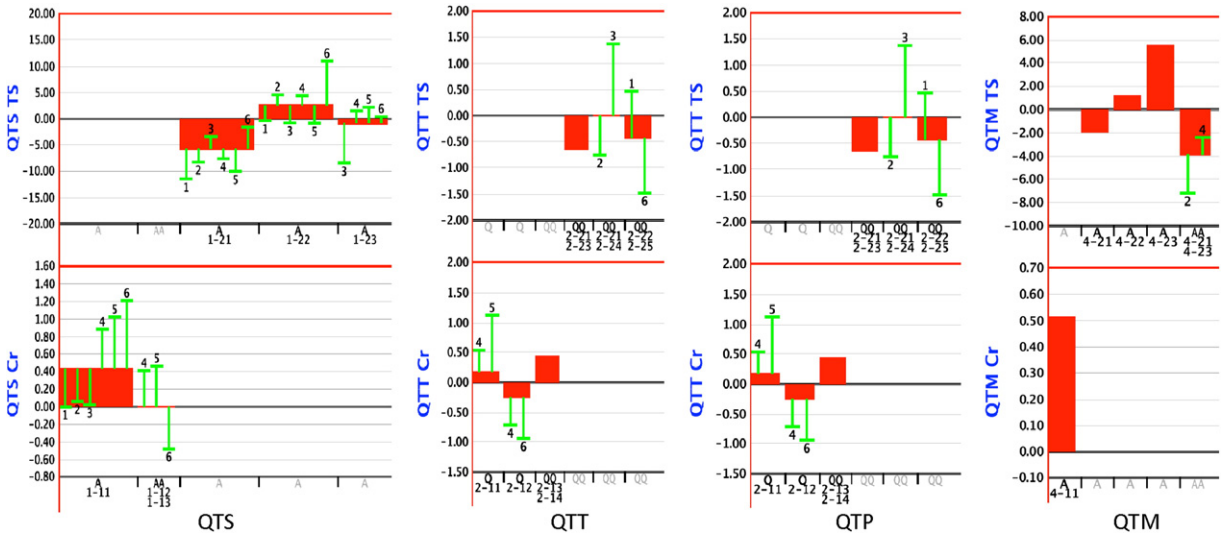


Fig. 2 – G × E plot of detected QTXs for chromium content (Cr) and total sugar (TS) in tobacco leaves. Red columns: general QTX effects for all six environments; Green lanes: environment-specific effects for 1 = e_1v_1, 2 = e_1v_2, 3 = e_1v_3, 4 = e_2v_1, 5 = e_2v_2, 6 = e_2v_3.

significant positive main effect ($-\log_{10}P = 231.75$) with high heritability ($h_q^2 = 37.34\%$). For additive × treatment interaction, the Zunyan 6 had positive effects in the two locations, while the other two varieties had negative effects. It was suggested that *Phm1132* of Zunyan 6 could increase total sugar content across different environments. Though *Phm1227* was reported with significantly positive *a* effect ($-\log_{10}P = 47.12$), its heritability ($h_q^2 = 7.28\%$) was not as large as *qe* interaction effects ($h_{qe}^2 = 21.07\%$). Three cultivars tended to have opposite effects in the two locations. Zunyan 6 had quite large positive *qe* in the second location ($-\log_{10}P = 88.43$) while the other two cultivars did not. *Phm1298* had positive *qe* interaction for three cultivars in Xingyi, but significantly negative *qe* interaction for the Zunyan 6 variety in Guiding ($-\log_{10}P = 71.61$).

Three loci with epistasis between mRNA and miRNA genes also controlled total sugar content. mRNA537 had negative main epistasis ($h_q^2 = 22.87\%$) jointly with miRNA652, and treatment-specific epistasis ($h_{qqe}^2 = 23.94\%$) in Guiding (negative for Hongda and positive for Zunyan 6) together with miRNA775. mRNA1218 × miRNA183 had negative main epistasis ($h_q^2 = 10.44\%$) and treatment-specific epistasis ($h_{qqe}^2 = 18.44\%$) in Xingyi for Zunyan 6. Therefore, epistasis might be useful as an efficient genetic tool for increasing total sugar content in tobacco leaf.

In QTP mapping, lysine was detected to have a large individual negative main effect (*q*) on total sugar content in tobacco leaves ($-\log_{10}P = 62.55$ and h_{qq}^2 46.90%), but positive epistasis effects (*qq*) along with phenylalanine ($-\log_{10}P = 53.47$ and $h_{qq}^2 = 33.27\%$) (Table 2, Figs. 1 and 2). Meanwhile, for QTM mapping, fructose was detected with large positive individual effects (*q*) ($-\log_{10}P = 80.45$ and h_{qq}^2 52.30%), while linolenic and linoleic acids had lower negative individual effects (*q*) ($-\log_{10}P = 13.20$ and h_{qq}^2 6.22%) (Table 2, Figs. 1 and 2). Epistasis effects of these two QTMs were also significant ($-\log_{10}P = 38.29$ and h_{qq}^2 26.02%).

4. Discussion

4.1. *QTXNetwork can reveal genetic architecture of complex traits at four -omics levels*

The principal feat of this research was to implement QTXNetwork, a software program based on a mixed linear model, for analysis of -omics data. This research was able to take advantage of an abundance of data on gene methylation, transcript expression, protein content and metabolite characterization to find associations of QTS, QTT, QTP and QTM with two complex traits. Our goal in these analyses was to directly estimate the genetic effects of each type of loci on the genetic architecture of these traits. We believe this to be the first time that these new methods have been used to detect genome methylated loci, transcripts, proteins and metabolites associated with chromium content and total sugar content in tobacco leaves. The results showed that various types of genetic effects contributed to the two traits at different levels of -omics data, but that the composition and proportion of each type varied among -omics levels (Tables 1 and 2). For example we observed that total heritability increased consecutively for genomic, transcriptomic, proteomic and metabolomic loci, which was consistent with the central genetic dogma of gene expression through transcripts and their resulting proteins and metabolites in the transfer of genetic information to phenotype.

4.2. *Epistasis and treatment interaction may play an important role in complex traits*

Another discovery of this study was that the proportion of total heritability of epistasis and treatment interaction was very significant in the combination of trait and -omic evaluation, and

Table 2 – Predicted QTX effects and QTX × environment interaction with significance and heritability for total sugar content in tobacco leaf of three cultivars in two locations.

Type	QTX[a]	Effect[b]	Predict	$-\log_{10}P$	h^2 (%)
QTS	Phm1132	q	5.68	231.75	37.34
		qe_1v_1	−5.79	42.16	16.27
		qe_1v_2	−2.55	8.82	
		qe_1v_3	2.31	7.41	
		qe_2v_1	−1.89	5.16	
		qe_2v_2	−4.32	23.93	
		qe_2v_3	4.10	21.68	
	Phm1227	q	2.51	47.12	7.28
		qe_1v_1	−2.85	10.89	21.07
		qe_1v_2	2.06	5.99	
		qe_1v_3	−3.26	14.03	
		qe_2v_1	1.86	4.98	
		qe_2v_2	−3.35	14.75	
		qe_2v_3	8.47	88.43	
	Phm1298	q	−0.85	6.10	0.84
		qe_1v_3	−7.59	71.61	14.31
		qe_2v_1	2.36	7.67	
		qe_2v_2	3.09	12.69	
		qe_2v_3	1.28	2.64	
QTT	mRNA537 × miRNA652	qq	−0.63	9.90	22.87
	mRNA537 ×	qqe_1v_2	−0.76	2.73	23.94
	miRNA775	qqe_1v_3	1.37	8.03	
	mRNA1218 ×	qq	−0.42	4.97	10.44
	miRNA183	qqe_1v_1	0.89	2.68	18.44
		qqe_2v_3	−1.05	3.27	
QTP	Threonine	q	0.97	1.63	0.85
	Lysine	q	−7.24	62.55	46.9
	Lysine × serine	qq	−2.19	7.40	4.31
	Lysine ×	qq	6.10	53.47	33.27
	phenylalanine	qqe_2v_2	−2.16	4.21	0.7
QTM	Fructose	q	5.47	80.45	52.3
	Linolenic acid + linoleic acid	q	−1.89	13.20	6.22
	Chlorogenic acid	q	1.15	5.32	2.31
	Fructose × linolenic	qq	−3.86	38.29	26.02
	acid + linoleic acid	qqe_1v_2	−3.33	3.85	3.84
		qqe_2v_1	1.46	1.91	

[a] mRNA537: GenBank accession number CS224885; mRNA1218: GenBank accession number FH198617; miRNA652: sequence 5′-GGACCGTATACAGATAGTCCCCTC-3′; miRNA775: sequence 5′-GTCATTTTCCCTCCATTTCCCATTC-3′; miRNA183: sequence 5′-GTCTTTGCTACTTCAGCCCCGTAT-3′.

[b] e_1: Guiding; e_2: Xingyi; v_1: K326; v_2: Hongda; v_3: Zunyan 6.

that the total proportion of heritability based on epistasis and treatment interaction was nearly equal to that of the main factors. There was one QTS epistasis detected only in location 2 (h^2_{qqe} 11.24%) for chromium content among the four -omics levels. The proportions of total treatment interaction (h^2_{qe+qqe}) were 35.97%, 20.46%, 0.70% and 3.84% in genomic, transcriptomic, proteomic, and metabolic levels, respectively. In the case of total sugar content, the proportions of epistasis (h^2_{qq}) were 37.58% and 26.02% for proteome and metabolism. However, no epistasis was detected for genome and transcriptome loci. In contrast, the proportions of epistasis and treatment interaction effects on heritability (h^2_{qe+qqe}) were 51.65%, but only 0.70% and 3.84% at the transcriptome, proteome and metabolome levels, respectively.

4.3. *Application potential of detected QTXs in crop breeding*

Molecular markers have enormous potential to improve the efficiency and precision of conventional plant breeding via marker-assisted selection [29]. The important challenge of applying genetic and -omics data to breeding is the identification of the genes underlying a trait of interest. We performed an integrated association mapping for chromium content and total sugar content based on genome, transcriptome, proteome and metabolites, and detected some QTSs, QTTs, QTPs and QTMs associated with two complex traits. The strategy to employ these molecular loci in the breeding practice should be considered prudently. For example, those QTX based on methylated loci of the genome were essentially directly useful as DNA markers and would be directly applicable in breeding practice.

In terms of marker assisted breeding for each of the two traits, chromium content could be selected with *Phm1376* which had a significant positive additive effect ($-\log_{10}P$ = 10.05, and h^2_q = 20.29), indicating that demethylation of this locus could reduce chromium content for three varieties in two locations. The *qe* (additive by treatment interaction) effect of *Phm1376* was negative in Guiding for all three varieties tested in this study, but in Xingyi they were positive. This suggests that reduction of chromium content in tobacco leaves could be achieved by methylation of this locus in Guiding but demethylation of the same locus in Xingyi. The epistasis of Phm1053 and Phm1471 only had *qqe* effects in Xingyi, positive for K326 and Hongda, but negative for Zunyan 6, indicating that the best genotypic state in the two loci was demethylation for varieties K326 and Hongda, but methylation for Zunyan 6.

In the cases of the QTTs and QTPs associated with the traits of interest, there might be two strategies to use in practical plant breeding. For the first strategy, bioinformatics analysis can be carried out to make sure that the function of the transcript or proteins is based on a functional gene association with the investigated traits, and then the representative gene of the transcript can be developed as a DNA based marker useful for marker-assisted-selection. This strategy is more valuable when the transcripts or proteins have large genetic effects and high heritability. The second strategy would be based on direct use of the transcript as an indicative marker, where the abundance of the transcript would predict the performance of the genotype for the traits. In this study, two transcripts and two miRNAs presented association with chromium content. The transcripts representing mRNA2713 (GenBank accession number FH405569) presented a stable positive main genetic effect (0.17) for the three varieties in two locations, with one unit reduction of transcript expression resulting in 0.17 units of chromium reduction. Another pair of transcripts, namely mRNA1119 (GenBank accession number FH569168) and miRNA445 (sequence: 5′-GAGCACGTACCCTGCTTCTCCA-3′), presented a high and positive interaction effect (0.43) with moderate heritability (26.25%) and no environment interaction effect, indicating that they can be used as markers for breeding without concern for the specific growing environment. Another smRNA based locus, namely miRNA644 (sequence: 5′-GCTTATCCATATTTGACCCGTTTTT-3′) showed a moderate negative main genetic effect (−0.25) and heritability (8.53%), but presented a higher negative environment

interaction effect (−0.58 on average) and heritability (16.29%) in Xingyi, which indicated that this marker would be an environment-specific indicator of chromium content.

Finally, some metabolites had significant impacts on trait inheritance. For example, Lysine was detected with large individual negative main effects on total sugar content, but positive epistasis effects on this trait in combination with phenylalanine (Table 2). This indicated that high concentration of individual lysine could reduce the concentration of total sugar content in tobacco leaves. One explanation for this observation could be based on the Maillard reaction in vivo [30], which is the result of a chemical reaction between amino acids and reducing sugar. But when the two kinds of amino acids (lysine and phenylalanine) worked together as a pair, the joint effects (*qq*) were positive on total sugar content. Further study is required to confirm this and other associations of the two traits with the metabolomic, proteomic, transcriptomic and genome methylation datasets. Furthermore, the same kind of analysis could be used for additional traits that are of complex inheritance but for which biochemical (mRNA, protein and metabolite) analysis is very salient and as important as just a genome wide association test with random DNA markers.

Acknowledgments

This study was partially supported by the National Basic Research Program of China (2011CB109306 and 2009CB118404), the Program of Introducing Talents of Discipline to Universities of China ("111" Project, B06014), and Research Programs (CNTC-D2011100, CNTC-[2012]146, NY-[2011]3047, QKHRZ [2013] 02). We are grateful for editing from Dr. M. W. Blair and the suggestions of two anonymous reviewers.

REFERENCES

[1] T.F. Mackay, The genetic architecture of quantitative traits, Annu. Rev. Genet. 35 (2001) 303–339.
[2] L.X. Ku, W.M. Zhao, J. Zhang, L.C. Wu, C.L. Wang, P.A. Wang, W.Q. Zhang, Y.H. Chen, Quantitative trait loci mapping of leaf angle and leaf orientation value in maize (*Zea mays* L.), Theor. Appl. Genet. 121 (2010) 951–959.
[3] Z. Wang, X. Wu, Q. Ren, X. Chang, R. Li, R. Jing, QTL mapping for developmental behavior of plant height in wheat (*Triticum aestivum* L.), Euphytica 174 (2010) 447–458.
[4] P. Langridge, D. Fleury, Making the most of 'omics' for crop breeding, Trends Biotechnol. 29 (2011) 33–40.
[5] T.F. Mackay, E.A. Stone, J.F. Ayroles, The genetics of quantitative traits: challenges and prospects, Nat. Rev. Genet. 10 (2009) 565–577.
[6] P. Martinez-Gomez, R. Sanchez-Perez, M. Rubio, Clarifying omics concepts, challenges, and opportunities for *Prunus* breeding in the postgenomic era, OMICS 16 (2012) 268–283.
[7] J.M. Thornsberry, M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, E.T. Buckler, *Dwarf8* polymorphisms associate with variation in flowering time, Nat. Genet. 28 (2001) 286–289.
[8] X. Huang, X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q.F. Zhang, J. Li, B. Han, Genome-wide association studies of 14 agronomic traits in rice landraces, Nat. Genet. 42 (2010) 961–967.
[9] J.A. Rafalski, Association genetics in crop improvement, Curr. Opin. Plant Biol. 13 (2010) 174–180.
[10] F. Tian, P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland, E.S. Buckler, Genome-wide association study of leaf architecture in the maize nested association mapping population, Nat. Genet. 43 (2011) 159–162.
[11] O. Zuk, E. Hechter, S.R. Sunyaev, E.S. Lander, The mystery of missing heritability: genetic interactions create phantom heritability, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 1193–1198.
[12] L.S. Yung, C. Yang, X. Wan, W.C. Yu, GBOOST: a GPU-based tool for detecting gene–gene interactions in genome-wide case control studies, Bioinformatics 27 (2011) 1309–1310.
[13] G.K. Chen, K. Wang, A.H. Stram, E.M. Sobel, K. Lange, Mendel-GPU: haplotyping and genotype imputation on Graphics Processing Units, Bioinformatics 28 (2012) 2979–2980.
[14] P. Winterhalter, R.L. Rouseff, Carotenoid-derived Aroma Compounds, An American Chemical Society Publication, 2002.
[15] A. Bombarely, K.D. Edwards, J. Sanchez-Tamburrino, L.A. Mueller, Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective, BMC Genomics 13 (2012) 406.
[16] B. Lei, X.H. Zhao, K. Zhang, J. Zhang, W. Ren, Z. Ren, Y. Chen, H.N. Zhao, W.J. Pan, W. Chen, H.X. Li, W.Y. Deng, F.Z. Ding, K. Lu, Comparative transcriptome analysis of tobacco (*Nicotiana tabacum*) leaves to identify aroma compound-related genes expressed in different cultivated regions, Mol. Biol. Rep. 40 (2013) 345–357.
[17] B.G. Xiao, X.P. Lu, F.C. Jiao, Y.P. Li, Y.H. Sun, Z.K. Guo, Preliminary QTL analysis of several chemical components in flue-cured tobacco (*Nicotiana tabacum* L.), Acta Agron. Sin. 34 (2008) 341–349.
[18] C. Goulet, C. Goulet, M.C. Goulet, D. Michaud, 2-DE proteome maps for the leaf apoplast of *Nicotiana benthamiana*, Proteomics 10 (2010) 2536–2544.
[19] E. Julio, B. Denoyes-Rothan, J.L. Verrier, F.D. de Borne, Detection of QTLs linked to leaf and smoke properties in *Nicotiana tabacum* based on a study of 114 recombinant inbred lines, Mol. Breed. 1 (2006) 69–91.
[20] G. Bindler, J. Plieske, N. Bakaher, I. Gunduz, N. Ivanov, R. Van der Hoeven, M. Ganal, P. Donini, A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development, Theor. Appl. Genet. 123 (2011) 219–230.
[21] H. Cui, S.T. Zhang, H.J. Yang, H. Ji, X.J. Wang, Gene expression profile analysis of tobacco leaf trichomes, BMC Plant Biol. 11 (2011) 76.
[22] W.Q. Lui, H.X. Guo, H. Li, Proteomics identification of differentially expressed proteins relevant for nicotine synthesis in flue-cured tobacco roots before and after decapitation, Agric. Sci. China 9 (2008) 1084–1090.
[23] X.P. Lu, Y.J. Gui, B.G. Xiao, Y.P. Li, Z.J. Tong, Y. Liu, X.F. Bai, W.R. Wu, L. Xia, E. Huttner, A. Kilian, L.J. Fan, Development of DArT markers for a linkage map of flue-cured tobacco, Chin. Sci. Bull. 58 (2012) 641–648.
[24] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.
[25] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, T.P. Speed, Exploration, normalization, and summaries of high density oligonucleotide array probe level data, Biostatistics 4 (2003) 249–264.
[26] Y. Guo, H. Liu, Z. Yang, J. Chen, Y. Sun, X. Ren, Identification and characterization of miRNAome in tobacco (*Nicotiana tabacum*) by deep sequencing combined with microarray, Gene 501 (2012) 24–32.

[27] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 14863–14868.
[28] J. Yang, J. Zhu, R.W. Williams, Mapping the genetic architecture of complex traits in experimental populations, Bioinformatics 23 (2007) 1527–1536.
[29] B.C. Collard, D.J. Mackill, Marker-assisted selection: an approach for precision plant breeding in the twenty-first century, Philos. Trans. R. Soc. Lond. B Biol. Sci. 363 (2008) 557–572.
[30] H. Nersten, The Maillard Reaction: Chemistry, Biochemistry and Implications, The University of Reading, Reading, UK, 2005.