Journal of Advanced Research 37 (2022) 43–60

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

A *de novo* regulation design shows an effectiveness in altering plant secondary metabolism

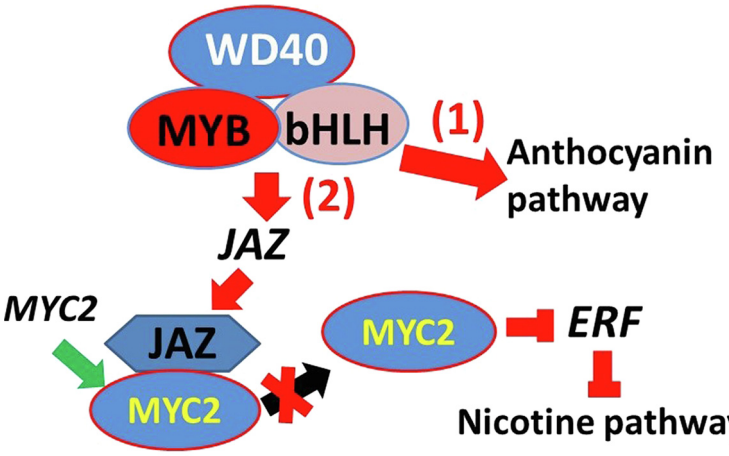
Mingzhuo Li [1], Xianzhi He [1], Christophe La Hovary [1], Yue Zhu, Yilun Dong, Shibiao Liu, Hucheng Xing, Yajun Liu, Yucheng Jie, Dongming Ma, Seyit Yuzuak, De-Yu Xie *

Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA

HIGHLIGHTS

- Arabidopsis PAP1 and TT8 transcription factors and cis-regulatory elements of tobacco *JAZs* are for the first time selected as molecular tools for a *De Novo* regulation design.
- This design creates a distant pathway-cross regulation (DPCR) that effectively downregulates the alkaloid biosynthesis in tobacco.
- The DPCR significantly reduces carcinogenic compounds.
- This design unearths two novel regulatory functions of PAP1 and TT8 and their complex and a new regulation mechanism of four *NtJAZs*.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 March 2021
Revised 26 May 2021
Accepted 17 June 2021
Available online 20 June 2021

Keywords:
Arabidopsis
Alkaloid
Anthocyanin
Molecular tools
Tobacco
Transcription factor

ABSTRACT

Introduction: Transcription factors (TFs) and *cis*-regulatory elements (CREs) control gene transcripts involved in various biological processes. We hypothesize that TFs and CREs can be effective molecular tools for *De Novo* regulation designs to engineer plants.
Objectives: We selected two Arabidopsis TF types and two tobacco CRE types to design a *De Novo* regulation and evaluated its effectiveness in plant engineering.
Methods: G-box and MYB recognition elements (MREs) were identified in four *Nicotiana tabacum JAZs* (*NtJAZs*) promoters. MRE-like and G-box like elements were identified in one nicotine pathway gene promoter. TF screening led to select Arabidopsis Production of Anthocyanin Pigment 1 (PAP1/MYB) and Transparent Testa 8 (TT8/bHLH). Two *NtJAZ* and two nicotine pathway gene promoters were cloned from commercial Narrow Leaf Madole (NL) and KY171 (KY) tobacco cultivars. Electrophoretic mobility shift assay (EMSA), cross-linked chromatin immunoprecipitation (ChIP), and dual-luciferase assays were performed to test the promoter binding and activation by PAP1 (P), TT8 (T), PAP1/TT8 together, and the PAP1/TT8/Transparent Testa Glabra 1 (TTG1) complex. A DNA cassette was designed and then synthesized for stacking and expressing PAP1 and TT8 together. Three years of field trials were performed by following industrial and GMO protocols. Gene expression and metabolic profiling were completed to characterize plant secondary metabolism.

Peer review under responsibility of Cairo University.
* Corresponding author.
E-mail address: dxie@ncsu.edu (D.-Y. Xie).
[1] Contributed equally (M.L., X.H, and C.L.H.).

https://doi.org/10.1016/j.jare.2021.06.017

Results: PAP1, TT8, PAP1/TT8, and the PAP1/TT8/TTG1 complex bound to and activated *NtJAZ* promoters but did not bind to nicotine pathway gene promoters. The engineered red P + T plants significantly upregulated four *NtJAZs* but downregulated the tobacco alkaloid biosynthesis. Field trials showed significant reduction of five tobacco alkaloids and four carcinogenic tobacco specific nitrosamines in most or all cured leaves of engineered P + T and PAP1 genotypes.
Conclusion: G-boxes, MREs, and two TF types are appropriate molecular tools for a *De Novo* regulation design to create a novel distant-pathway cross regulation for altering plant secondary metabolism.

Introduction

Plant MYB and bHLH proteins are two large TF families with diverse regulation functions involved in various biological processes [1–4]. To date, TFs that regulate the anthocyanin biosynthesis have gained a comprehensive characterization in both model and non-model plants. MYB and bHLH TFs together with a WD40 protein have been demonstrated to form MBW complexes to regulate the biosynthesis of anthocyanins [5–7]. In particular, the MBW complex formed by Arabidopsis Production of Anthocyanin Pigment 1 (PAP1), Transparent Testa 8 (TT8), and Transparent Testa Glabra 1 (TTG1) has been characterized to be a main regulatory complex. *PAP1*, *TT8*, and *TTG1* encode a R2R3-MYB TF (namely MYB75) [8], bHLH TF (namely bHLH42) [9], and a WD40 protein [10], respectively. Moreover, the functions of these three genes in Arabidopsis and their homologs in other plants have been intensively characterized in genetics and biochemistry [11]. PAP1 and TT8 bind to MYB recognizing elements (MREs) and G-box elements in promoters of anthocyanin pathway genes (such as *DFR*, *ANS* and *3-GT*, Fig. 1A), respectively [12–14]. The current dogma is that on the one hand, PAP1, TT8, and TTG1 form a master regulatory MYB-bHLH-WD40 (MBW) complex to only activate the anthocyanin biosynthesis in plants (Fig. 1 A) [6,15–17]; on the other hand, Arabidopsis JAZ repressors can interact with PAP1, TT8, and their MBW complex to negatively regulate the anthocyanin biosynthesis in the absence of jasmonate (JA) [18,19]. Moreover, *PAP1* has been used to create red crop varieties to engineer anthocyanins, such as tobacco [20,21], tomato [22,23], hop [24], and canola [25]. All these studies reported that PAP1 alone could activate the anthocyanin biosynthesis in different tested plants. However, whether PAP1, its partner TT8, and their homologs can regulate other plant secondary metabolisms remains open for studies.

Tobacco is an ideal model plant to study plant secondary metabolism, given that it produces thousands of metabolites in the alkaloid, phenolic, and terpenoid families [21,26–28]. Nicotine is the main alkaloid of tobacco (*Nicotiana tabacum*) and harmfully addictive [29–36]. Other harmful tobacco alkaloids include nornicotine, anabasine, anatabine, and myosmine (Fig. 1B). Tobacco specific nitrosamines (TSNAs) are carcinogenic compounds in tobacco smoke. Five common TSNAs are N-nitrosonornicotine (NNN), nicotine-derived nitrosamine ketone (NNK), N'-nitrosoanatabine (NAT), N-nitrosoanabasine (NAB), and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) [37–39]. Given that NNN is a severe carcinogen, the Food and Drug Administration has proposed a guideline that the average level of NNN must be less than 1.0 ppm (dry weight) in all final smokeless tobacco products [40].

To reduce these harmful compounds, numerous achievements have been accomplished in elucidating the biosynthesis of nicotine and other tobacco alkaloids [41–47]. Nicotine is biosynthesized from two distinct pathways (Fig. 1B), the steps of which are catalyzed by enzymes encoded by 10 genes *AO*, *QS*, *QPT2*, *ODC*, *ADC*, *AIH*, *PMT*, *MPO*, *A622* and *BBL*, (Fig. 1B) [42,48,49]. In addition, *CYP82E4* and its homologs have been characterized to involve the demethylation of nicotine to form nornicotine [50–52]. The mechanisms of the anatabine and anabasine formation are unknown [42]. One hypothesis is that these two alkaloids result from the BBL catalysis of nicotinic acid. The other is that anabasine is derived from lysine, a different pathway (Fig. 1B) [42]. Four main TSNAs, NNK, NNN, NAT, and NAB (Fig. 1B), result from the nitrosation of nicotine, nornicotine, anatabine, and anabasine during the leaf curing process, respectively [42]. Meanwhile, the regulation of the nicotine pathway has gained an intensive elucidation. Myelocytomatosis oncogene (MYC2) and APETALA 2/ethylene response factor (AP2/ERF) of tobacco are two groups of transcription factors (TFs) that directly bind to promoters of nicotine pathway genes [42,48,53–55]. Moreover, jasmonate (JA), a plant hormone, essentially controls the nicotine biosynthesis (Fig. 1B) [56–59]. JA regulates the nicotine biosynthesis via a co-receptor complex consisting of coronatine insensitive 1 (COI1), a Skp/Cullin/F-box (SCF^{COI1}) complex, and JA ZIM-domain (JAZ) repressor [53–55,60]. The tobacco genome has been reported to have 15 NtJAZs, in which NtJAZ1, NtJAZ3, NtJAZ7, and NtJAZ10 are associated with the regulation of the nicotine biosynthesis [54,60,61]. When tobacco roots lack isoleucine-JA (JA-Ile), NtJAZ1 and NtJAZ3 bind to NtMYC2 and then turn off the nicotine biosynthesis; when JA-Ile exists, it induces the interaction of SCF^{COI1} and JAZ to form a complex, which leads to the ubiquitination of JAZs to release NtMYC2. The freed NtMYC2 then activates the biosynthetic pathway of nicotine (Fig. 1B) [54,55,60]. Compared with the appropriate characterization of NtJAZs' functions, how their transcription in tobacco is regulated remains open for investigation. To date, based on those pathway and regulatory genes, on the one hand, antisense and RNAi have been reported to down-regulate gene expression (such as *BBL*, *ODC*, *QPT*, and nicotine demethylase gene) to reduce nicotine, nornicotine, or TSNAs in leaves [49,55,62,63]. On the other hand, these multiple successes revealed a challenge that the reduction of nicotine was traded with the increase of anatabine and other alkaloids [64,65]. These successful engineering data indicate that continuous efforts are necessary to diminish TSNAs, nicotine, and other alkaloids simultaneously.

Since 2000 when more than 1500 TFs were annotated in the genome of *Arabidopsis thaliana* [66], 320,370 TFs had been identified from 156 sequenced plant genomes and sequences of nine additional plants by 2017 [67]. Likewise, 21,997,501 TF binding sites (motifs or regulatory elements) (TFBSs) were determined in 63 plant genomes and nearly 2,493,577 TFBSs (*cis*-regulatory elements) were localized in the promoters [68]. Given that the interaction of TFs and TFBSs regulates plant development, growth, responses to stresses, and metabolisms, we hypothesize that the large numbers of *cis*-regulatory elements and TFs are useful molecular tools for *de novo* designs to create novel regulations for plant engineering. For proof of concept, herein, we selected tobacco to test our hypothesis, given that as described above, the success might enhance diminishing harmful nicotine, other tobacco alkaloids (OTAs), and TSNAs for human health. As introduced, given that *NtJAZ1* and *NtJAZ3* encode two main repressors of the nicotine biosynthesis, we hypothesized that their *cis*-regulatory elements were appropriate molecular tools to screen other plant TFs for *de novo* regulation designs and positive designs could be tested for

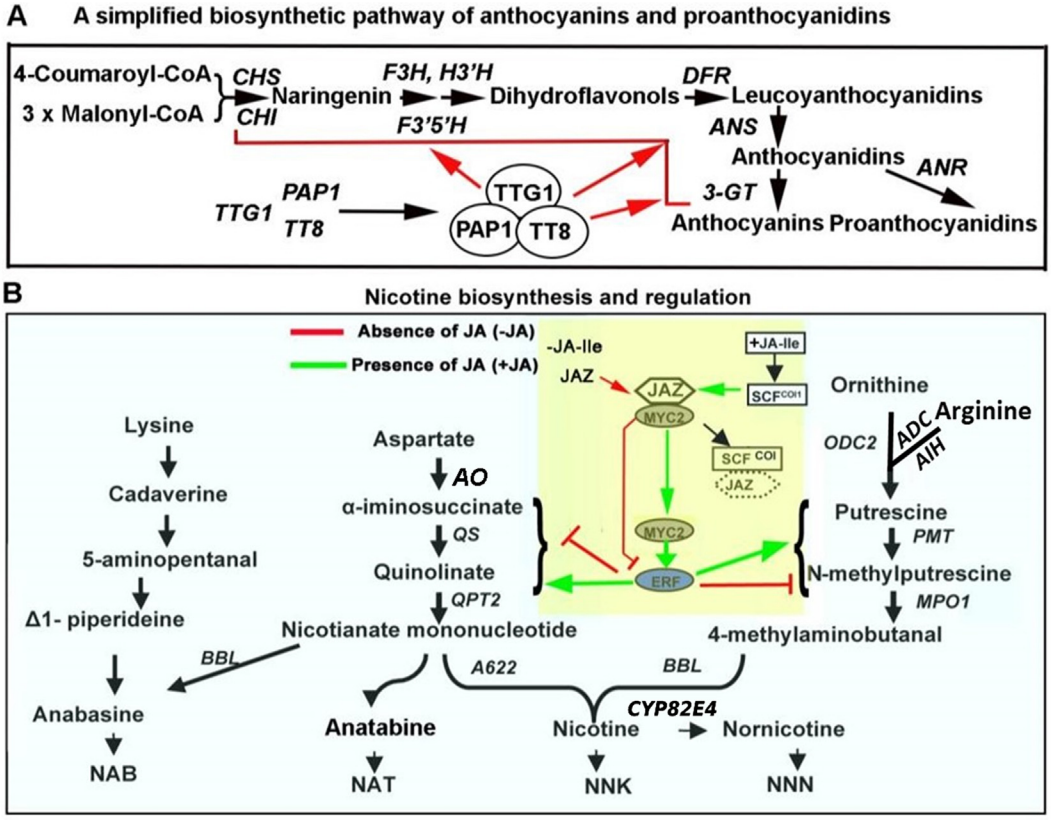


Fig. 1. Biosynthetic pathways of nicotine, other tobacco alkaloids, and anthocyanin and their regulations. A, the biosynthetic pathway of Arabidopsis anthocyanin from the step catalyzed by CHS. *PAP1*, *TT8*, and *TTG1* encode MYB75, bHLH, and WD40 to form a MYB-bHLH-WD40 complex that activates late pathway genes. Gene abbreviations are *CHS* and *CHI*: chalcone synthase and isomerase, *F3H*: flavonone-3-hydroxylase, *F3′H* and *F3′5′H*: flavonoid-3′-hydroxylase and 3′5′hydroxylase, *DFR*: dihydroflavonol reductase, *ANS*: anthocyanidin synthase, *3-GT*: glycosyltransferase. *ANR* encodes anthocyanidin reductase, a key enzyme toward proanthocyanidin biosynthesis. **B,** the biosynthetic pathway of nicotine, nornicotine, and tobacco specific nitrosamines (TSNAs) and the regulation mechanism. MYC2 and ERF are two positive transcription factors that activate the expression of most pathway genes. In the absence of isoleucine-jasmonate (Ile-JA), NtJAZ1 and NtJAZ3 repressors bind to MYC2 to inhibit the activation of the pathway genes (red lines). In the presence of bioactive Ile-JA (green arrows), SCF^{COI1} perceives this plant hormone to form a new JA-SCF^{COI1} complex to bind to NtJAZs and lead them to proteasomal ubiquitination. The degradation of NtJAZs releases MYC2 to activate pathway gene expression toward the biosynthesis of nicotine. Gene abbreviations are *AO*: aspartate oxidase, *QS*: quinolinate synthase, *QPT2*: quinolinate phosphoribosyltransferase 2, *ADC:* arginine decarboxylase, *AIH*: agmatine iminohydrolase, *ODC*: ornithine decarboxylase, *MPT1*: N-methylputricine transferase 1, *MPO1*: methylputricine oxidase 1, *A622*: an isoflavone reductase like enzyme, *BBL*: berberine bridge enzyme-like protein, and *NND*: nicotine N-demethylase.

the downregulation of nicotine biosynthesis in tobacco plants. Then, we analyzed *cis*-regulatory elements of all *NtJAZ* promoters and cloned MRE(s) and/or G-box(es) from four *NtJAZ* genes. In addition, we analyzed and identified MRE-like and G-box like elements of one nicotine pathway gene. Meanwhile, we analyzed Arabidopsis TFs that have a regulatory function to bind to these types of elements. As a result, PAP1 and TT8, two master regulators of Arabidopsis anthocyanin biosynthesis, were identified for this design. Mechanistic, transgenic, transcriptional, and metabolic quantification, and field farming practices showed that this design was successful to create a distant pathway-cross regulation (DPCR) crossing from the anthocyanin to nicotine pathway. The constructed DPCR showed an effectiveness in significantly reducing nicotine, four OTAs, and four TSNAs in tobacco plants, simultaneously. Moreover, the resulting data disclose new regulatory functions of both PAP1 and TT8 and new activators of four *NtJAZs*.

Materials and methods

Tobacco varieties

Six genotypic tobacco (*Nicotiana tabacum*) cultivars were used in this study. Two commercial dark tobacco varieties were Narrow Leaf Madole (NL) and KY171 for smokeless products (Andersen et al., 1990; Pearce et al., 2015). Two red transgenic genotypes were generated from NL and KY171 described below. Two others were *N. tabacum* Xanthi (an oriental tobacco variety) and a red PAP1 tobacco variety that is a novel homozygous red Xanthi variety (Fig. S22B) created from the overexpression of the *Arabidopsis PAP1* (MYB75) (Xie et al., 2006). *Nicotiana benthamiana* was grown for dual-luciferase assay. Growth conditions in the greenhouse and phytotron are listed in supporting materials.

Cloning of NtJAZ1, NTJAZ3, PMT2, and ODC2 promoters from Narrow leaf Madole and KY171 cultivars

The promoter sequences of tobacco *NtJAZ1, NtJAZ3, NtJAZ7, NtJAZ10, PMT2*, and *ODC2* were identified from the genomic sequences of the NT90 cultivar curated at NCBI (https://www.ncbi.nlm.nih.gov/). The sequences identified were 1000 bp or 1500 bp. Based on these sequences, primers (Table S1) were designed for PCR to amplify *NtJAZ1, NtJAZ3, PMT2*, and *ODC2* promoters from the genomic DNA of wild type NL and KY171 tobacco plants, and then the amplified genomic DNA fragments were cloned to pEASY-T1 for sequencing. Briefly, the DNA fragments from PCR were purified using a gel-extraction Kit (Thermo Fisher, Waltham, MA) by following the manufacturer's instructions and then ligated to the pEASY-T1 plasmid using a T4-ligase system

(10 μl reaction system: 1 μl T4 buffer, 1 μl T4-ligase, 1.0 μl pEASY-T1 linear plasmid and 7.0 μl purified PCR product). The resulting recombinant plasmid was transformed into the competent *E. coli* Top10 cells, which were screened on agar-solidified LB medium supplemented with 50 mg/l ampicillin. Positive colonies were selected for sequencing at Eton Bio (Durham, NC, USA). All promoter sequences obtained were 1000 or 1500 bp (Figs. S1, S2, and S4). Analysis of promoter sequences and identification of regulatory elements such as AC- and AC-like elements (Figs. S1, S2, and S4) were completed with the PLACE (http://www.dna.affrc.go.jp/PLACE/) and PlantPan (http://plantpan.itps.ncku.edu.tw/) tools. The *NtJAZ1, NtJAZ3, PMT2*, and *ODC2* promoters were used for electrophoretic mobility shift assay, dual-luciferase, and ChIP assays described below.

Electrophoretic mobility shift assay of PAP1 and TT8 binding to promoters

To perform electrophoretic mobility shift assay (EMSA), we designed gene-specific primers (Supporting Table S1) for PCR to amplify the binding domains of both *PAP1* and *TT8*. The amplified *PAP1* binding domain from the N-terminus was 534 bp, which included the *R2R3-MYB* domain sequences for Gate-way cloning. The amplified *TT8* binding domain franking the *bHLH* region was 252 bp, in which both BamHI and SacII restriction sites were added to two ends for further cloning. After the confirmation of sequence accuracy, three expression vector systems, pRSF-Dute, pet-Dute, and pDest-His-MBP were comparatively tested to induce soluble recombinant proteins. We selected the pDest-His-MBP vector, a Gateway cloning system, for the *R2R3* domain of *PAP1*. We tested and then identified that the pRSF-Dute vector with a His-tag in the N-terminus was appropriate for expressing the bHLH domain of TT8 after several vectors were tested. As reported previously [69], detailed steps are listed in Method S1.

Dual-luciferase assays

Four reporter and four effector vectors were developed to perform dual-luciferase assays that were completed to analyze the regulatory activity of PAP1 alone, TT8 alone, PAP1 and TT8 together, and PAP1/TT8/TTG1 (WD40) together in activating *NtJAZ1*, *NtJAZ3*, *PMT2,* and *ODC2* promoters. First, the pGreenII-0800 vector was used to develop four reporter vectors. Next, four effector vectors were developed from the PK2GW7 vector [20], to which the ORFs of *GFP*, *PAP1*, *TT8*, and *TTG1* were cloned. Each gene was driven by a 35S promoter. Then, four reporter and four effector vectors were introduced into *Agrobacterium tumefaciens* strain GV3101, respectively. Last, the activated effector and reporter *Agrobacterium* GV3101 cells were used to inoculate leaves of *N. benthamiana* for transient expression assay. Detailed steps are described in Method S2.

Cross-linked chromatin immunoprecipitation (ChIP) assays

A Gateway cloning system was used to develop two plasmids, in which *GFP* was fused to the 3′-end of *PAP1* and *TT8* ORFs without their stop codon for genetic transformation. A pair of primers (Supporting Table S1) was designed to amplify the *PAP1* ORF without its stop codon. The resulting PCR product was purified as described above and then cloned into the pDonr221 vector via a BP reaction. The *PAP1* fragment was further cloned to the upstream of *GFP* in the destination vector pGWB5 via a LR reaction. This cloning step generated a recombinant vector pGWB5-PAP1-GFP, in which *PAP1-GFP* was driven by a 35S promoter. The pGWB5-PAP1-GFP vector was introduced into the competent *E. coli* Top10 cells. Positive colonies were selected on LB medium containing 50 mg/l kanamycin. One colony was used for suspension culture on the same shaker as described above. Cultured *E. coli* cells were harvested to isolate the pGWB5-PAP1-GFP binary plasmid, which was further introduced into competent *Agrobacterium* GV301 cells. To develop a binary vector expressing *TT8*, a pair of primers (Supporting Table S1) was designed for PCR. The reverse primer was designed to include a HA-tag sequence. The resulting PCR product was cloned to the plasmid PK2GW7 as described above. Next, the resulting binary vector PK2GW7-TT8-HA-tag was introduced into competent *Agrobacterium* GV301 cells. *Agrobacterium* GV301 cells harboring *PAP-GFP* or *TT8-HA-tag* cDNA were activated and then injected to leaves of 5-week old wild type NL and KY171 seedlings that were grown in the phytotron. After the injection, plants were allowed to grow 72 hrs. The infected leaves were then harvested to liquid nitrogen to isolate nuclei and ChIP experiments. Detailed steps are listed in Method S3.

Synthesis of T-DNA, development of binary constructs, and genetic transformation of two dark tobacco varieties

PAP1 and *TT8* of *A. thaliana* encode a R2R3-MYB and bHLH TF, respectively. A DNA cassette was designed and then synthesized to stack *PAP1* and *TT8* for their coupled overexpression. From 5′ to 3′ end, the cassette was composed of *attL1*, the *PAP1* cDNA (NP_176057), a *NOS* (terminator), a *35S* promoter, the *TT8* cDNA (CAC14865), and *attL2* (Fig. S7 A). The synthesized sequence was introduced to the entry vector pUC57 and then cloned to the destination binary vector pK2GW7 by attL × attR combination (LR) reaction with LR clonase. The resulting binary vector, namely PAP1-TT8-pK2GW7 (Fig. S7 B), was further introduced into the competent cells of *Agrobacterium tumefaciens* strain GV3101. A positive colony was selected for genetic transformation of tobacco. In addition, the pK2GW7 vector was used for genetic transformation as control. These constructs were transformed into two globally commercial dark tobacco varieties, NL and KY171. Detailed steps (Fig. S7 C) for transformation and selection of T0 transgenic plants followed our protocols reported previously [20]. Transgenic plants from NL and KY171 were labeled as P + T-NL and P + T-KY, respectively. More than 20 T0 transgenic plants were obtained for each variety and grown in the greenhouse to select T0 seeds, 12 of which were further selected on medium supplemented with antibiotics to screen antibiotic resistant T1 progeny. A 3:1 ratio of seed germination to seed death indicated one single copy of T-DNA that was in the genomic DNA of T0 plants. This ratio was observed in the T1 progeny of T0 plants. In this study, we focused on lines with one single copy of transgenes. A large number of seeds were obtained from the T1 progeny of each line. Seeds from these types of T1 progeny were screened on an agar-solidified MS medium supplemented with 50 mg/l kanamycin. If all seeds from a T1 plant could germinate T2 seedlings, it indicated that this T1 plant was a homozygous line. Accordingly, red homozygous T2 plants (Fig. S7 M) were obtained and then used for growth in the greenhouse and field trial. In addition, PCR and qRT-PCR were performed to demonstrate the presence of transgene in the genome of T0 and T2 transgenic plants and the expression of both *PAP1* and *TT8* in both P + T-NL and P + T-KY plants.

DNA and RNA extractions

Genomic DNA was extracted using a DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). RNA was extracted using Trizol extraction reagents. Fresh plant samples were ground into fine powder in liquid nitrogen. Extraction steps were followed the manufacturers' protocols. Detailed steps are listed in Method S4.

Polymerase chain reaction and quantitative reverse transcript-polymerase chain reaction

All real-time quantitative RT-PCR (qRT-PCR) analyses were performed for wild-type, vector control transgenic, and red transgenic plants. SYBR-Green PCR Master-mix (Bio-Rad, Hercules, CA) was used for qRT-PCR experiments on a Step-One Real-time PCR System (Thermo Fisher, Waltham, MA). Detailed steps are listed in Method S5.

Assays of anthocyanin and proanthocyanidin in seedlings

Fresh leaves and roots (100 mg) of wild type, P + T-NL1, P + T-KY1, and vector control seedlings grown hydroponically (Supporting Figure S9) were used to extract anthocyanins and proanthocyanidins (PAs). Methods were followed our anthocyanin [20] and proanthocyanidin assay protocols [70]. Detailed steps are listed in Method S6.

Field trial of P + T-NL1 and P + T-KY1 T2 progeny and air curing

Homozygous T2 progeny of P + T-NL1 and P + T-KY1 were planted at the research station in Oxford, North Carolina in 2015. In addition, NL and KY171 plants were planted as controls. Prior to planting transgenic plants in the field, we applied for GMO permits from USDA-APHIS for this field trial. North Carolina State University was licensed to grow transgenic tobacco plants in the field. Every single step strictly followed the protocol requested in the license. Field planting also followed the protocol of commercial tobacco production in North Carolina. These steps included seedling growth in a contained greenhouse, selection and design of field, planting, phenotypic observation, field management, topping, harvest of leaves, air curing, and cleanup of plant residues from the field (Table S2 and Supporting Figures S15-S19). Detailed steps are listed in Method S7 (Table S2, Figs S15-S19).

Field trial of red PAP1 tobacco plants and flue-curing

PAP1 tobacco is the progeny of a homologous red PAP1 Xanthi plant (#292) that is generated by the overexpression of *PAP1* [21]. The field trials for PAP1 tobacco were conducted in 2011 and 2012. Field design, seed germination, plantation, field management, and plant growth management were the same as described above for P + T-NL1 and P + T-KY1 genotypes. The main differences were that leaves were harvested at different dates and flue-cured in barns. Detailed steps are listed in Method S8 (Fig. S22A-E).

Determination of TSNAs in cured leaves of field-grown P + T-NL1 and P + T-KY1 tobacco plants with high performance liquid chromatography coupled with triple quadrupole tandem mass spectrometer

Accurate quantification of tobacco specific nitrosamines (TSNA) was completed with high performance liquid chromatography coupled with triple quadrupole tandem mass spectrometer (HPLC-QQQ-MS/MS). This analysis was performed using industrial protocols on an Agilent 1200 series HPLC System coupled with AB Sciex Triple quadrupole mass spectrometer (API 5000 with Electrospray) at RJ Reynolds. Four TSNAs, N-nitrosonornicotine (NNN), N-nitrosoanatabine (NAT), N-nitrosoanabasine (NAB), and nicotine-derived nitrosamine ketone (4(methylnitrosamino)-1-(3-pyridyl)-1-butanone, NNK) in air-cured leaves, were quantified. One hundred milligrams of dry leaf powder were extracted in an aqueous ammonium acetate solution (100 mM aqueous ammonium acetate containing four deuterium analogs for TSNA) and filtered through disposable PVDF syringe filters into autosampler vials. The extract was separated in a Phenomenex Gemini C18 column with 3.0 μm particle 2.0 x150 mm. The injection volume was 2.0–5.0 μl. TSNAs were detected by multiple reaction monitoring (MRM) of the precursor ion to a product ion specific for each compound. Quantification was achieved using an internal standard calibration comprised of ten points. A separate internal standard was used for each analyte by using a mixture of four stable isotope-labeled analytes. Results are reported in units of ng/g (ppb). Data were determined to be acceptable if the correlation coefficient for the calibration curve was greater than 0.99, standards were within 85% to 115% of expected concentration values, check solution values were within 75% to 125% of target, peak shape and resolution were acceptable (based on historical data), and values for the QC samples were within established control limits. The instrument parameters for MS, HPLC gradient, the quadrupole mass spectrometer parameters, reagents and standards are listed in in Supporting Tables S3–S6.

Determination of nicotine and other tobacco alkaloids in cured leaves of field grown P + T-NL1 and P + T-KY1 tobacco plants by GC

Analysis of gas chromatograph (GC) equipped with a flame ionization detector (FID) was performed using a standard industry method to quantify nicotine and other tobacco alkaloids at RJ Reynolds. In this study, analysis of nicotine and other tobacco alkaloids was performed on an Agilent 6890 gas chromatograph (GC) equipped with a flame ionization detector (FID) and an Agilent 7683 automatic sampler. One hundred milligrams of air-cured dry leaf powder were alkalinized in 2 mM sodium hydroxide (NaOH). This solution was extracted with methyl-tertbutyl ether (MTBE) spiked with an internal standard using a wrist-action shaker. The mixture was allowed to separate to two layers. The resulting MTBE layer was transferred to an autosampler vial for analysis of nicotine, anabasine, nornicotine, and myosmine with GC-FID. Quantification was achieved using an internal standard calibration comprised of six points. Data were determined to be acceptable if the correlation coefficient of the calibration curve was greater than 0.998, the response factors (RF) were consistent, the percent relative standard deviation (%RSD) for RF was equal or less than 5%, the check solution values were within 10% of target, values for the quality control (QC) samples were within established control limits, and chromatograms had appropriate identification of peaks. The instrument parameters, the oven program, and reagents and conditions used in alkaloid analysis are listed in Supporting Table S7–S9. The results were reported as percentiles (%, g/g, dry weight).

Determination of TSNAs and tobacco alkaloids in flue-cured leaves of PAP1 plants with HPLC-ESI-MS

HPLC-ESI-MS analysis was performed on a 2010EV LC/UV/ESI/MS instrument (Shimadzu, Japan). One gram of powdered samples was extracted in 20 ml of methanol in a flask placed on a shaker for one hour at room temperature. The samples were centrifuged twice at 4000 rpm and aliquots of the supernatant were transferred to a 2.0 ml glass vial. The samples were separated on an Eclipse XDB-C18 analytical column (250 mm × 4.6 mm, 5 μm, Agilent, Santa Clara, CA, USA). The mobile phase solvents were composed of 1% acetic acid in water (solvent A; HPLC-grade acetic acid and LC–MS grade water) and 100% LC–MS grade acetonitrile (solvent B). To separate metabolites, the following gradient solvent system was used with ratios of solvent A to B: 90:10 (0–5 min), 90:10–88:12 (5–10 min), 88:12–80:20 (10–20 min), 80:20–75:25 (20–30 min), 75:25–65:35 (30–35 min), 65:35–60:40 (35–40 min), 60:40–50:50 (40–55 min), and 50:50–10:90 (55–60 min). The column was then washed for 10 min with 10% solvent

B. The flow rate and the injection volume were 0.4 ml/min and 20 μl, respectively. The total ion chromatograms of positive electrospray ionization were recorded from 0 to 60 min by mass spectrum detector and mass spectra were scanned and stored from *m*/*z* of 120–1,600 at a speed of 1,000 amu/s. Standards of NAB, NAT, NNAL, NNK, and NNN (Sigma, St-Louis, MO) were used to establish standard curves with a coefficient value at least 98%. The peak values were used to estimate the levels of nicotine, nornicotine, anabasine, and anatabine.

Determination of nicotine and nornicotine in leaves and roots of seedlings with HPLC-qTOF-MS

Seedlings of P + T-NL1, P + T-KY1, vector control, and wild type plants were hydroponically grown in the phytotron. Leaves and roots of 45-day old seedlings (Supporting Figure S9) were used to analyze nicotine and nornicotine in both leaves and roots. Fresh tissues were ground into fine powder in liquid nitrogen and dried in a 45 °C oven. One hundred mg of powdered samples was extracted with 2.0 ml methanol. Other steps were as described above. The analysis of nicotine and nornicotine was performed with HPLC-qTOF-MS on Agilent 6520 time-of-flight LC-MS/MS instrument (Agilent Technologies, Santa Clara, CA, USA). Mobile solvents and column used for separation of metabolites were solvent A and B described above. The gradient solvent system was composed of ratios of solvent A to B: 95:5 (0–5 min), 95:5 to 90:10 (5–10 min), 90:10 to 85:15 (10–15 min), 85:15 to 45:55 (15–20 min), 45:55 to 25:75 (20–25 min), 25:75 to 95:5 (25–26 min). After the last gradient step, the column was equilibrated and washed for 10 min with solvents A: B (95:5). The flow rate was 0.4 ml/min. The injection volume of samples was 5.0 μl. The dry gas flow and the nebulizer pressure were set at 12 l/min and at 50 psi, respectively. Metabolites were ionized with the negative mode. The mass spectra were scanned from 100 to 3000 *m*/*z*. The acquisition rate was three spectra per second. Other MS conditions included fragmentor: 150 V, skimmer: 65 V, OCT 1 RF Vpp: 750 V, and collision energy: 30 psi. Standard curves were developed for both nicotine and nornicotine. The results were reported as mg/g, dry weight.

Statistical analysis

Metabolite values for each group of samples such as B1 were averaged from samples of three plots. One-way ANOVA was used to evaluate the statistical significance of the contents of nicotine, nornicotine, anabasine, NNN, NNK, NAT, and NAB between red and wild type leaves. Standard deviation was used to show the variation of metabolite contents. For gene expression, protein analysis, nicotine and nornicotine contents in seedlings, all values are averaged from four or five biological replicates. Student *t*-test was performed to evaluate statistical significance. P-values less than 0.05, 0.01, and 0.001 mean significant, very significant, and extremely significant.

Results

MRE and G-box elements identified in the promoters of NtJAZ1, 3, 7, and 10

NtJAZ1 and NtJAZ3 are two main repressors of nicotine biosynthesis [54,55,60]. In a recent study, totally 15 *NtJAZs* were mined from the tobacco genomes, showing that *NtJAZ7*, and *NtJAZ10* were also associated with the regulation of nicotine biosynthesis [61]. We identified the promoter sequences of 13 *NtJAZs* from the genome of NC90 tobacco curated at NCBI and named these as *NtJAZ1-pro* through *NtJAZ13pro*. One or 1.5 kb nucleotides in the proximal upstream from the start codon (ATG) were obtained for sequence characterization. Sequence analysis revealed that *NtJAZ1pro*, *NtJAZ3pro*, and *NtJAZ7pro* had both MYB recognizing element (MRE) and G-box elements, while *NtJAZ10pro* only had G-box elements (Figs S1-S2). Next, we cloned the promoter sequences of *NtJAZ1pro* and *NtJAZ3pro* from two commercial smokeless dark tobacco cultivars, Narrow Leaf Madole (NL) and KY171 (KY). Their sequences were identical to those of *NtJAZ1pro* and *NtJAZ3pro* from TN90 (Fig. S1), therefore, we used the same names in this report. Sequence analysis further substantiated the MRE and G-box elements in *NtJAZ1pro* and *NtJAZ3pro*. *NtJAZ1pro* has two MRE and two G-box elements (Fig. 2A and Fig. S1). The two MRE elements are MRE1: ACCCACC at position −180 and MRE2: ACCCCAC at position −172. The two G-boxes are CACGTG at positions −232 and AACGTG at −395. *NtJAZ3pro* has one MRE3 element: AACTACC at position −169 and one G-box: CACGTG at position −960 (Fig. 2A and Fig. S1).

EMSA shows the binding of PAP1 and TT8 to the NtJAZ1 and NtJAZ3 promoters in vitro

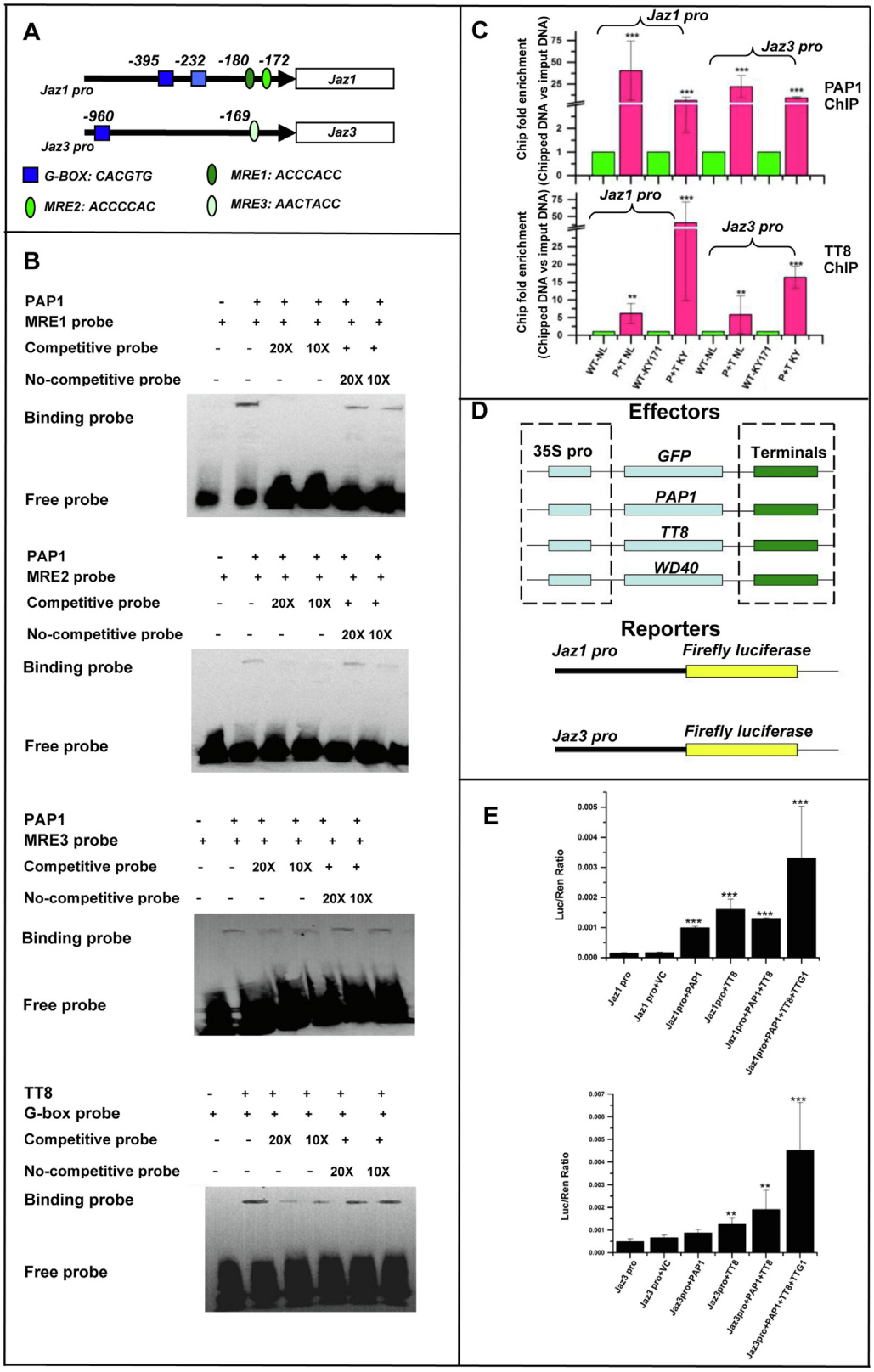
Electrophoretic mobility shift assay (EMSA) was carried out to test whether PAP1 and TT8 could bind to the MRE and G-box elements of $NtJAZ1_{pro}$ *and* $NtJAZ3_{pro}$ *in vitro*. The R2R3-MYB binding domain (amino acids 1–178) of Arabidopsis PAP1 was cloned to the pDEST-HisMBP vector, in which it was fused with the His-MBP-Tag at the N-terminus (Fig. S3A). The bHLH binding domain (amino acid 359–443) of TT8 was cloned to the PRSF-Dute vector, in which it was fused with the His-Tag at the N-terminus (Fig. S3B). Two constructs were introduced to *E. coli* to induce recombinant proteins, which were further purified with a His-tag purification system (Fig. S3B). Biotin probes were prepared for MRE1 (ACCCACC), MRE2 (ACCCCAC), MRE3 (AACTACC), and G-box (CACGTG). Both competitive and non-competitive probes were used as controls for EMSA. The resulting data showed that the recombinant R2R3 binding domain directly bound to three MRE element probes and the recombinant bHLH binding domain directly bound to the G-box elements (Fig. 2B). The binding activity of both recombinant R2R3 and bHLH binding domains was competed by competitive probes but not competed by non-competitive probes (Fig. 2B). These data demonstrate that both PAP1 and TT8 bind to the $NtJAZ1_{pro}$ *and* $NtJAZ3_{pro}$.

ChIP-Q-PCR shows the binding of PAP1 and TT8 to the NtJAZ1 and NtJAZ3 promoters in vivo

ChIP-Q-PCR analysis was performed to examine whether PAP1 and TT8 could bind to *NtJAZ1pro* and *NtJAZ3pro in vivo*. In ChIP experiments, *PAP1* was fused at the 5′-end of *GFP* in the binary vector pGWB5 to obtain pGWB5-PAP1-GFP, which was introduced into leaves of wild type NL and KY171 seedlings for transient expression via injection of *Agrobacterium*. *TT8* was fused at the 5′-end of an *HA-tag* sequence in the binary vector PK2GW7 to obtain PK2GW7-TT8-HA-tag, which was also introduced into leaves of wild type NL and KY171 seedlings for transient expression by injection of *Agrobacterium*. The IP experiments of PAP1 and TT8 were performed by using GFP and HA monoclonal antibodies. The ChIP PCR results showed that the amplified *NtJAZ1* and *NtJAZ3* promoter fragments were increased 5.0–40.4 times by PAP1 and 6.18–37.5 times by TT8 (Fig. 2C).

Dual-luciferase assay demonstrates the activation of NtJAZ1 and NtJAZ3 promoters by PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1

A dual-luciferase assay was completed to demonstrate whether PAP1 alone, TT8 alone, PAP1 and TT8 together, and PAP1/TT8/TTG1


A
-395 -232 -180 -172
Jaz1
Jaz1 pro
-960 -169
Jaz3
Jaz3 pro
G-BOX: CACGTG
MRE1: ACCCACC
MRE2: ACCCCAC
MRE3: AACTACC
B
PAP1
MRE1 probe
Competitive probe
No-competitive probe
Binding probe
Free probe
20X 10X
MRE2 probe
MRE3 probe
TT8
G-box probe
C
Jaz1 pro
Jaz3 pro
PAP1 ChIP
TT8 ChIP
Chip fold enrichment
(Chipped DNA vs imput DNA)
WT-NL
P+T NL
WT-KY171
P+T KY
D
Effectors
35S pro
GFP
Terminals
PAP1
TT8
WD40
Reporters
Firefly luciferase
E
Luc/Ren Ratio
Jaz1 pro
Jaz1 pro+VC
Jaz1pro+PAP1
Jaz1pro+TT8
Jaz1pro+PAP1+TT8
Jaz1pro+PAP1+TT8+TTG1
Jaz3 pro
Jaz3 pro+VC
Jaz3pro+PAP1
Jaz3pro+TT8
Jaz3pro+PAP1+TT8
Jaz3pro+PAP1+TT8+TTG1

(WD40) together could activate the *NtJAZ1* and *NtJAZ3* promoters in tobacco. PAP1, TT8, and TTG1 were used as effectors. The ORF sequences of *PAP1*, *TT8*, *TTG1*, and *GFP* were respectively cloned into the PCB2004 plasmid to generate four constructs, in which each gene was driven by a 35S-promoter (Fig. 2D). The *NtJAZ1pro* and *NtJAZ3pro* were cloned into the plasmid pGreenII-0800 to drive the luciferase gene to obtain *NtJAZ1pro*-luciferase and *NtJAZ3pro*-luciferase reporters (Fig. 2D). Effectors and reporters were incubated for dual-luciferase assays. Promoters alone and a binary vector were used as controls. The resulting data showed that compared with controls, PAP1 and TT8 alone enhanced the activity of *NtJAZ1pro* 3.8 and 7.2 times, respectively, and increased the activity of *NtJAZ3pro* 1.9 (slightly) and 2.1 (significantly) times, respectively (Fig. 2E). PAP1 and TT8 together enhanced the activity of *NtJAZ1pro* and *NtJAZ3pro* approximately 4.0 and 6.3 times (Fig. 2E). Furthermore, PAP1, TT8, and TTG1 together increased 12.7 and 15.6 times of the activity of both *NtJAZ1pro* and *NtJAZ3pro*, much higher than PAP1 alone, TT8 alone, and PAP1 and TT8 together (Fig. 2E). These results demonstrate that PAP1 and TT8 alone, PAP1 and TT8 together, PAP1-TT8-TTG1 complex transcriptionally activate both *NtJAZ1* and *NtJAZ3*.

PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1 do not activate promoters of ODC2 and PMT of tobacco

NtODC2 and *NtPMT2* are two key upstream pathway genes of the nicotine biosynthesis. We obtained 1.0 kb long sequences for these two genes' promoters from the tobacco genome database curated at NCBI. The promoter sequence of *NtPMT2* has an MRE-like element (AACAACC at position −530) and a G-box like (CACGTT at position −186) element, while the sequences of the *NtODC2* promoter does not have MRE and G-box elements (Fig. S4).

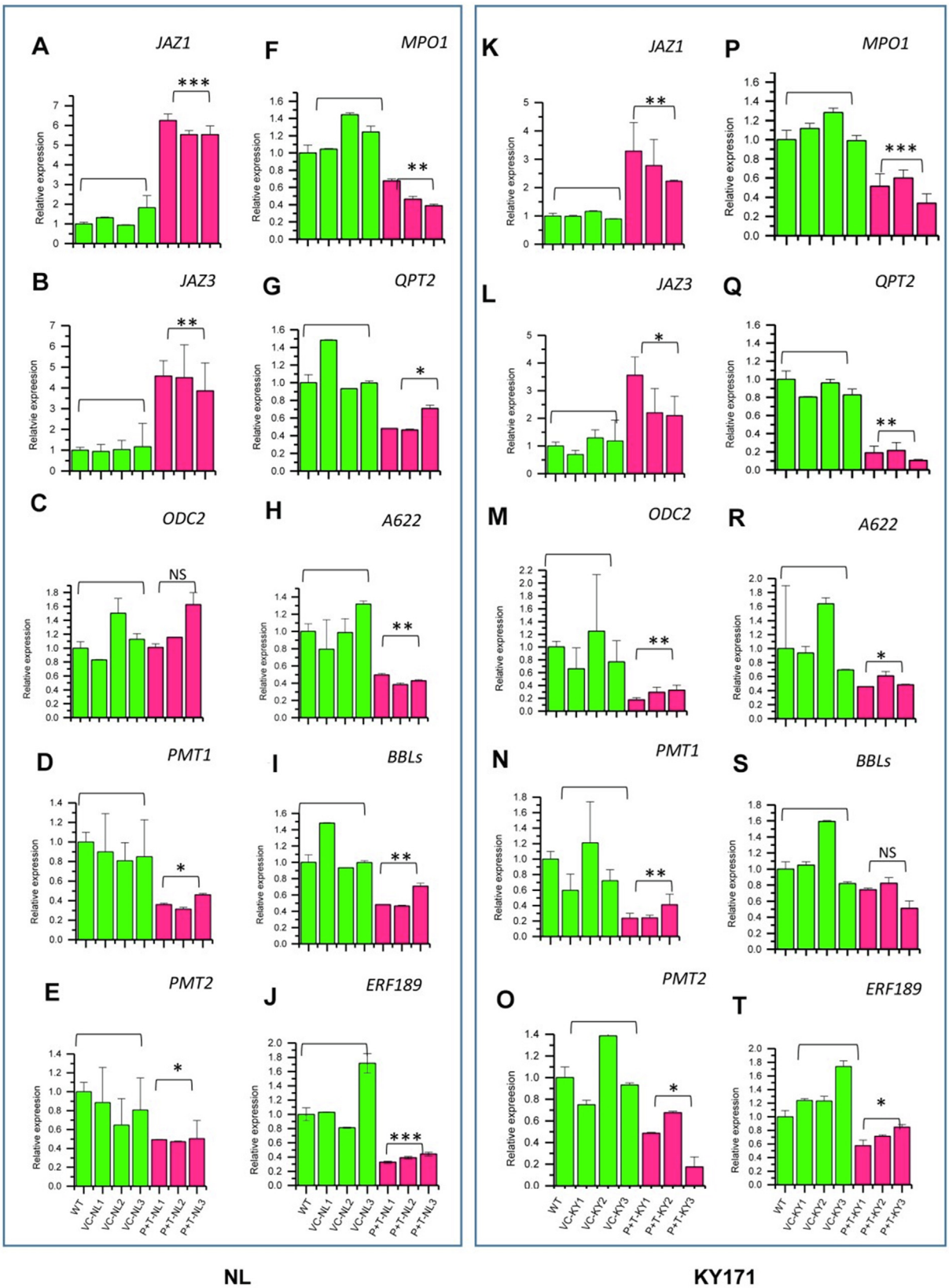
EMSA experiments were performed to test whether PAP1 and TT8 could bind to MRE-like and G-box-like elements. The EMSA results indicated that the TT8 binding domain could bind to the G-box-like (CACGTT) element. However, when a competitive probe was added, the binding signal was lost (Fig. S5 A). Furthermore, compared with the positive control G-box elements from *NtJAZ1* and *NtJAZ3* promoters, the resulting data showed that the signal from G-box like element was much weaker (Fig. S5 A). This result indicates that TT8 can weakly bind to the promoter of *NtPMT2*. When the MRE-like element probe was prepared for EMSA, the resulting data showed that PAP1 could not bind to the MRE-like probe (AACAACC), while PAP1 strongly bound to the positive control MRE1 and MRE3 probes of *NtJAZ1* and *NtJAZ3* promoters (Fig. S5 B).

Dual-luciferase assay and ChIP-Q-PCR were performed to test whether PAP1 and TT8 could activate the *ODC2* and *PMT2* promoters. The dual-luciferase assay data did not show that PAP1 alone, TT8 alone, PAP1/TT8 together, and the PAP1/TT8/TTG1 complex could activate the activity of the two promoters (Fig. S6 A). Meanwhile, the ChIP-Q-PCR data showed that PAP1 and TT8 could not enrich binding elements of the two promoters *in vivo* (Fig. S6 B). These data indicate that PAP1, TT8, PAP1/TT8, and PAP1/TT8/TTG1 cannot activate the promoters of these two key pathway genes.

Engineering of red genotypic P + T-NL and P + T-KY tobacco plants

An expression cassette was synthesized to stack the ORF sequences of *PAP1* (P) and *TT8* (T) and then used to generate 20 T0 transgenic plants from each of Narrow Leaf Madole (NL) and KY171 (KY), two major commercial varieties (Fig. S7 A, B, C, E, and G-H). In addition, vector control transgenic plants were generated (Fig. S7 D and F). The first 12 red T0 transgenic plants from the two cultivars were tagged as P + T-NL1 to 12 and P + T-KY1 to 12, respectively. The first 12 T0 vector control transgenic plants from the two cultivars were tagged as NL1 to 12 and KY1 to 12, respectively. These T0 plants were selected to obtain seeds to screen T1 and homozygous T2 progeny (Fig. S7 I-M), during which the copy number of the transgenes was determined. Five P + T-NL lines (1–5) and five P + T-KY lines (1–5) were confirmed to contain one single copy of the stacked two transgenes. Quantitative RT-PCR showed that homozygous T2 progeny highly expressed *PAP1* and *TT8* (Fig. S8). Furthermore, qRT-PCR was performed to understand the expression of *CHS*, *CHI, F3H, F3′H*, *DFR*, and *ANS* involved in the biosynthesis of both anthocyanin and proanthocyanidin (PA) and the expression of *3-GT* involved in the glycosylation of anthocyanidins (Fig. 1 A) in T2 seedlings (Fig. S9). The resulting data showed that the expression levels of these seven genes were up-regulated 2–1000 times (Fig. S10 A-B) in pink roots and red leaves of the P + T-NL and P + T-KY genotypes. Further UV spectrometry-based measurements showed that the roots and leaves of analyzed P + TNL1 and P + T-KY1 genotypes produced anthocyanins, while those of control plants did not (Fig. S11). It was noticeable that the levels of anthocyanins were significantly higher in leaves than in roots of both P + T-NL1 and P + T-KY1 plants (Fig. S11). To understand this differentiation, the expression of *ANR* specifically involved in the colorless PA formation [71] was analyzed. The resulting data showed that the expression of *ANR* was significantly activated in roots but not in leaves of the two red genotypes (Fig. S10). To substantiate this, root extracts of both P + T-NL1 and P + T-KY1 seedlings were treated with DMACA. A bluish color was immediately developed from the reaction (Fig. S12 A-B), showing the presence of PAs. However, this bluish color was not developed from the extracts of both P + T-NL1 and P + T-KY1 leaves and wild type roots and leaves, indicating the lack of PAs. Further measurement of DMACA-treated samples at 640 nm showed that the ABS values were 5–6 times higher in the root extracts than in the leaf extracts of P + T-NL1 and P + T-KY1 and in the leaf and root extracts of wild type (Fig. S12 C-D). These data showed that the less red pigmentation in roots resulted from the conversion of red anthocyanidins to colorless PAs by ANR. In addition, the less red

Fig. 2. Binding and activation of *NtJAZ1* and *NtJAZ3* promoters by PAP1 and TT8 alone, PAP1 and TT8 together, and PAP1-TT8-WD40 complex. A, three MYB response element (MRE) types and G-box were identified in the *NtJAZ1* and *NtJAZ3* promoters; B, electrophoretic mobility shift assay (EMSA) showed that PAP1 and TT8 bound to three MRE types and G-box of the NtJAZ1 and NtJAZ3 promoters, respectively. Competitive and non-competitive probes were used for binding of both TFs. 20 × and 10×: concentrations of tested competitive and non-competitive probes 20 and 10 times as those MRE and G-box probes. C, fold change values from ChIP-qPCR showed that both PAP1 and TT8 bound to the *NtJAZ1* and *NtJAZ3* promoters *in vivo*. Approximately 150 bp *NtJAZ1* and *NtJAZ3* promoter fragments containing both MRE and G-box were enriched with anti-HA antibodies in qRT-PCR analysis. The regions of *NtJAZ1* and *NtJAZ3* promoters that do not contain a MRE and G-box were used as negative controls. Green and red bars represent wild type and red tobacco plants, respectively. D, schematic diagrams shows four effector constructs (PK2GW7-GFP, PAP1, TT8 and TTG1) and two reporter constructs (pGreenII-0800-Jaz1pro and Jaz3 pro) used for dual-luciferase assay. E, luciferase (LUC)/renilla (REN) luminescence ratios from dual luciferase assays showed that PAP1 and TT8 alone bound to and activated the promoters of both *NtJAZ1* and *NtJAZ3*, two TFs together increased promoter activity, and two TFs and WD40 together increased the highest activity. The promoters were fused to a firefly luciferase (reporter). The promoters of *NtJAZ1* and *NtJAZ3* genes were used in dual luciferase assay. Values represent the mean ± S.D. (n = 5). Asterisks on top of bars indicate that the values are significantly higher than those of controls (*P less than 0.05, **P less than 0.01, ***P less than 0.001). Error bars indicate standard errors.


A
JAZ1
***
Relative expression
F
MPO1
**
Relative expression
B
JAZ3
**
Relative expression
G
QPT2
*
Relative expression
C
ODC2
NS
Relative expression
H
A622
**
Relative expression
D
PMT1
*
Relative expression
I
BBLs
**
Relative expression
E
PMT2
*
Relative expression
J
ERF189
***
Relative expression
WT
VC-NL1
VC-NL2
VC-NL3
P+T-NL1
P+T-NL2
P+T-NL3
NL
K
JAZ1
**
Relative expression
P
MPO1
***
Relative expression
L
JAZ3
*
Relative expression
Q
QPT2
**
Relative expression
M
ODC2
**
Relative expression
R
A622
*
Relative expression
N
PMT1
**
Relative expression
S
BBLs
NS
Relative expression
O
PMT2
*
Relative expression
T
ERF189
*
Relative expression
WT
VC-KY1
VC-KY2
VC-KY3
P+T-KY1
P+T-KY2
P+T-KY3
KY171

pigmentation in roots might result from the lack of light in the hydroponic condition and secretion of anthocyanins.

Upregulation of NtJAZ1, 3, 7, and 10 and downregulation of the nicotine biosynthesis in red P + T-NL and P + T-KY plants

Both transcriptional and metabolic profiling were performed to characterize the biosynthesis of nicotine in homozygous T2 seedlings (Fig. S9) of different genotypes selected, including P + T-NL1, P + T-NL2, P + T-NL3, P + T-KY1, P + T- KY 2, and P + T-KY3. Quantitative RT-PCR was performed to characterize the transcriptional profiles of 16 genes (Fig. 1 B) involved in the nicotine biosynthesis. The expression levels of *NtJAZ1, NtJAZ3*, *NtJAZ7,* and *NtJAZ10* were significantly increased in roots of both P + T-NL and P + T-KY genotypes (Fig. 3 A-B and K-L, Fig. S13). The expression levels of *PMT1*, *PMT2*, *MPO*, *QPT2*, *A622*, *BBL* and *ERF189* were significantly reduced in roots of both P + T-NL (Fig. 3 D-J) and P + T-KY plants (Fig. 3 N-T). The expression levels of *OCD2* was not altered in the P + T-NL1, 2, and 3 lines (Fig. 3 C) but significantly decreased in the P + T-KY1, 2 and 3 lines (Fig. 3 M), indicating the different responses of this gene expression in two commercial varieties to transgenes. The expression levels of *MYC1a*, *MYC1b*, *MYC2a*, and *MYC2b* were not altered in roots of both P + T-NL and P + T-KY plants (Fig. S14). HPLC-MS analysis showed that the contents of nicotine, nornicotine, and the total amount of the two compounds in roots and leaves of P + K-NL1, P + K-NL2, P + K-NL3, P + K-KY1, P + K-KY2, and P + K-KY3 plants were significantly reduced 1.5–3.4, 1.3–2.5, and 1.9–3.3 fold, respectively (Fig. 4).

Field trials and reduction of nicotine, nornicotine, anatabine, anabasine, myosmine, and four TSNAs in red P + T-NL1 and P + T-KY1 tobacco plants

Field trials that follow the commercial production practices of tobacco are essential to examine the effects of new varieties or genes on the contents of nicotine, other tobacco alkaloids, and TSNAs. In addition, field farming practices were necessary to test the agricultural significance of this *de novo* regulation design. The field practice exactly followed the protocols of both industry tobacco and GMO tobacco growth. We planted T2 homozygous P + T-NL1 and P + T-KY1 progeny and their corresponding wild-type tobacco variety in the field at the research station in Oxford, North Carolina (Fig. 5 A-C and Figs S15-S18). Tobacco leaves were air-cured in a contained ventilation barn (Fig. 5D, Figs S18-S19).

Quantification with GC-FID showed the significant reduction of nicotine, nornicotine, anabasine, anatabine, and myosmine in most or all cured leaves of red P + T-NL1 and P + T-KY1 tobacco plants compared to those of non-transgenic controls. The contents of nicotine were significantly reduced by 45–51% in all cured leaf groups of P + T-NL1 (Fig. 6A) and by 19–30% in upper three groups of cured leaves of P + T-KY1 (Fig. 6B). The contents of nornicotine were significantly decreased by 39–44% and 30–40% in all cured leaf groups of T-P-NL1 and P + T-KY1, respectively (Fig. 6C and D). The contents of anabasine, anatabine, and myosmine were significantly reduced by 27–45%, 55–66%, and 19–25% in all cured leaf groups of P + T-NL1 (Fig. 6E, G, and I), respectively, and were significantly reduced by 27–45%, 32-56%, and 29–35% in all cured leaf groups of P + T-KY1, respectively (Fig. 6F, H, and J).

Quantification with HPLC-QQQ-MS/MS showed that the contents of four carcinogenic TSNAs were significantly diminished in most or all leaf groups of P + T-NL1 and P + T-KY1 plants. The content of NNN was significantly decreased by 63–74% to a level of 0.136–0.375 ppm in all four groups of cured P + T-NL1 leaves and by 67–79% to a level of 0.208–0.552 ppm in three groups of cured P + T-KY1 leaves (Fig. 7A and B). The content of NNN in the B3 leaf group of P + T-KY1 was reduced by 63% to 0.488 ppm with a slight significance. The content of NNK was reduced by 38–72% to a level of 0.1–0.173 ppm and by 30–74% to a level of 0.151–0.3 ppm in four groups of cured P + T-NL1 and P + T-KY1 leaves, respectively, in which the reduction in three leaf groups of each genotype was significant (Fig. 7C and D). The content of NAT was significantly decreased by 77–92% to a level of 0.093–0.159 ppm and by 70–80% to a level of 0.159–0.285 ppm in all leaf groups of P + T-NL1 and P + T-KY1, respectively (Fig. 7E and F). The content of NAB was significantly decreased by 72–80% to a level of 0.011–0.021 ppm and by 65–77% to a level of 0.015–0.031 ppm in all leaf groups of P + T-NL1 and P + T-KY1, respectively (Fig. 7G and H). Furthermore, the contents of NNN, NNK, NAT, and NAB were summed to obtain the total TSNA contents in each leaf group. The total contents of four TSNAs in each group of wild type NL and KY171 leaves were 1.6–2.2 ppm and 2.2–2.6 ppm, while these values were significantly reduced to be 0.4–0.6 ppm and 0.6–1.0 ppm in each group of P + T-NL1 and P + T-KY1 leaves (Fig. 7I, J). The average contents of four TSNAs in the entire plant leaves of wild type NL and KY171 were approximately 1.77 ppm and 2.25 ppm, while these values were significantly reduced to be 0.69 ppm and 0.87 ppm in those of P + T-NL1 and P + T-KY1 (Fig. 7K,L).

Furthermore, the total contents of tobacco alkaloids in each group of leaves were summed from nicotine, four other alkaloids, and total TSNAs. The resulting data showed that the contents of total alkaloids were significantly decreased by 44–51% in four leaf groups of P + TNL1 and 22–35% in the upper three group leaves of P + T-KY1 (Fig. S20A,B).

Field trials and reduction of nicotine, nornicotine, anatabine, anabasine, myosmine, and four TSNAs in flue-cured leaves of the red PAP1 tobacco

Our red PAP1 tobacco genotype (Fig. S22B) is a homozygous red variety generated from Xanthi by the overexpression of *PAP1* [21]. Gene expression analysis showed that the expression of *PAP1* upregulated *NtAn1a* and *NtAn1b* (Fig. S21), two *TT8* homologs [72]. This result indicated that *PAP1* alone activated tobacco anthocyanin biosynthesis via upregulating the endogenous *N. tabacum TT8* (*NtTT8*) homologs. Therefore, to substantiate our *de novo*

Fig. 3. Upregulation of two *NtJAZs* and downregulation of nicotine pathway genes and *ERF189* in roots of P + T transgenic seedlings. A-J, in roots of P + T-NL transgenic lines, the expression levels of *NtJAZ1* (A) and *NtJAZ3* (B) were significantly upregulated; the expression level of *ODC2* (C) was not changed; while the expression levels of *PMT1* (D), *PMT2* (E) *MPO* (F), *QPT* (G), *A622* (H), *BBLs* (I) (primer pairs designed for all three BBLs), and *ERF189* (J) were significantly downregulated. K-T, in roots of P + T-KY transgenic lines, the expression levels of *NtJAZ1* (K) and *NtJAZ3* (L) were significantly upregulated, while, the expression levels *ODC2* (M), *PMT1* (N), *PMT2* (O) *MPO* (P), *QPT* (Q), *A622* (R), *BBLs* (S), and *ERF189* (T) were significantly downregulated. Data of three transgenic lines are shown for both P + T-NL and P + T-KY genotypes. Wild type samples were pooled from five plants. Data of three lines are shown for vector control transgenic NL and KY171 plants. WT: wild type, P + T NL1, P + T NL2, P + T NL3: three lines of stacked *PAP1* and *TT8* transgenic NL plants, VC-1, 2, and 3-NL: three vector control transgenic NL lines, P + T KY1, P + T KY2, P + T KY3: three lines of stacked *PAP1* and *TT8* transgenic KY171 plants. VC-1, 2, and 3-KY: three vector control transgenic KY171 lines. Green and red bars represent wild type and vector control transgenic, and red P + T-NL and P + T-KY plants, respectively. Values represent the mean ± S.D. (n = 3). Asterisks on top of the bars indicate that values are significantly lower or higher in red transgenic lines than in wild type plants (*P < 0.05, **P < 0.01, ***P < 0.001). Error bars indicate standard errors.

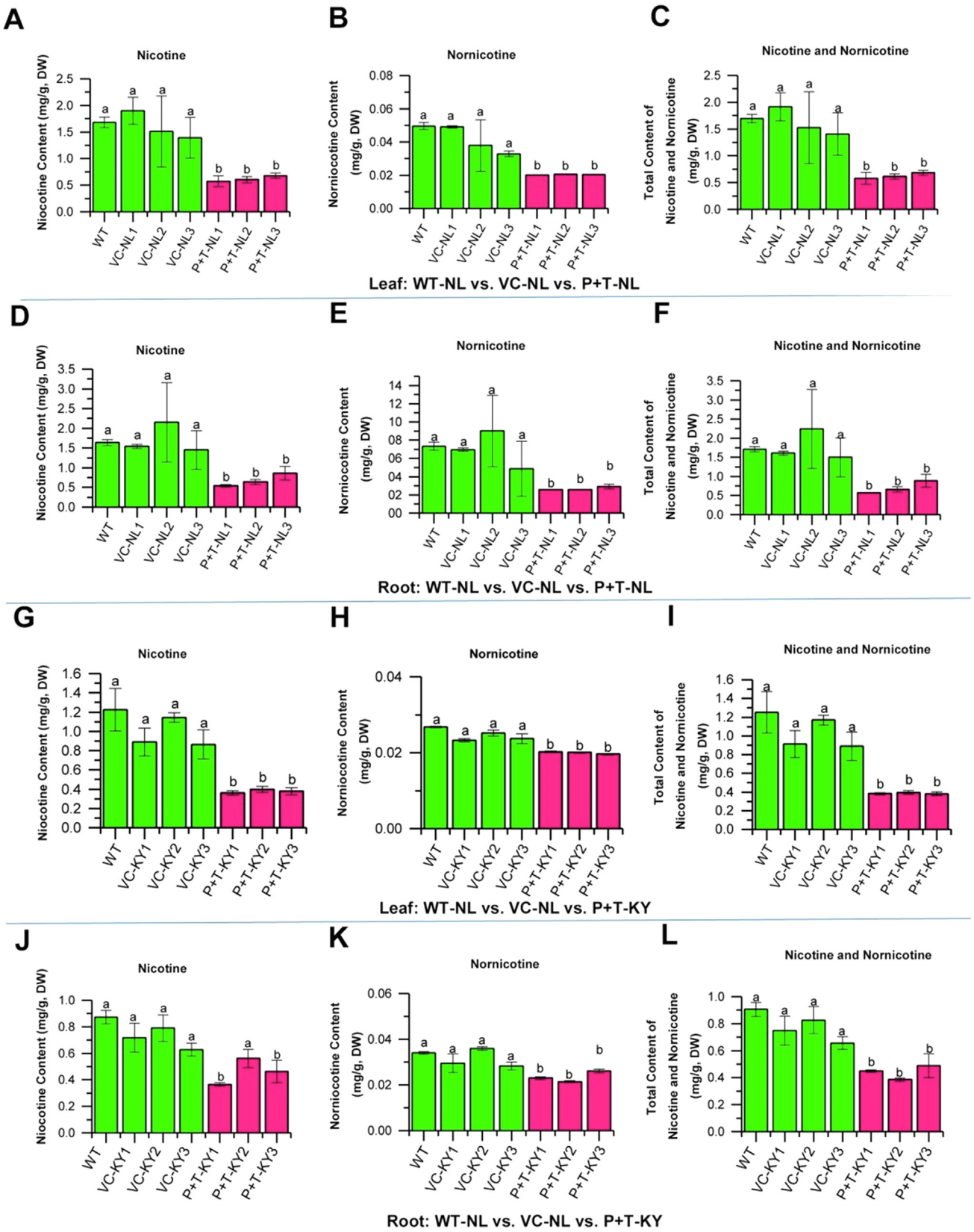


Fig. 4. Significant reduction of nicotine and nornicotine contents in leaves and roots of P + T-NL and P + T-KY seedlings. The contents of nicotine and nornicotine were significantly reduced in roots and leaves of both transgenic P + T-NL and P + TKY lines. A-C: reduction of nicotine (A), nornicotine (B), and total nicotine and nornicotine (C) in leaves of three P + T-NL lines (1, 2, and 3); D-F: reduction of nicotine (D), nornicotine (E), and total nicotine and nornicotine (F) in roots of three P + T-NL lines (1, 2, and 3); G-I: reduction of nicotine (G), nornicotine (H), and total nicotine and nornicotine (I) in leaves of three P + T-KY lines (1, 2 and 3); J-L: reduction of nicotine (J), nornicotine (K), and total nicotine and nornicotine (L) in roots of three P + T-NL lines (1, 2, and 3). Bars labeled with "a" and "b" indicate p-value less than 0.05 and bars labelled with the same lowercase "a" or "b" indicate no significant difference.

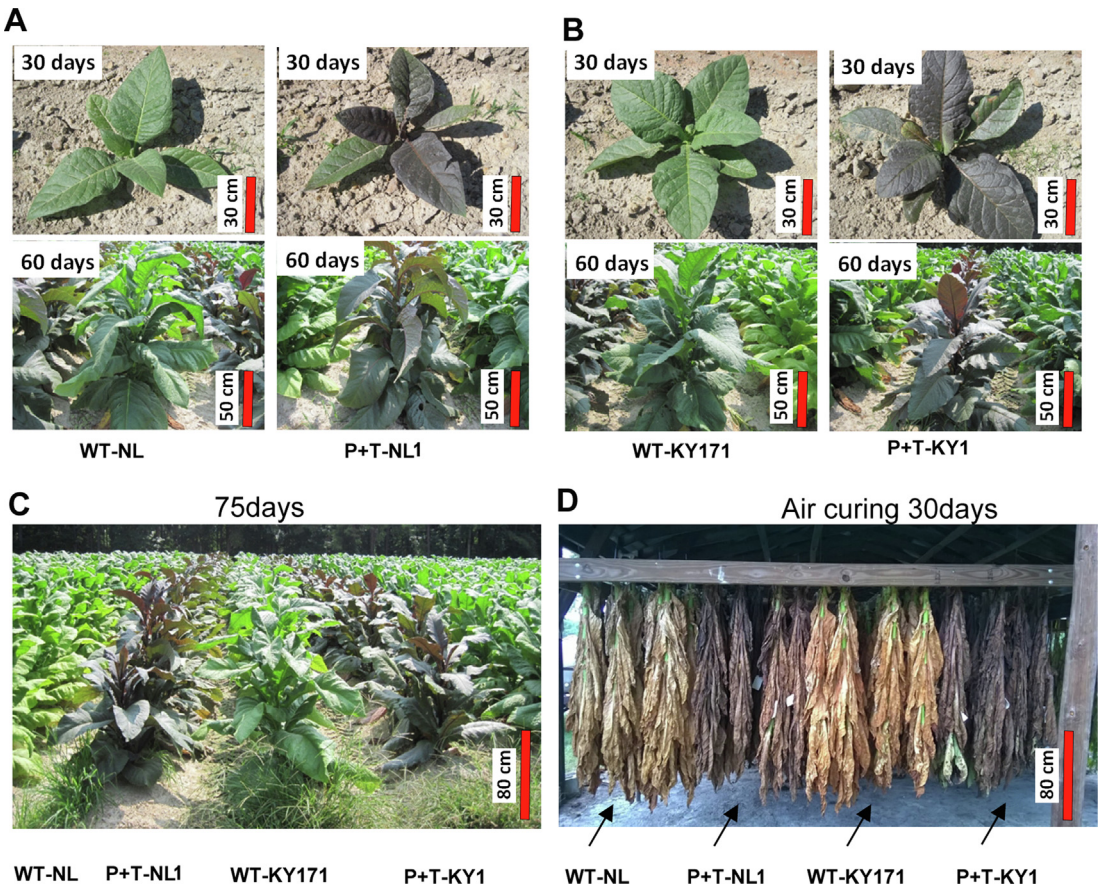


Fig. 5. Phenotypes of P + T-NL1 and P + T-KY1 plants versus their corresponding wild-type tobacco plants in the field and of air-cured leaves. Field farming practice of four genotypes was performed in Oxford, North Carolina. A, phenotypes of 30-day and 60-day old WT-NL vs. red P + T-NL1 plants. B, phenotypes of 30-day and 60-day old WT-KY171 vs. red P + T-KY1 plants. C-D, phenotypes of topped plants (C) and leaves in air curing (D). Plant name abbreviations are WT-NL: wild-type Narrow Leaf Madole (NL) variety, P + T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WT-KY171: wild-type KY171 variety, and P + T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.

design, we used this genotype to test whether PAP1 and activated NtTT8 homologs could reduce nicotine, other tobacco alkaloids, and TSNAs.

Two years' field trials were completed. The field design (Fig. S22A,B) and cropping management were the same as described above. When leaves were harvested, they were harvested into five groups (Fig. S22C,D) and then flue-cured in barns (Fig. S22E). Analysis with HPLC-ESI-MS obtained peak values that showed the reduction of nicotine, nornicotine, anabasine, and anatabine. The levels of nicotine were significantly or slightly reduced by 10–50% and 5–25% in each group of leaves in 2011 and in 2012, respectively (Fig. S23A,B). The levels of nornicotine, anabasine, and anatabine were significantly or slightly reduced in all or most group leaves by 25–70% and 5–55%, 30–60% and 5–20%, and 5–60% and 10–40% in 2011 and 2012, respectively (Fig. S23A,B). In addition, the total cumulative peak level of each alkaloid was summed from all leaf groups of the entire PAP1 tobacco plant. The resulting data showed that the total cumulative levels of nicotine, nornicotine, anabasine, and anatabine in the entire plant were significantly reduced by 30%, 45%, 50%, and 40% in 2011 and 10%, 20%, 10%, and 30% in 2012 respectively (Fig. S24A,B).

The contents of NNN, NNK, NAB, and NAT were reduced in flue-cured leaf groups (Fig. S25A,B). The contents of NNN were reduced by 20–60% and 25–60% in all leaf groups in two years. The average content of NNN was diminished to a level less than 0.5 ppm in most leaf groups (Fig. S25A,B). The contents of NNK were reduced by 5–50% in all leaf groups in 2011 and 5–60% in four of five leaf groups in 2012. The contents of NAB were reduced by 20–70% and 15–50% in all leaf groups in two years. The contents of NAT were reduced by 50–65-% and 35–70% in all leaf groups in two years. The contents of each NNN, NNK, NAB, and NAT were averaged for all leaf groups of each plant. The resulting data showed that the average contents of each TSNA were significantly reduced by 25–60% (P-value less than 0.05) in 2011 and 2012 (Fig. S25A,B). The contents of four TSNAs were summed for all leaf groups of each plant. The resulting data showed that the total contents of four TSNAs were significantly reduced by 57% and 43% (P-value <0.05) in 2011 and 2012, respectively (Fig. S26).

Discussion

Our findings indicate that *cis*-regulatory elements and TFs are useful molecular tools for a *de novo* design to create a novel regulation of plant secondary metabolism. Herein, based on the regulation mechanism elucidated (Fig. 2), we constructively term the designed regulation to be a *Distant Pathway-Cross Regulation* (DPCR) of PAP1 and TT8. The rationale is that the pathway of nicotine does not exist in Arabidopsis and two anthocyanin TFs are metabolically characterized to be distant from the tobacco alkaloid pathway. The other reason is that our findings are distinct from the general pleiotropic effects that have been generally used to explain gene functions without a mechanistic elucidation. This DPCR definition is characterized by the mechanisms that PAP1, TT8, and their complex formed with WD40 negatively regulate the tobacco alkaloid biosynthesis via activating *NtJAZ* repressor genes. Accord-

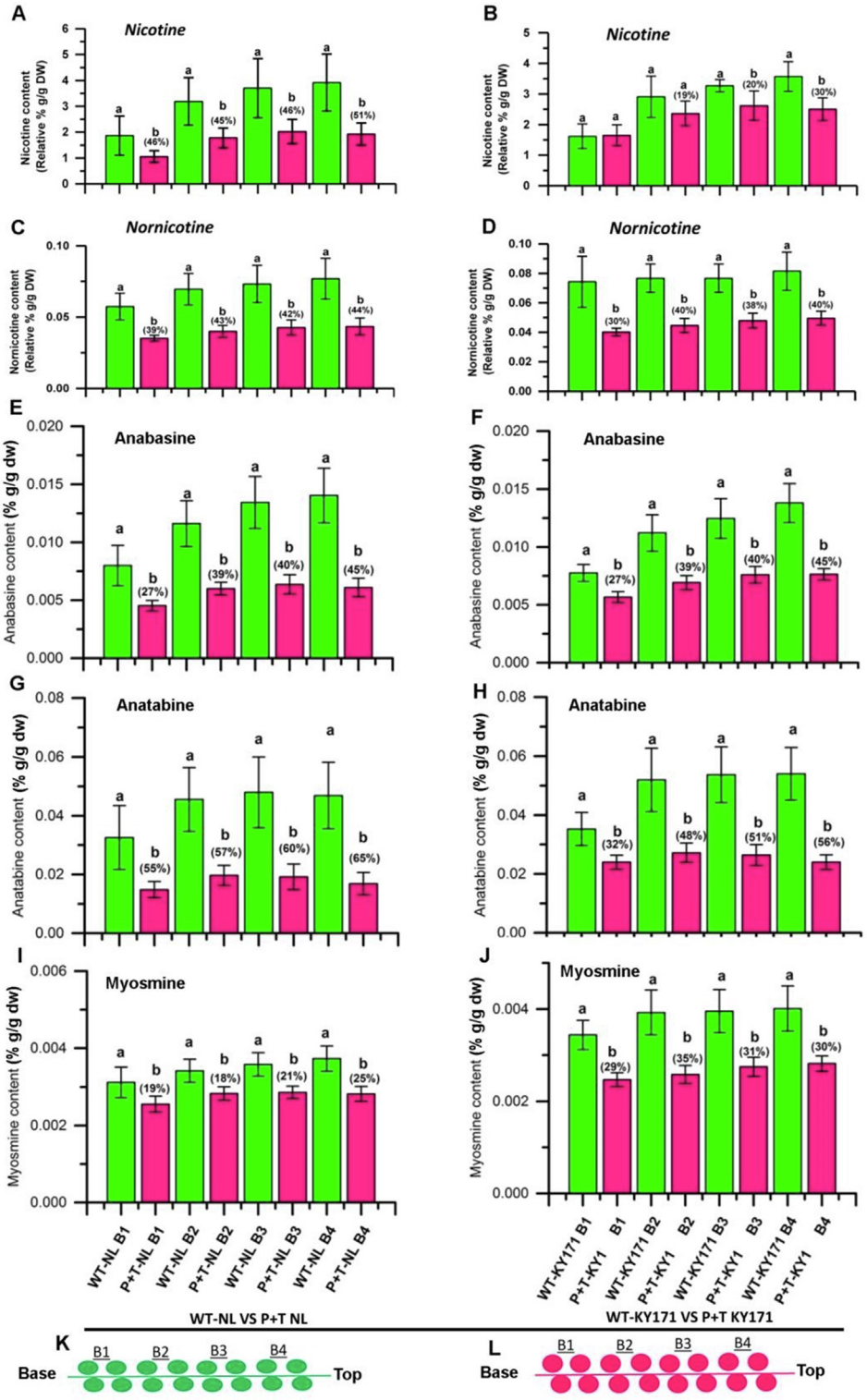


Fig. 6. Reduction of nicotine and four other alkaloids in cured leaves of red P + T-NL1 and P + T-KY1 tobacco plants. A-J, reduction of nicotine (A-B), nornicotine (C-D), anabasine (E-F), anatabine (G-H), and myosmine (I-J) in four leaf groups of P + T-NL (A, C, E, G, and I) and P + T-KY plants (B, D, F, H, and J). K-L, cartoons showing leaf groups harvested from both wild type (K) and transgenic red plants (L). Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of tobacco plant, respectively. Each metabolite in each leaf group was quantified from leaves harvested from 120 tobacco plants in three plots. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Bars labelled with lowercase "a" and "b" indicate significant difference between paired green and red bars (n = 120, P-value less than 0.05). Percentages labeled in the top of red bars mean reduction compared to green bars. The error bars indicate standard deviation. Abbreviations, WT-NL: wild-type Narrow Leaf Madole (NL) variety, P + T-NL1: Stacked *PAP1* and *TT8* transgenic NL line 1, WTKY: wild-type KY171 variety, and P + T-KY1: stacked *PAP1* and *TT8* transgenic KY line 1.

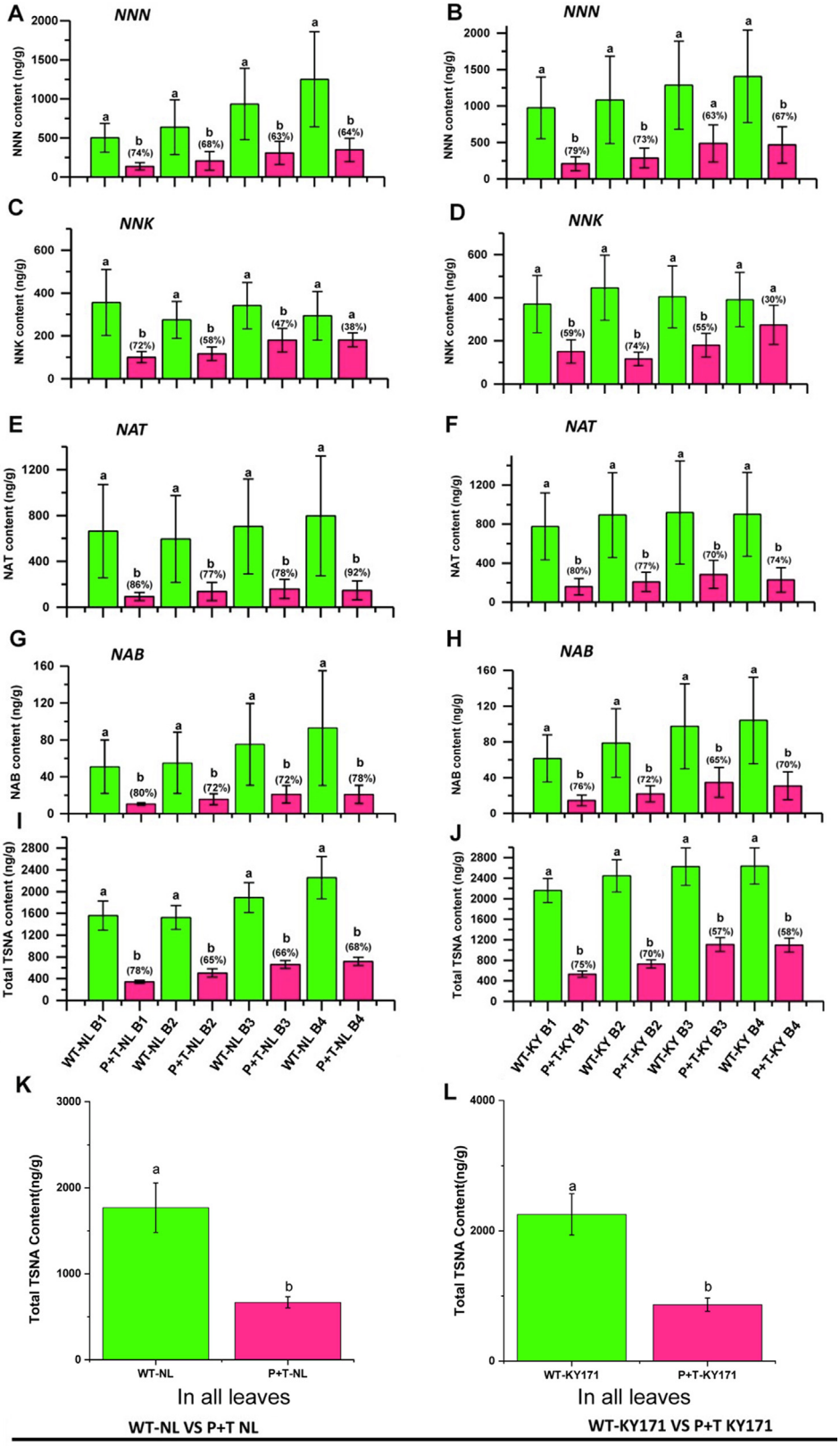


Fig. 7. Reduction of four individual and total TSNAs in leaves of red P + T-NL1 and P + T-KY1 tobacco plants. A-H, reduction of NNN (A-B), NNK (C-D), NAT (E-F), and NAB (G-H) in each leaf group of P + T-NL1 (A, C, E, and G) and P + TKY1 (B, D, F, and H) plants; I-J, reduction of total TSNA contents summed from four individual TSNAs of each leaf group of P + T-NL1 (I) and P + T-KY1 (J) plants; K-L, reduction of total TSNA contents summed from four individual TSNAs of all leaves of P + T-NL1 (K) and P + T-KY1 (L) plants. Green and red bars represent wild type and red tobacco plants, respectively. B1, B2, B3, and B4 labels represent the 1st, 2nd, 3rd, and 4th group of leaves from the base to the top of plants, respectively. Each metabolite in each group of leaves was quantified from 120 tobacco plants. One-way ANOVA test included in the SAS software was used to evaluate statistical significance. Paired green and red bars labelled with lowercase ''a'' and ''b'' indicate significant difference (n = 120, P-value less than 0.05). Percentages labeled on the top of red bars indicate reduction levels compared to green bars. Abbreviation, WT-NL: Narrow Leaf Madole, WT-KY171: wild type KY171, P + T-NL: stacked PAP1 and TT8 transgenic NL line 1, and P + T-KY1: stacked PAP1 and TT8 transgenic KY171 line 1.

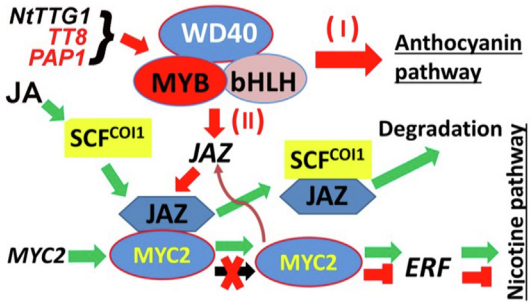


Fig. 8. A model showing the distant pathway-cross regulation (DPCR) created for the downregulation of the biosynthesis of tobacco alkaloids in red tobacco plants. Regulation pathways created in red tobacco plants are featured by red colored solid arrows, "X" shape, and "T" blocks. The stacked *PAP1* and *TT8* as well as the endogenous *NtTTG1* encode MYB, bHLH, and WD40 proteins, respectively, which form a complex to regulate two biosynthetic pathways. The regulation pathway (function I) is the activation of biosynthesis of anthocyanin. The regulation pathway (function II) is the DPCR that the MYB-bHLH-WD40 complex upregulates the endogenous *NtJAZs* leading to the reduction of MYC2 and the downregulation of *EAR* and the nicotine biosynthesis. In addition, The JA-JAZ-MYC regulatory pathway is indicated with green arrows starting with JA (greenish arrows). The feedback regulation of NtMYC2 indicated by a brownish bending arrow is proposed to activate the expression of NtJAZs. SCF: Skp/Cullin/F-box and COI1: coronatine insensitive 1.

ingly, this definition distinguishes from the master regulatory function of PAP1, TT8, and their MBW complex that activates the anthocyanin biosynthesis in Arabidopsis [8,14,73,74] or red tobacco shown here via directly activating most or late pathway genes, such as *DFR*, *ANS*, and *3-GT* (Fig. S10). Based on these data, we propose a model to characterize the DPCR created from the *de novo* design and the regulation of tobacco alkaloid biosynthesis in red tobacco plants (Fig. 8). Given that *NtTTG1* is expressed in roots, leaves, and flowers, this model proposes that PAP1 and TT8 likely anchor to the endogenous tobacco NtTTG1 (WD40) to form a heterogeneous MBW complex, which has a dual regulatory function that regulates tobacco metabolism. First, the MBW complex directly activates the anthocyanin biosynthesis (function I), which has been intensively elucidated. Second, the MBW complex upregulates four *NtJAZ1*, *NtJAZ3*, *NtJAZ7*, and *NtJAZ10*, which increase NtJAZ repressors (function II). Next, the increased NtJAZs bind to more MYC2 and block the release of this TF. Then, the reduction of free MYC2 results in downregulating *EFR* and six or seven nicotine pathway genes (Fig. 3). Finally, the biosynthesis of nicotine, OTAs, and four TSNAs is significantly diminished in red tobacco plants (Figs. 4, 6, and 7). In addition, based on the report that AtMYC2 solely regulates the *AtJAZ* transcription via a feed-back regulation [75], this model proposes that PAP1/TT8 and NtMYC2 co-activate the expression of NtJAZs in red tobacco plants. When JA is absent, PAP1 and TT8 play the main regulatory role to activate the expression of four *NtJAZs*. To date, in addition to tobacco, PAP1 and its homologs have been introduced to a few crops to engineer anthocyanins for value-added traits, such as tomato [22,23], hop [24], canola [25], and rose [76]. Although whether and how PAP1 and TT8 can regulate non-anthocyanin or non-flavonoid pathways in these plants remains unknown, it is interesting that PAP1 transgenic rose plants have been reported to produce more than 6.5-time scent terpenoids [76]. Based on our findings, it is interesting to investigate whether a similar DPCR occurs to increase scent terpenoids in rose in the future. Moreover, different TF families are a large part of plant genomes [77–80], our findings suggest that it is necessary to investigate DPCR events of other TFs to fully understand their regulatory functions. In particular, as shown in our data that carcinogenic TSNAs are significantly reduced in red tobacco plants, investigations of potential DPCR of TFs likely create value-added traits for crop improvement and human health.

This *de novo* regulation design with PAP1/TT8 TFs and *NtJAZs* shows a promising application in enhancing efforts for the simultaneous reduction of most harmful tobacco alkaloids. As introduced above, in the past decades, multiple studies have been undertaken to develop technologies to reduce harmful tobacco alkaloids, such as gene suppression or silencing technologies with anti-sense or RNAi of pathway genes (Fig. 1 B), including *PMT* [65], *ODC* [64], *QPT* [63], *NDM* [81], A622 [82], *BBL* [49], and others [42]. On the one hand, most studies successfully reported the decrease of the contents of nicotine, OTAs, or TSNAs in tobacco plants grown in the greenhouse, growth chambers, or field. On the other hand, these previous successes revealed a challenge that all tobacco alkaloids could not be diminished simultaneously. For example, the decrease of nicotine led to either the increase of anatabine or non-reduction of other tobacco alkaloids [64,65]. In addition, the reduction of harmful compounds was relatively limited. For instance, an RNAi of A622 could mainly reduce NNN and total TSNAs [81]. A gene silencing of *BBL* was reported to mainly decrease the nicotine content [49]. Furthermore, all these successes have shown that the reduction of nicotine and other alkaloids to a practical level requires more studies, such as elucidation of metabolic regulation and unknown genes. As intensively characterized, the regulation of the nicotine biosynthesis is via three cohort layers, the presence/absence of an active JA form, the JA and JAZ signaling pathway, and the release of MYC2 for activation of ERF189 and pathway genes [42,48,55] (Fig. 1B). To date, an important unanswered question is how tobacco plants regulate *NtJAZs*. In Arabidopsis, the expression of *AtJAZs* is solely activated by AtMYC2 via a feed-back regulation [75]. Based on this, although NtMYC2 members have not been reported to regulate the transcription of *NtJAZs* in tobacco plants, it can be postulated that NtMYC2 members may also perform a feed-back regulation of *NtJAZs* (Fig. 8). Our *de novo* design with PAP1 and TT8 and MRE and G-box elements was used to target four *NtJAZs*. As designed, PAP1, TT8, PAP1 and TT8 together, and their MBW complex bound and activated promoters of the four NtJAZs (Fig. 2) and the stacked overexpression of *PAP1* and *TT8* led to the upregulation of four *NtJAZs* (Fig. 3A,B and K,L, Fig. S13). As the positive result, the biosynthesis of nicotine was downregulated in red P + T genotypic plants (Figs. 3 and 4). The contents of nicotine, four OTAs, and four TSNAs were simultaneously reduced in most or all cured leaf groups and entire plant leaves of P-T-NL, P-T-KY, and PAP1 genotypes (Figs. 6 and 7, Figs. S20). The NNN contents were particularly diminished to a level less than 0.55 ppm in most of red tobacco leaves (Fig. 7 A-B). This low content is significant to human health, because NNN is a severe tobacco carcinogen and its content in any smokeless tobacco products is limited to 1.00 pm proposed by FDA. In addition, other TSNAs were reduced to a level lower than 0.5 ppm in most or all leaves of red P + T genotypic tobacco plants. To further substantiate the function of this *de novo* regulation design, we tested our PAP1 red tobacco variety, in which the expression of *PAP1* activated two *NtTT8* homologs (Fig. S21). Although this variety was created from Xanthi, a low-nicotine and flue-cured cultivar, the two years' field trials showed the reduction of nicotine, OTAs, and TSNAs in PAP1 plants (Figs. S23-24 and S25-26). These data indicate that PAP1 can partner with the endogenous tobacco NtTT8 homologs to form a DPCR to downregulate the nicotine biosynthesis. Taken together, these findings provide a new platform to eliminate tobacco carcinogens to a non-harmful level and to alter plant metabolism in other crops for novel traits.

The plant kingdom has a large number of TFs that regulate plant development, structure, responses to stresses, and plant metabolisms. In 2000, when the genome of *Arabidopsis thaliana* was sequenced [83], gene annotation and sequence analysis revealed more than 1500 TFs (accounting for more than 5% of annotated

genes). Since then, more TFs have been reported from numerous sequenced plant genomes, such as rice [84,85]and soybean[86], and multiple browser servers have been developed to search TF families and their regulatory functions [87–93]. By 2017, 156 plant genomes were sequenced, including 16 from Chlorophytae, one from Charophyta, one from Marchantiophyta, two from Byrophyta, one from Lycopodiophyta, two from Coniferophyta, one from Basa Magnoliophyta, 38 from monocots, and 95 Eudicots. [67]. These 156 genomes together with sequences of nine additional plant species allowed the annotation of 320,370 plant TFs, which were grouped into 58 families [67]. Likewise, the plant genomes have a large number of binding sites (motifs) (BS), such as G-box and MRE in promoters, to which TFs bind. In recent, a public PlantRegMap server was developed from 63 plant species (out of 156 genome-sequenced plants) to search TFBSs and the interactions between TF and BS [68]. This PlantRegMap characterized a conserved regulatory landscape that included 21,997,501 TFBSs, 21,346 TFs, and more than two million interactions between them [68]. It can be speculated that as more plant species will be sequenced, the number of plant TFs will be increased. To our knowledge, no TF or BS has been used as a molecular tool for a novel regulatory design to engineer better valuable traits in crops. Based on our first proof-of-concept study with PAP1 vs. MRE and TT8 vs. G-box that are effective molecular tools to design novel regulation modes to create a DPCR in tobacco plant, it can be speculated that the large number of plant TFs and BSs provides an extremely rich molecular tool source for novel regulatory designs, which will be fundamental for plant engineering.

Conclusion

The findings show that PAP1, TT8, MREs, and G-box elements are useful molecular tools to design a *De Novo* regulation, a distant pathway-cross regulation (DPCR) of plant secondary metabolism in plants. The findings unearth novel regulatory functions of PAP1 and TT8 that are two positive regulators of *NtJAZs* and negatively regulate tobacco alkaloid biosynthesis in red tobacco plants. This *De Novo* regulation design significantly reduces harmful nicotine, other tobacco alkaloids, and all TSNAs in all or most leaves of tobacco plants. Our findings indicate that the great number of TFs and BSs provides an extremely rich molecular tool source for novel regulatory designs, which will be fundamental in plant engineering for value-added crops and products.

Credit Author Statement

De-Yu Xie: perceived and designed the entire project, developed all experimental plans, selected field, supervised analysis of all data and figure preparation, and drafted and finalized the manuscript. **Mingzhuo Li:** contributions to experiments of EMSA, dual-luciferase, ChIP, gene expression profiling, promotor cloning, sequencing, identification of *cis*-elements, genetic transformation, PCR, qRT-PCR, anthocyanin analysis, proanthocyanidin analysis, analysis of all these experimental data, greenhouse experiments, HPLC-MS analysis, figure preparation, and manuscript drafting. **Xianzhi He:** contributions to gene stacking design, gene synthesis, construct development, genetic transformation, and selection of P + T-transgenic plants, screening of T1 and homozygous T2 progeny, waterbed nursery of P + T seedlings in greenhouse, field preparation, planting, management of plant growth, plant harvest, air-curing, and sample preparation for HPLC-MS analysis, sample preparation for HPLC-QQQ-MS/MS and GC-FID experiments, analysis of nicotine, nornicotine, and TSNAs, figure preparation, and manuscript drafting. **Christophe La Hovary:** contributions to waterbed nursery of PAP1 seedlings, field preparation, planting, management of plant growth, plant harvest, flue-curing, HPLC-ESI-MS, analysis of nicotine, nornicotine, and TSNAs in PAP1 plants, data analysis, figure preparation, and manuscript preparation. **Yue Zhu:** analysis of nicotine and nornicotine in seedlings. **Yilun Dong:** extraction of RNA and performance of qRT-PCR. **Shibiao Liu**: **Hucheng Xing, Yajun Liu**, **Yucheng Jie, Dongming Ma, and Seyit Yuzuak:** contributions to selection of T2 progeny, greenhouse experiments, planting, plant harvest, and collection of field data, sampling, sample preparation for metabolite analysis, and discussion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

R.J. Reynolds Tobacco Company provided funding (#571426) and analytical support for metabolite analysis of field samples in this project. We are grateful to Dr. Jannell Rowe for analyzing nicotine, TSNAs and other alkaloids in P + T-NL and P + T-KY samples. We are thankful to our colleague professor Ramsey Lewis for providing us seeds of KY171 and Narrow Leaf Madole cultivars from the tobacco germplasm bank at NCSU and for suggestions about field trial regulation of transgenic tobacco plants and leaf curing. We are grateful to staff at Research Station in Oxford, North Carolina for helping planting, sucker control, irrigation, and field fertilization as well as providing both flue-curing barns and air-curing facility. We are thankful to Drs. Emmett Hiatt and Dr. Jannell Rowe from R. J. Reynolds for critically reading this manuscript. We are grateful to Professors Ian Mackay and Zhao-Bang Zeng for advising statistical analysis.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2021.06.017.

References

[1] Du H, Feng BR, Yang SS, Huang YB, Tang YX. The R2R3-MYB Transcription Factor Gene Family in Maize. PLoS ONE 2012;7(6):12.
[2] Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci 2010;15(10):573–81.
[3] Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 2011;66 (1):94–116.
[4] Wei K, Chen H. Comparative functional genomics analysis of bHLH gene family in rice, maize and wheat. BMC Plant Biol 2018;18(1):309.
[5] Saigo T, Wang T, Watanabe M, Tohge T. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. Curr Opin Plant Biol 2020;55:93–9.
[6] Shi M-Z, Xie D-Y. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. Recent Pat Biotechnol 2014;8(1):47–60.
[7] Springob K, Nakajima H, Yamazaki M, Saito K. Recent advances in the biosynthesis and accumulation of anthocyanins. Nat Prod Rep 2003;20:288–303.
[8] Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 2000;12 (12):2383–94.
[9] Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. Plant Cell 2000;12:1863–78.
[10] Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasdan N, Blundell TL, et al. The *TRANSPARENT TESTA GLABRA 1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. Plant Cell 1999;11:1377–11349.
[11] Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, et al. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB–bHLH–WDR complexes and their targets in Arabidopsis seed. New Phytol 2014;202 (1):132–44.

[12] Thévenin J, Dubos C, Xu W, Le Gourrierec J, Kelemen Z, Charlot F, et al. A new system for fast and quantitative analysis of heterologous gene expression in plants. New Phytol 2012;193(2):504–12.
[13] Xu WJ, Grain D, Le Gourrierec J, Harscoet E, Berger A, Jauvion V, et al. Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression. Arabidopsis New Phytol 2013;198(1):59–70.
[14] Zhu ZX, Wang HL, Wang YT, Guan S, Wang F, Tang JY, et al. Characterization of the cis elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. J Exp Bot 2015;66(13):3775–89.
[15] Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 2004;39 (3):366–80.
[16] Ramsay NA, Glover BJ. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci 2005;10(2):63–70.
[17] Shi MZ, Xie DY. Engineering of red cells of *Arabidopsis thaliana* and comparative genome-wide gene expression analysis of red cells versus wild-type cells. Planta 2011;233(4):787–805.
[18] Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, et al. The jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. Plant Cell 2011;23(5):1795–814.
[19] Qi TC, Huang H, Wu DW, Yan JB, Qi YJ, Song SS, et al. Arabidopsis DELLA and JAZ proteins bind the WD-Repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. Plant Cell 2014;26(3):1118–33.
[20] He X, Li Y, Lawson D, Xie D-Y. Metabolic engineering of anthocyanins in dark tobacco varieties. Physiol Plant 2017;159:2–12.
[21] Xie D-Y, Sharma SB, Wright E, Wang Z-Y, Dixon RA. Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J 2006;45(6):895–907.
[22] Butelli E, Titta L, Giorgio M, Mock H-P, Matros A, Peterek S, et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotech 2008;26(11):1301–8.
[23] Zuluaga DL, Gonzali S, Loreti E, Pucciariello C, Degl'Innocenti ED, Guidi L, et al. Arabidopsis thaliana MYB75/PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. Funct Plant Biol 2008;35(7):606–18.
[24] Gatica-Arias A, Farag MA, Stanke M, Matousek J, Wessjohann L, Weber G. Flavonoid production in transgenic hop (*Humulus lupulus* L.) altered by PAP1/MYB75 from *Arabidopsis thaliana* L. Plant Cell Rep 2012;31(1):111–9.
[25] Li X, Gao MJ, Pan HY, Cui DJ, Gruber MY. Purple canola: Arabidopsis PAP1 increases antioxidants and phenolics in *Brassica napus* leaves. J Agric Food Chem 2010;58(3):1639–45.
[26] Ma DM, Gandra SVS, Manoharlal R, La Hovary C, Xie DY. Untargeted metabolomics of *Nicotiana tabacum* grown in United States and India characterizes the association of plant metabolomes wiith natural climate and geography. Front Plant Sci 2019;10:18.
[27] Nugroho LH, Verpoorte R. Secondary metabolism in tobacco. Plant Cell Tissue Organ Cult 2002;68(2):105–25.
[28] Rodgman A, Perfetti TA. The Chemical Components of Tobacco and Tobacco Smoke, CRC Press. Boca Raton, London, New York: Taylor & Francis Group; 2009.
[29] Arany I, Taylor M, Fulop T, Dixit M. Adverse effects of chronic nicotine exposure on the kidney: Potential human health implications of experimental findings. Int J Clin Pharmacol Ther 2018;56(11):501–6.
[30] Grando SA. Connections of nicotine to cancer. Nat Rev Cancer 2014;14 (6):419–29.
[31] Greillier L, Cortot AB, Viguier J, Brignoli-Guibaudet L, Lhomel C, Eisinger F, et al. Perception of Lung Cancer Risk: Impact of Smoking Status and Nicotine Dependence. Current Oncol Rep 2018;20:7.
[32] Lee PN, Forey BA, Hamling JS, Thornton AJ. Environmental tobacco smoke exposure and heart disease: A systematic review. World J Meta-Anal 2017;5 (2):14–40.
[33] Sanner T, Grimsrud TK. Nicotine: carcinogenicity and effects on response to cancer treatment - a review. Front Oncol 2015;5:10.
[34] Santoro A, Tomino C, Prinzi G, Lamonaca P, Cardaci V, Fini M, et al. Tobacco Smoking: Risk to Develop Addiction, Chronic Obstructive Pulmonary Disease, and Lung Cancer, Recent Patents on Anti-Cancer. Drug Discovery 2019;14 (1):39–52.
[35] Thorndike AN, Rigotti NA. A tragic triad: coronary artery disease, nicotine addiction, and depression. Curr Opin Cardiol 2009;24(5):447–53.
[36] Tidey JW, Davis DR, Miller ME, Pericot-Valverde I, Denlinger-Apte RL, Gaalema DE. Modeling nicotine regulation: A review of studies in smokers with mental health conditions. Prev Med 2018;117:30–7.
[37] Hecht SS. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. Chem Res Toxicol 1998;11(6):559–603.
[38] Hecht SS, Stepanov I, Carmella SG. Exposure and Metabolic Activation Biomarkers of Carcinogenic Tobacco-Specific Nitrosamines. Acc Chem Res 2016;49(1):106–14.
[39] Konstantinou E, Fotopoulou F, Drosos A, Dimakopoulou N, Zagoriti Z, Niarchos A, et al. Tobacco-specific nitrosamines: A literature review. Food Chem Toxicol 2018;118:198–203.
[40] FDA, Tobacco Product Standard for N-Nitrosonornicotine Level in Finished Smokeless Tobacco Products, in: H.A.H. SERVICES (Ed.) Food and Drug Administration, 2017.
[41] Dawson RF. An experimental analysis of alkaloid production in *Nicotiana* - The origin of nornicotine. Am J Bot 1945;32(7):416–23.
[42] Dewey RE, Xie JH. Molecular genetics of alkaloid biosynthesis in Nicotiana tabacum. Phytochemistry 2013;94:10–27.
[43] Ladesic B, Tso TC. Biochemical studies on tobacco alkaloids. 6. Biosynthesis of nicotine through nornicotine. Phytochemistry 1964;3(4):541–5.
[44] Mizusaki S, Tanabe Y, Noguchi M, Tamaki E. Phytochemical studies on tobacco alkaloids. 16. Changes in activities of ornitheine decarboxylase, putrescine N-methyltransferase and N-methyl-putrescine oxidase in tobacco roots in relation to nicotine biosynthesis. Plant Cell Physiol 1973;14(1):103–10.
[45] Stepka W, Dewey LJ. Conversion of nicotine to nornicotine in harvested tobacco - fate of methyl group. Plant Physiol 1961;36(5):592–1000.
[46] Tso TC, Jeffrey RN. Studies on tobacco alkaloids. 1. Changes in nicotine and nornicotine content in *Nicotiana*. Plant Physiol 1956;31(6):433–9.
[47] Tso TC, Jeffrey RN. Studies on tobacco alkaloids. 2. The formation of nicotine and nornicotine in tobacco supplied with N15. Plant Physiol 1957;32 (2):86–92.
[48] Shoji T, Hashimoto T. Smoking out the masters: transcriptional regulators for nicotine biosynthesis in tobacco. Plant Biotechnol 2013;30(3):217–24.
[49] Lewis RS, Lopez HO, Bowen SW, Andres KR, Steede WT, Dewey RE. Transgenic and mutation-based suppression of a berberine bridge enzyme-like (BBL) gene family reduces alkaloid content in field-grown tobacco. PLoS ONE 2015;10(2).
[50] Chakrabarti M, Bowen SW, Coleman NP, Meekins KM, Dewey RE, Siminszky B. CYP82E4-mediated nicotine to nornicotine conversion in tobacco is regulated by a senescence-specific signaling pathway. Plant Mol Biol 2008;66 (4):415–27.
[51] Gavilano LB, Siminszky B. Isolation and characterization of the cytochrome P450 gene CYP82E5v2 that mediates nicotine to nornicotine conversion in the green leaves of tobacco. Plant Cell Physiol 2007;48(11):1567–74.
[52] Lewis RS, Bowen SW, Keogh MR, Dewey RE. Three nicotine demethylase genes mediate nornicotine biosynthesis in *Nicotiana tabacum* L.: Functional characterization of the CYP82E10 gene. Phytochemistry 2010;71(17–18):1988–98.
[53] De Boer K, Tilleman S, Pauwels L, Vanden Bossche R, De Sutter V, Vanderhaeghen R, et al. APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. Plant J 2011;66(6):1053–65.
[54] Shoji T, Hashimoto T. Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by way of the NIC2-locus ERF genes. Plant Cell Physiol 2011;52(6):1117–30.
[55] Zhang HB, Bokowiec MT, Rushton PJ, Han SC, Timko MP. Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. Mol Plant 2012;5(1):73–84.
[56] Baldwin IT, Zhang ZP, Diab N, Ohnmeiss TE, McCloud ES, Lynds GY, et al. Quantification, correlations and manipulations of wound-induced changes in jasmonic acid and nicotine in *Nicotiana sylvestris*. Planta 1997;201: 397–404.
[57] Imanishi S, Hashizume K, Nakakita M, Kojima H, Matsubayashi Y, Hashimoto T, et al. Differential induction by methyl jasmonate of genes encoding ornithine decarboxylase and other enzymes involved in nicotine biosynthesis in tobacco cell cultures. Plant Mol Biol 1998;38(6):1101–11.
[58] Shoji T, Yamada Y, Hashimoto T. Jasmonate induction of putrescine N-methyltransferase genes in the root of *Nicotiana sylvestris*. Plant Cell Physiol 2000;41(7):831–9.
[59] Xu BF, Timko MP. Methyl jasmonate induced expression of the tobacco putrescine N-methyltransferase genes requires both G-box and GCC-motif elements. Plant Mol Biol 2004;55(5):743–61.
[60] Shoji T, Ogawa T, Hashimoto T. Jasmonate-induced nicotine formation in tobacco is mediated by tobacco COI1 and JAZ genes. Plant Cell Physiol 2008;49 (7):1003–12.
[61] Zhang HY, Li WJ, Niu DX, Wang ZJ, Yan XX, Yang XL, et al. Tobacco transcription repressors NtJAZ: Potential involvement in abiotic stress response and glandular trichome induction. Plant Physiol Biochem 2019;141:388–97.
[62] DeBoer KD, Dalton HL, Edward FJ, Ryan SM, Hamill JD. RNAi-mediated down-regulation of ornithine decarboxylase (ODC) impedes wound-stress stimulation of anabasine synthesis in *Nicotiana glauca*. Phytochemistry 2013;86:21–8.
[63] Xie J, Maksymowicz SWW, Wx JWCKC, Ma CCKJC. Biotechnology: A tool for reduced risk tobacco products - The nicotine experience from test tube to cigarette pack, Recent Advances in Tobacco. Science 2004:17–37.
[64] DeBoer KD, Dalton HL, Edward FJ, Hamill JD. RNAi-mediated down-regulation of ornithine decarboxylase (ODC) leads to reduced nicotine and increased anatabine levels in transgenic *Nicotiana tabacum* L. Phytochemistry 2011;72 (4–5):344–55.
[65] Chintapakorn Y, Hamill JD. Antisense-mediated down-regulation of putrescine N-methyltransferase activity in transgenic *Nicotiana tabacum* L.can lead to elevated levels of anatabine at the expense of nicotine. Plant Mol Biol 2003;53:87–105.
[66] Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among Eukaryotes. Science 2000;290:2105–10.
[67] Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2016;45(D1):D1040–5.

[68] Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G. PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res 2019;48(D1):D1104–13.
[69] Li MZ, Li YZ, Guo LL, Gong ND, Pang YZ, Jiang WB, et al. Functional characterization of tea (*Camellia sinensis*) MYB4a transcription factor using an integrative approach. Front Plant Sci 2017;8:17.
[70] Wang P, Liu Y, Zhang L, Wang W, Hou H, Zhao Y, et al. Functional demonstration of plant flavonoid carbocations proposed to be involved in the biosynthesis of proanthocyanidins. Plant J 2020;101(1):18–36.
[71] Xie D-Y, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science 2003;299(5605):396–9.
[72] Bai YH, Pattanaik S, Patra B, Werkman JR, Xie CH, Yuan L. Flavonoid-related basic helix-loop-helix regulators, NtAn1a and NtAn1b, of tobacco have originated from two ancestors and are functionally active. Planta 2011;234(2):363–75.
[73] Baudry A, Caboche M, Lepiniec L. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J 2006;46(5):768–79.
[74] Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J 2008;53(5):814–27.
[75] Chini A, Fonseca S, Fernandez G, Adie B, Chico JM, Lorenzo O, et al. The JAZ family of repressors is the missing link in jasmonate signalling. Nature 2007;448(7154):666-+.
[76] Zvi MMB, Shklarman E, Masci T, Kalev H, Debener T, Shafir S, et al. PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. New Phytol 2012. doi: https://doi.org/10.1111/j.1469-8137.2012.04161.x.
[77] Feng K, Hou XL, Xing GM, Liu JX, Duan AQ, Xu ZS, et al. Advances in AP2/ERF super-family transcription factors in plant. Crit Rev Biotechnol 2020;40(6):750–76.
[78] Millard PS, Kragelund BB, Burow M. R2R3 MYB transcription factors - functions outside the DNA-binding domain. Trends Plant Sci 2019;24(10):934–46.
[79] Pireyre M, Burow M. Regulation of MYB and bHLH transcription factors: a glance at the protein level. Mol Plant 2015;8(3):378–88.
[80] M.K. Skinner, A. Rawls, J. Wilson-Rawls, E.H. Roalson, Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature, Differentiation; research in biological diversity 80(1) (2010) 1-8.
[81] Lewis RS, Jack AM, Morris JW, Robert VJM, Gavilano LB, Siminszky B, et al. RNA interference (RNAi)-induced suppression of nicotine demethylase activity reduces levels of a key carcinogen in cured tobacco leaves. Plant Biotechnol J 2008;6(4):346–54.
[82] DeBoer KD, Lye JC, Aitken CD, Su AKK, Hamill JD. The A622 gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Mol Biol 2009;69(3):299–312.
[83] Initiative TAG. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000;408:796–815.
[84] Yu J. A draft sequuence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 2002;296:79–92.
[85] Goff SA. A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002;296:92–100.
[86] Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. PNAS 2010;107(51):22032–7.
[87] Guo AY, Chen X, Gao G, Zhang H, Zhu QH, Liu XC, et al. PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res 2008;36:D966–9.
[88] Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res 2013;42(D1):D1182–7.
[89] Guo AY, He K, Liu D, Bai SN, Gu XC, Wei LP, et al. DATF: a database of Arabidopsis transcription factors. Bioinformatics 2005;21(10):2568–9.
[90] Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, et al. RARTF: Database and tools for complete sets of Arabidopsis transcription factors. DNA Res 2005;12(4):247–56.
[91] Gao G, Zhong YF, Guo AY, Zhu QH, Tang W, Zheng WM, et al. DRTF: a database of rice transcription factors. Bioinformatics 2006;22(10):1286–7.
[92] Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B. PlnTFDB: an integrative plant transcription factor database. BMC Bioinf 2007;8:10.
[93] Fredslund J. DATFAP: A database of primers and homology alignments for transcription factors from 13 plant species. BMC Genomics 2008;9:10.