



## Features and genetic basis of chlorogenic acid formation in diploid potatoes

Huiqin Yang<sup>a,1</sup>, Qinggang Liao<sup>b,1</sup>, Ling Ma<sup>a,1</sup>, Wei Luo<sup>a</sup>, Xiaofeng Xiong<sup>a</sup>, Youwei Luo<sup>c</sup>, Xiong Yang<sup>d</sup>, Chunyong Du<sup>e</sup>, Yufeng He<sup>f</sup>, Xiaolei Li<sup>g</sup>, Dongli Gao<sup>a</sup>, Xiaofeng Xue<sup>h,\*</sup>, Yi Shang<sup>a,\*</sup>

<sup>a</sup> Yunnan Key Laboratory of Potato Biology, The CAAS-YNNU-YINMORE Joint Academy of Potato Sciences, Yunnan Normal University, Kunming 650500, China

<sup>b</sup> China Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518116, China

<sup>c</sup> Center for Agricultural Technology Extension of Dehong State, Mangshi 678400, China

<sup>d</sup> Dali Institute of Agricultural Sciences, Dali 671005, China

<sup>e</sup> Agricultural Sciences and Technology Extend Station of Liushao Town, Xundian 655200, China

<sup>f</sup> Yunnan Lian-Da Science Technology Development Co., Ltd., Yunnan Normal University, Kunming 650500, China

<sup>g</sup> Research Center for Analysis and Measurement, Kunming University of Science and Technology, Kunming 652094, China

<sup>h</sup> Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, China

### ARTICLE INFO

#### Keywords:

Diploid potato  
Chlorogenic acid  
Customer preference  
Genome-wide association study

### ABSTRACT

The concentration of chlorogenic acids (CGAs), is tightly associated with the appearance, taste, and nutrient content of potato tubers. Manipulation of tuber CGA concentrations allows for the breeding of quality traits in potatoes. Currently, a hybrid potato breeding system that aims to convert tetraploid potato into a diploid seed crop represents a new development in potato breeding. Unfortunately, however, a systematic study of CGA formation is very limited in diploid potatoes. Here, using a diverse panel of diploid potatoes, including 40 ancestors and 374 landraces, we analyzed the influence of location, environment, genetic basis, as well as expression of enzymes, in affecting the CGA concentrations in diploid lines. We revealed a selection of the decreased CGA level of tuber flesh in the domestication of diploid potatoes. Moreover, we identified 18 SNPs associated with tuber CGA levels using re-sequenced genome data. This study provides a basis for the breeding of high-quality potato by taking into consideration customer preferences.

### 1. Introduction

Potatoes (*Solanum tuberosum* L.) are grown in ~ 125 countries and are consumed as the third most important staple food by over 1 billion people (Devaux, Goffart, Petsakos, Kromann, & Hareau, 2020). Potato tubers provide not only large amounts of carbohydrates as an energy source, but also nutrient-rich elements including minerals, vitamins, and fiber to human diets. Potatoes represent the largest supply of vegetable phenolics and antioxidants in the American diet (Song et al., 2010). Unfortunately, quality traits such as micronutrients and flavors are commonly neglected and are often lost in breeding processes that prioritize high yield (Sands, Morris, Dratz, & Pilgeram, 2009; Tieman et al., 2017). This can result in consumer complaints and unmet market demands. Thus, potato breeders need to develop elite cultivars with high-quality traits and more desirable flavors.

The flavor, nutrient content, and disease resistance of crops is commonly associated with a vast array of specialized metabolites, such as phenolics, terpenoids, and alkaloids. These compounds are biosynthesized by plants in order to mediate the interactions with the constantly changing environment. In potato tubers, chlorogenic acids (CGAs), ester of caffeic acid and quinic acid, represent over 80% of all phenolic compounds (Friedman, 1997a; Im et al., 2008; Riciputi et al., 2018; Valinas, Lanteri, Have, & Andreu, 2017). They exist in three major isomers, CGA (5-O-caffeoylquinic acid, the predominant isomer), neochlorogenic acid (3-O-caffeoylquinic acid), and cryptochlorogenic acid (4-O-caffeoylquinic acid), respectively (Maldonado, Mudge, Gänzle, & Schieber, 2014). Given their high antioxidant activities, CGAs exhibit potential health benefits including anti-inflammatory, anti-diabetic, anti-carcinogenic, and anti-obesity effects, which can prevent cardiovascular and other degenerative diseases (Onakpoya, Spencer,

\* Corresponding authors.

E-mail addresses: [xue\\_xiaofeng@126.com](mailto:xue_xiaofeng@126.com) (X. Xue), [shangyi@ynnu.edu.cn](mailto:shangyi@ynnu.edu.cn) (Y. Shang).

<sup>1</sup> These authors contributed equally.

Thompson, & Heneghan, 2015; Tajik, Tajik, Mack, & Enck, 2017). Thus, potato is an ideal crop to provide natural dietary antioxidants for human consumption. In addition, as CGAs are produced by plants to improve their resistance against pathogens, and/or oxidative stress (Harrison et al., 2008; Wojciechowska et al., 2014); Niggeweg, Michael, & Martin, 2004), a high concentration of CGAs in tuber peel and other tissues would likely confer native disease resistance for growing potatoes. However, as CGA levels are positively correlated with total phenolic compounds including flavonoids in potato tubers (Ru, Pang, Gan, Liu, & Bao, 2019; Valinas et al., 2017), higher contents of CGAs are usually observed in colored potatoes which can negatively influence a customer's preference. Moreover, over accumulation of CGAs in tubers (>120 mg/100 g) may induce unwanted sourness and bitterness, severely affecting the flavors of the potato tuber (Jansky, 2010). Since the concentration of chlorogenic acids (CGAs), is tightly associated with the appearance, taste, and nutrient content of potato tubers, CGA levels should be controlled to provide bioactive nutrients, but not to detrimentally influence tuber flavors.

Currently, the breeding process to introduce the desired traits to the potato is labor-intensive and time-consuming, as most potato cultivars are tetraploids. This is a significant challenge, as potato feeds a rapidly-growing population, especially under the threat of global climate change (Stokstad, 2019). A possible pathway to circumvent this impediment of potato breeding could be the development of a hybrid potato breeding system that converts tetraploid potato into a diploid seed crop by the crossing of two inbred lines (Jansky et al., 2016; Lindhout et al., 2011). It is more efficient to introduce and fix beneficial alleles to diploid varieties, thereby shortening the breeding period (Jansky et al., 2016; Lindhout et al., 2011). Systematical study of CGA formation is very limited in diploid potatoes, and manipulation of tuber CGA levels to improve the quality traits of diploid potatoes is still far from being realized.

Here, we analyzed the influence of location, environment, genetic basis, as well as the expression of key enzymes, in affecting the abundance of the major isoform of CGAs (5-O-caffeoylquinic acid, CGA), using a natural variation diploid population including 40 wild species and 374 landraces. These results provide a basis to understand the mechanisms underlying the CGA formation in diploid potatoes, and also serve as guidance for the future breeding of high-quality potato by taking into consideration customer preferences.

## 2. Material and methods

### 2.1. Plant material

414 diploid potatoes, including *Solanum candolleianum* (40), *Solanum stenotomum* (113), *Solanum goniocalys* (38), and *Solanum phureja* (223) were planted in Xundian (102°41'-103°33'E, 25°20'-26°01'N, in March 2018) and Dehong (97°31'-98°43'E, 23°50'-25°20'N, in November 2018 and 2019) in the Yunnan Province. Each line was cultivated in fields containing ten replicates, and tubers from each line were simultaneously collected at end of the growing season. For each line, skin or flesh from five tubers were pooled together to generate a sample, while young leaves from five plants were also pooled together to generate a sample. The samples were immediately frozen in liquid nitrogen, then freeze-dried and stored at 4 °C until analysis. Three samples were used as three biological repeats in the following assays.

### 2.2. Sample preparation

200 mg powdered freeze-dried sample was homogenized with 2 mL of 75% ethanol (v/v) and incubated for 30 min in an ultrasonic bath, before being centrifuged at 13,945g for 2 min at 4 °C. The supernatant was then filtered (0.22 µm microporous organic filtration membrane) and transferred into a clean tube for analysis.

### 2.3. Quantification of chlorogenic acid, caffeic acid, ferulic acid, and antioxidant activity

Chlorogenic acid (5-O-caffeoylquinic acid), caffeic acid, and ferulic acid were determined by High-Performance Liquid Chromatography (HPLC) as described by Valinas et al. (2017) with minor modifications, which was carried out using an ACQUITY UPLC system (Waters, Milford, USA). A flow rate of 0.5 mL/min was used and 10 µL samples were injected onto a C-18 Phenomenex Luna column (250 × 3 mm i.d.; 5 µm particle size) at 25 °C. The mobile phases were (A) HPLC-grade water acidified with 4% H<sub>3</sub>PO<sub>4</sub> (Damao, Tianjin, China), and (B) acetonitrile (Damao, Tianjin, China). The solvent program was as follows: 0 to 25 mins, gradient from 10% to 25% B, then linear 30% B until 36.5 mins, return to 10% B between 36.5 and 37 mins and linear at 10% B until 40 mins. Simultaneous monitoring was set at 320 nm. 5-O-caffeoylquinic acid (Pusi, Chengdu, China), caffeic acid (Pusi, Chengdu, China), and ferulic acid (Pusi, Chengdu, China) were used as standards for the calibration curve. Chlorogenic acid, caffeic acid, and ferulic acid concentrations are expressed in µg/g DW. In addition to HPLC analysis, a UV-Vis spectrophotometer (Younike, Shanghai, China) was used to determine the content of chlorogenic acid at 327 nm for population samples. Antioxidant activity was measured using the 2,2-diphenyl-1-picrylhydrazyl (DPPH) assay (Valinas, Lanteri, Have, & Andreu, 2015) with minor modification. Briefly, 150 µL of extracts were added to 2850 µL DPPH (Yuanye, Shanghai, China) solution (100 µM) freshly prepared in methanol. The reaction was kept in the dark at room temperature for 24 h, and the absorbance at 515 nm was measured using a UV-Vis spectrophotometer (Younike, Shanghai, China). Trolox (Yuanye, Shanghai, China) was used as the standard for the calibration curve and the results are expressed as µg of Trolox equivalents g<sup>-1</sup> DW. For all determinations, three independent extractions from a pool of five tubers were conducted.

### 2.4. RNA extraction, cDNA synthesis, and qRT-PCR experiments

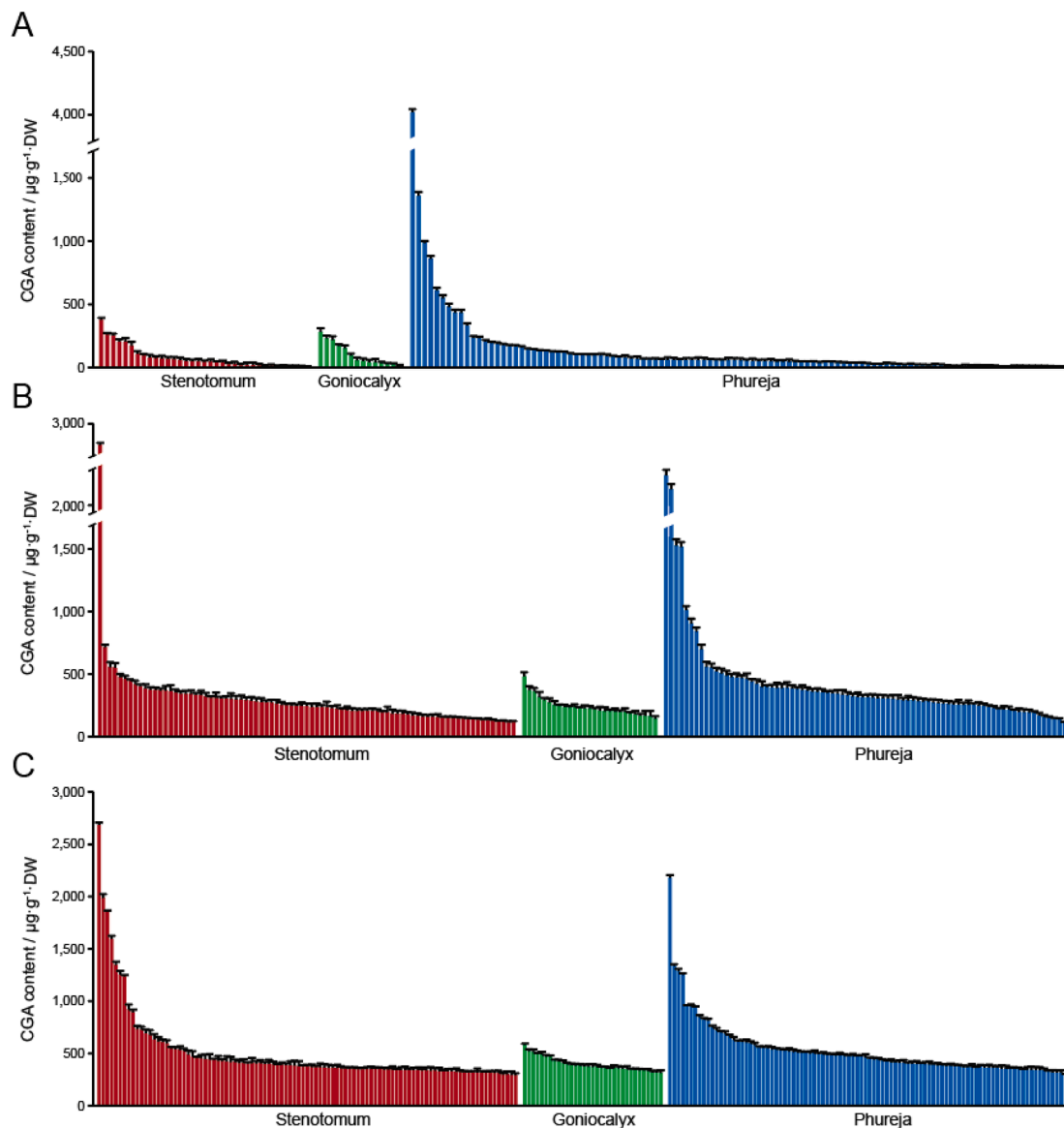
Total RNA from tubers was extracted using RNAPrep Pure polysaccharide polyphenol plant total RNA extraction kit (Tiangen, Beijing, China) following the manufacturer's instruction. First-strand cDNA synthesis was prepared using 1 µg total RNA with PrimerScript RT reagent kit (TaKaRa, Kusatsu, Japan). qRT-PCRs were performed using SYBR Premix (Takara, Kusatsu, Japan) on a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster, USA). Three independent biological experiments were performed in all cases. *ACTIN* was used as an internal control gene for normalization. The relative gene expression was calculated using 2<sup>-ΔΔCt</sup>. The primers used for qRT-PCRs are listed in [Supplementary Table 1](#).

### 2.5. Statistical analysis of phenotypic data

The variation assessments were compared before and after filter the data from the *Goniocalyx* group. The boxplot of the CGA content trait was plotted using Boxplot in R (<https://cran.r-project.org>). The best linear unbiased predictions (BLUPs) can eliminate the environmental deviation and estimate the real individual value. Therefore, the BLUPs were determined using the 'lme4' package of the R3.5.3 software ([www.r-project.org](http://www.r-project.org)), with year and location selected as random effects in the model  $[Y = \text{lmer}(X - (1|\text{Line}) + (1|\text{Year}) + (1|\text{Line}:\text{Year})$ . The descriptive statistics of the CGA content, correlations between BLUP and the measured values among different years, and those among BLUPs were analyzed using correlation in R3.5.3 and [corrrplot0.8.4](https://cran.r-project.org/web/packages/corrrplot) (<https://cran.r-project.org/web/packages/corrrplot>).

### 2.6. Genome-wide association studies (GWAS) for the CGA content of tuber flesh, skin, and leaf

PCA was conducted to evaluate the genetic structure of 271 diploid



**Fig. 1.** Comparison of CGA level of tuber flesh in different diploid potatoes. A-C, (A) Tubers of 158 species including *S. stenotomum* (109), *S. goniocalyx* (14), and *S. phureja* (35) were successfully collected from Xundian. (B) Tubers of 190 species including *S. stenotomum* (83), *S. goniocalyx* (27), and *S. phureja* (80) were harvested from Dehong in 2019, (C) while 226 lines including *S. stenotomum* (99), *S. goniocalyx* (33), and *S. phureja* (94) were harvested from Dehong in 2020. Varieties from each of the three Andean landrace groups are listed in a descending order based on its CGA level of tuber flesh. Data are presented as means  $\pm$  SD ( $n = 3$  biological replicates).

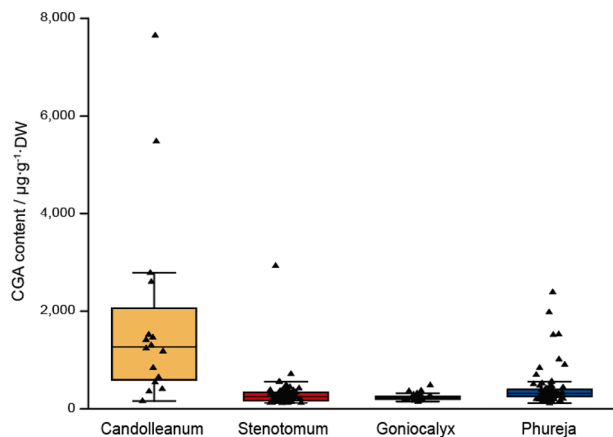
lines using Genome-wide Complex Trait Analysis (GCTA) software. The top five principal components were used to construct the covariate matrix, with which we performed the association analysis. GWAS were performed using the Emmax method (Valinas et al., 2015). Single nucleotide polymorphisms (SNPs) collected from the Resequencing project (Kang et al., 2010) were filtered with a missing rate  $\leq 10\%$ , and a Minor Allele Frequency (MAF)  $\geq 5\%$ . The BLUPs of tuber CGA content from plants grown in 2018 and 2019, and CGA content of the leaf and the tuber skin from plants grown in 2019 were utilized for the association study. GWAS threshold was set using a Bonferroni correction in the analysis of the tuber CGA content, and GWAS threshold was set as  $1 \times 10^{-6}$  in the analysis of the CGA content of leaf and tuber skin. Manhattan plots were generated using the qqman package in R (<https://github.com/stephenturner/qqman>). The effectiveness of the significant SNPs was calculated using gcta-1.93.2 (<http://cns.genomics.com/software/gcta>). To fully cover the candidate region, we extended the region to calculate the pairwise linkage disequilibrium (LD) using the SNPs

revealed by the GWAS analysis ( $p$  value  $< 0.001$  for chromosome 4 and 8, and  $p$  value  $< 0.0001$  for chromosome 10). Only the LD blocks containing at least one significant SNP were regarded as significant loci.

### 3. Results and discussion

#### 3.1. Selection of decreased CGA level of tuber flesh in the domestication of diploid potatoes

To systematically investigate the characteristics of CGA formation in diploid potatoes, a diversity panel consisting of a domestication progenitor *Solanum candolleianum* (40 lines) and the Andean landraces *S. tuberosum* groups Stenotomum (113 lines), Goniocalyx (38 lines), and Phureja (223 lines) were planted at Xundian ( $102^{\circ}41' - 103^{\circ}33'E$ ,  $25^{\circ}20' - 26^{\circ}01'N$ , in March 2018) and Dehong ( $97^{\circ}31' - 98^{\circ}43'E$ ,  $23^{\circ}50' - 25^{\circ}20'N$ , in November 2018 and 2019) in Yunnan Province (Supplementary Table 2). After the growth season, tubers of 158 species including

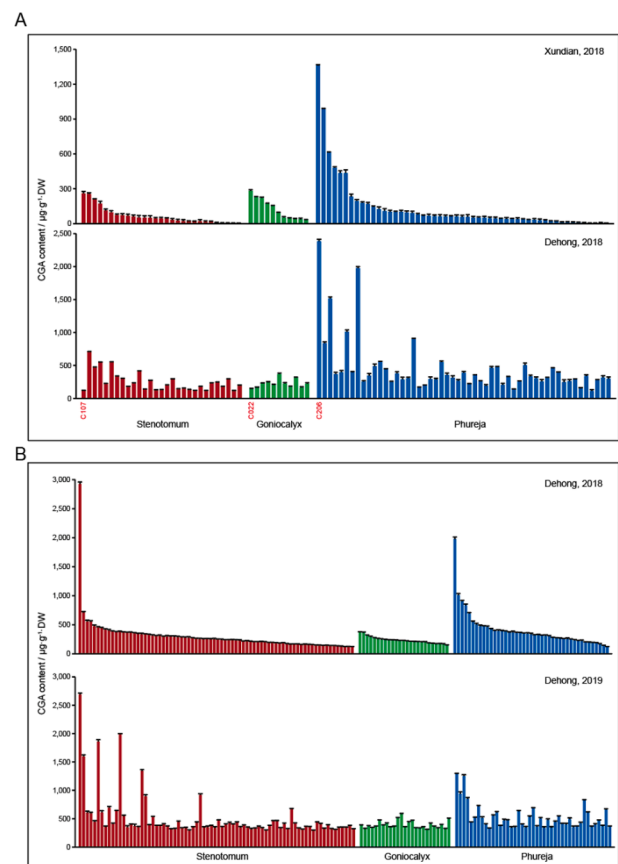


**Fig. 2.** Comparison of CGA level of tuber flesh between the ancestor and the landrace groups. CGA level of tuber flesh of 16 species from the immediate ancestor *S. candolleianum* group harvested at Kunming (2019), was compared to that of the 190 landrace varieties harvested at Dehong (2019). The ancestor group had a much higher CGA level of tuber flesh than the other three landrace groups.

*S. stenotomum* (109), *S. goniocalyx* (14), and *S. phureja* (35) were successfully gathered from Xundian. In contrast, tubers of 190 species including *S. stenotomum* (83), *S. goniocalyx* (27), and *S. phureja* (80) were harvested from Dehong in 2019, while 226 lines including *S. stenotomum* (99), *S. goniocalyx* (33), and *S. phureja* (94) were harvested from Dehong in 2020 because both field management and weather conditions were much better at Dehong, a well-known winter farming area in China, than that of Xundian. This study found that a large number of our chosen diploid species, especially those from the progenitor *S. candolleianum* group, failed to produce any tubers in each growing season, likely due to slow plant growth and pest infection. Thus, we had to use flowerpots and greenhouses to re-culture these *S. candolleianum* lines in Kunming (102°10'–103°40'E, 24°23'–26°22'N, July 2019), and finally, tubers from 16 progenitor species were successfully collected (Supplementary Fig. 1).

Tuber CGA level was determined using the Ultraviolet–visible spectroscopy method as well as HPLC, as both methods have been shown as efficient and accurate ways to determine CGA content in plants (Friedman, 1997b). The CGA concentrations in *S. phureja* tubers grown in the Xundian area, ranged from 0.60 to 4020.73 µg/g DW, while those in *S. goniocalyx* and *S. stenotomum* tubers ranged from 5.80 to 435.27 µg/g DW and 0.24 to 987.16 µg/g DW, respectively (Fig. 1A). A similar result was also observed for the tubers collected from both growing seasons at Dehong (Fig. 1B,C). More intriguingly, the average CGA concentrations of tuber flesh (1848.11 µg/g DW) are much higher in the immediate ancestor *S. candolleianum* group, when compared to that of the three Andean landraces (Fig. 2), although it should be noted that these ancestral lines were not grown in the same field as the other diploid lines.

A clear decrease in the abundance of CGAs present in Andean landrace tubers suggests a tuber CGA level was selected during the domestication of potato. This is consistent with previous studies that have found greater concentrations of phenolic compounds in wild species of potato and eggplant (Meyer et al., 2015; Nzaramba, Bamberg, Scheuring, & Miller, 2006). While other studies have also found comparable or even higher phenolic level in potato cultivars than in wild lines (Navarre, Pillai, Shakya, & Holden, 2011; Sołtys-Kalina, Murawska, Strzelczyk-Żyta, Wasilewicz-Flis, & Marczewski, 2019), this discrepancy may be due to a limited number of wild and cultivated lines used in this research. Given that domestication selects for certain metabolic processes, including the shikimate pathway which generates a tryptophan sink, and tryptophan catabolism which produces kynurenic acid (Hardigan et al., 2017), the metabolic direction of tryptophan to kynurenic

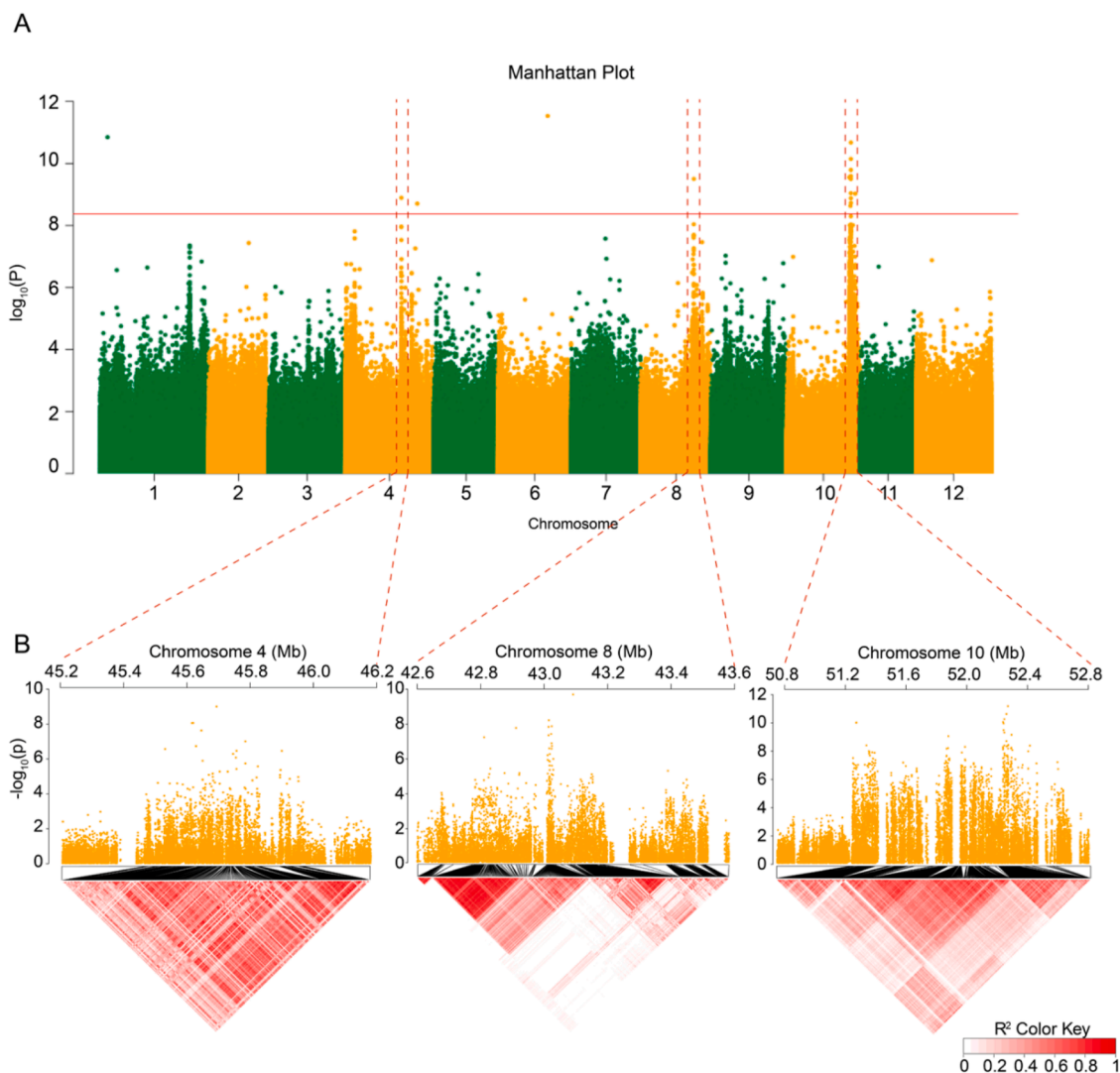


**Fig. 3.** Potential factors in affecting the CGA level of tuber flesh in diploid potatoes. (A) Location was an important influencing factor for the CGA concentration of tuber flesh in diploid lines. Tubers of 93 diploid lines inducing *S. stenotomum* (29), *S. goniocalyx* (11), and *S. phureja* (53) were successfully gathered at Xundian and Dehong after the growing season in 2018. Varieties from each of the three Andean landrace groups are listed in a descending order based on its CGA level of tuber flesh gathered from Xundian. For the lines grown at Xundian, the distribution patterns of tuber CGA contents are different from those of the plants grown at Dehong, especially for the 11 species from *S. goniocalyx* group. Three lines with featured CGA level of tuber flesh in the two growing seasons were indicated in red. (B) Planting time had less effect in shaping the CGA level of tuber flesh in diploid lines, when compared to location. Tubers of 144 diploid potatoes inducing *S. stenotomum* (76), *S. goniocalyx* (25), and *S. phureja* (43) were collected after two growth seasons at Dehong (2018 and 2019). Varieties from each of the three Andean landrace groups are listed in a descending order based on its CGA level of tuber flesh gathered from the 2018 growing season. Data are presented as means  $\pm$  SD ( $n = 3$  biological replicates). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

acid may decrease the abundance of tryptophan and phenylalanine in tubers (Yao & Brisson, 1995). This would limit the metabolic flux forward phenolic compounds such as CGAs and lignin.

Another possible explanation for the decreased CGA content is customer preference. Similar to the result in tetraploid potatoes (Orsák, Hamouz, Lachman, & Kasal, 2019; Ru et al., 2019; Valinas et al., 2017), higher CGA levels of tuber flesh and peel were observed in the colored diploid lines than that of ordinary potatoes (Supplementary Table 3-5). Our results (Supplementary Fig. 2) in accordance with previous studies (Albishi, John, Al-Khalifa, & Shahidi, 2013; Głosek-Sobieraj, Cwalina-Ambroziak, Waśkiewicz, Hamouz, & Perczak, 2019; Orsák et al., 2019; Valinas et al., 2017), found that purple and red-pigmented tubers have a greater concentration of total phenolics, anthocyanins, and antioxidant activities. Generally, US consumers prefer white-flesh varieties of potato rather than colored flesh (Navarre et al., 2011). This





**Fig. 4.** The genetic basis of CGA level of tuber flesh in diploid potatoes. (A) Manhattan plot of GWAS analysis for the CGA content of tuber flesh using the mixed linear model. The red line represents the significant threshold (adjusted by the Bonferroni method). 18 loci exceeding the threshold were identified in this analysis. The Quantile-Quantile plot of the p value shown in **Supplementary Fig. 5**. The regions containing significant polymorphisms were further restricted into three blocks (indicated by the dotted red line) by LD analysis. (B) Zoom in of the Manhattan plots and the local LD heatmap in chromosomal region Chr4:45.2–46.2 Mb (pvalue < 0.001), Chr8:42.6–43 (pvalue < 0.001), Mb, and Chr10:50.8–52.8 Mb (pvalue < 0.0001). Pairwise LD between the SNPs was estimated with  $r^2$  statistics. The pairwise LD  $r^2$  value from 0 to 1 is colored with white and dark red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

consumer preference likely encourages potato breeders to cultivate varieties with a low tuber CGA content to avoid pigmented tubers.

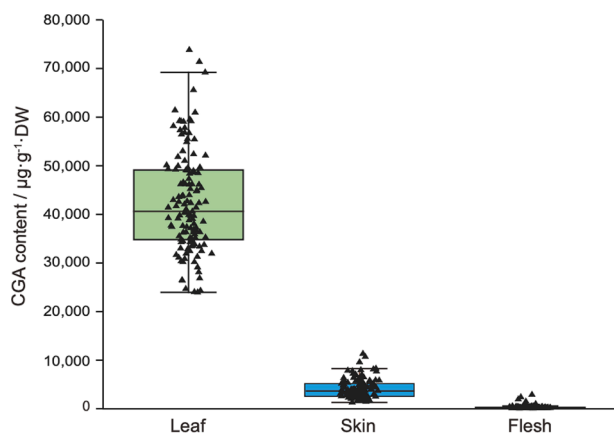
### 3.2. Major factors in affecting the CGA level of tuber flesh in diploid potatoes

Consistent with a previous study using tetraploid potatoes (Reddivari, Hale, & Miller, 2007), the most significant determinant of CGA levels in diploid potatoes was the effect of genotype (Fig. 2). A 2.80, 9.19, 1.87, and 7.23 fold change of the CGA level of tuber flesh were observed for the ancestor group *S. candolleianum*, the landrace groups *S. stenotomum*, *S. goniocalyx*, and *S. phureja*, respectively. This diversity in tuber CGA concentrations could only be explained by the significant phylogenetic distance between these diploid lines.

Following genetic variation, the second major determinant of the CGA content of tuber flesh is the geographic location. For all lines grown at Xundian, the distribution patterns of tuber CGA contents were different from those of the plants grown at Dehong, especially for the 11 species of *S. goniocalyx* (Fig. 3A). The most contrasting examples were

C107 and C22, whose tuber flesh had the highest CGA level in *S. stenotomum* and *S. goniocalyx* group, respectively, when they were grown at Xundian. However, once these two lines were planted at Dehong, their tuber flesh accumulated the lowest CGA in their respective group (Fig. 3A). This result further confirmed the role of location in determining the CGA level in potatoes (Reddivari et al., 2007). However, we also documented lines from *S. stenotomum* and *S. phureja* group, whose CGA levels were relatively stable and were only weakly influenced by the location. For instance, tube flesh of C206 from *S. phureja* accumulated the highest CGA content at both locations (Fig. 3A), indicating the genotype of this line is the decisive factor in controlling its CGA level.

Moreover, planting time was the third most influential factor in influencing the CGA content of tuber flesh in diploid potato. CGA levels were relatively stable when 144 diploid potatoes were collected after two growing seasons at Dehong, although inconsistent results were also observed for a few of the lines (Fig. 3B). Science a year is the least influential factor in controlling CGA content in diploid potatoes, the CGA contents of tuber flesh collected at between 2018 and 2019 were



**Fig. 5.** Comparison of the CGA level of different tissues in diploid potatoes. For 128 diploid lines planted at Dehong (2019), the CGA contents of leaf, tuber skin, and flesh were determined and compared. The CGA levels of leaves and tuber skins were much higher than that of tuber flesh in diploid potatoes.

stable, and thus used in the following GWAS analysis to further reveal the potential genetic basis underlying the CGA formation in diploid potatoes.

### 3.3. Potential genetic basis of the CGA formation in diploid potato

To identify the key quantitative trait locus (QTL) associated with CGA content in diploid potatoes, it was important to select populations that were not strongly genetically structured and yet exhibited clear phenotypic diversity. Since CGA levels of tubers fluctuated in the Goniocalyx group when they were planted in different places (Fig. 3A), we just focused on the 271 diploid varieties including two Andean landrace groups: *Stenotomum* (109) and *Phureja* (162). The phenotypic diversity of CGA concentrations in *Stenotomum* and *Phureja* groups was almost comparable to that of the global collection (35 Goniocalyx lines included) (Supplementary Fig. 3A). In addition, the PCA based on the 12,163,843 SNPs (unpublished data) revealed that the population structure of these 271 varieties exhibited a continuous distribution, without any distinct clusters (Supplementary Fig. 3B). Furthermore, as the extent of the linkage disequilibrium (LD) is another important factor in determining the efficiency of GWAS, we also examined the LD status of the three populations. The decay of LD with the physical distance between SNPs occurred at 22 kb, with  $r^2$  reaching half of its maximum  $r^2$  value (Supplementary Fig. 3C), which was comparable to that of a previous study that examined a more genetically diverse population (Li et al., 2018). Thus, the populations we selected were likely suitable for GWAS analysis.

To locate loci associated with the CGA content in diploid potatoes, the best linear unbiased prediction (BLUP) value was obtained from the CGA content of each accession grown at Dehong (2018 and 2019). The BLUP can eliminate environmental variation and estimate the real individual value for breeding, thus it has gradually become a common application used by breeders who want to generate precise estimates of genotypic values (Piepho, Möhring, Melchinger, & Büchse, 2008). The correlation rates between BLUP and the measured values obtained in 2018 and 2019 were 0.70 and 0.90, respectively (Supplementary Fig. 4), indicating that the BLUP values were significantly correlated with the CGA levels of individuals. We identified 18 loci exceeding the significant threshold using a Bonferroni correction (Fig. 4A). To measure the contribution of genotype to the CGA content trait, the genetic architecture of the trait was inspected further. The heritability of the SNPs within the candidate region was 0.21 (95% confidence region: 0.10 ~ 0.32) (Yang, Lee, Goddard, & Visscher, 2011), suggesting that the variation of CGA content could be partly explained by these 18 SNPs.

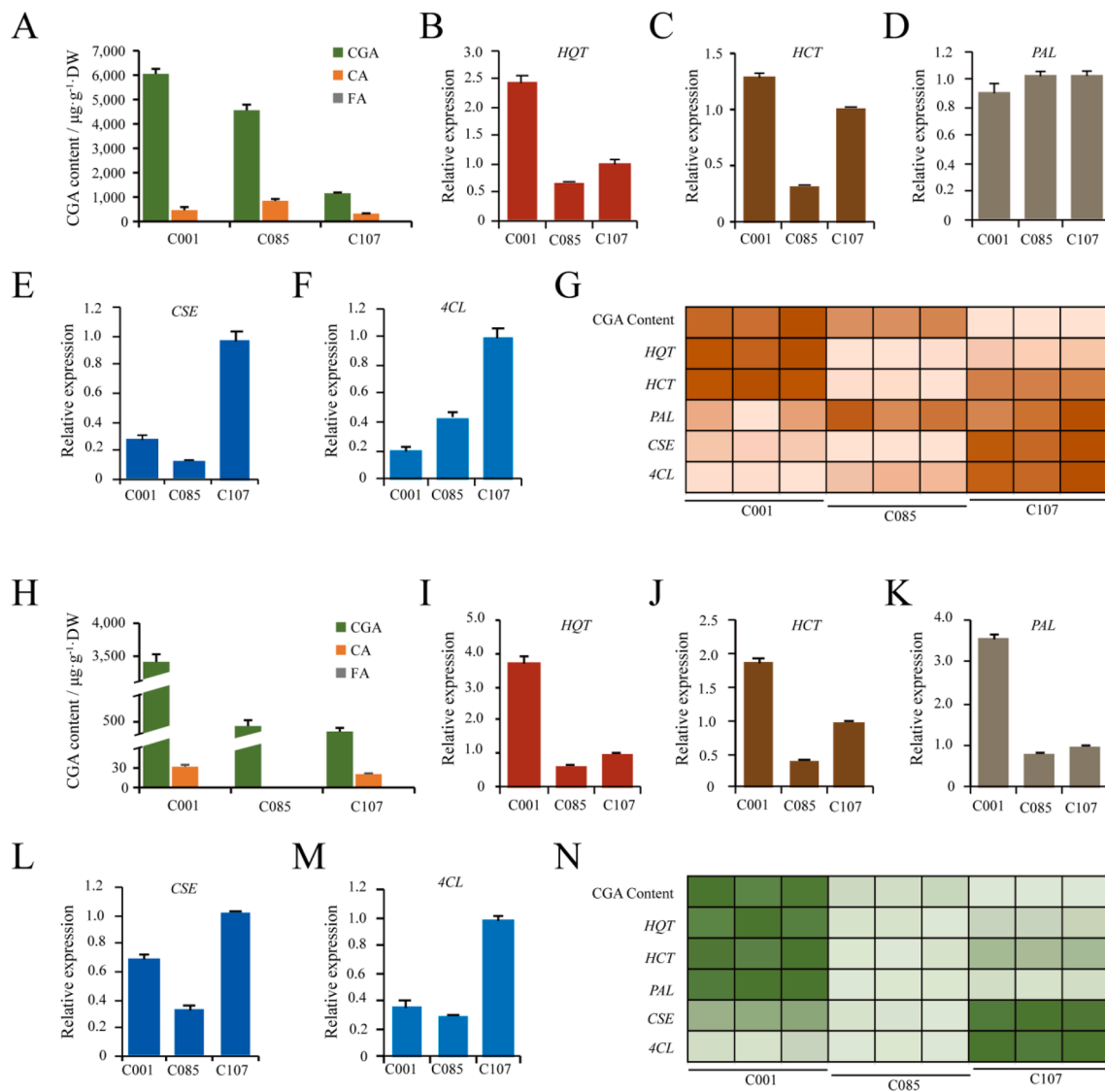
The regions containing significant polymorphisms were further restricted into three blocks on Chr. 4, 8, and 10 by LD analysis (Shin, Blay, Mcnenny, & Graham, 2006), which harbored 68 candidate genes (Fig. 4B; Supplementary Table 6).

Further, we also determined the CGA contents of tuber skin and leaf using the diploid lines planted at Dehong (2019, Fig. 5A). The result was consistent with previous studies that examined the CGA levels of leaves and tuber skins, documenting a higher concentration than that in the tuber flesh in tetraploid potatoes (Deusser, Guignard, Hoffmann, & Evers, 2012; Valinas et al., 2015, 2017). Moreover, since CGA could be transported from tuber flesh to skin (Valinas et al., 2015), we further tested whether CGA levels of leaf, tuber peel, and flesh were co-related in diploid potato. The low Pearson correlation values indicated that CGA biosynthesis in different tissues were independently regulated in diploid potatoes, except for flesh and peel (Supplementary Fig. 6). Using GWAS analysis, we additionally identified a total of 28 significant loci ( $p$  value  $< 1 \times 10^{-6}$ ) mainly distributed in chromosome 5 for the CGA abundance in leaves, and chromosome 10 for the CGA abundance in peels (Supplementary Fig. 7). Intriguingly, the polymorphisms associated with the CGA content of tuber flesh and peel overlapped on chromosome 10 (Supplementary Fig. 8), indicating the CGA biosynthesis may be co-regulated in these two tissues.

### 3.4. Expression patterns of the key enzymes involved in the CGA metabolism

In addition to utilizing GWAS analysis to investigate the potential genetic basis in controlling the CGA levels in diploid potatoes, three lines with a low, medium, or high CGA content of tuber peel and flesh were selected from *S. stenotomum* group to study the expression of the CGA biosynthetic enzymes. We focused on the enzymes, such as phenylalanine ammonia-lyase (PAL), 4coumarate-CoA ligase (4CL), caffeoyl shikimate esterase (CSE), hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase (HQT), and hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT), that are involved in driving the metabolic flux toward CGA or byproducts, such as caffeic acid (CA) and ferulic acid (FA). Among these enzymes, PAL and 4CL are responsible for the biosynthesis of *p*-Coumaroyl-CoA, a common precursor for CGA production and lignin biosynthesis (Howles et al., 1996; Niggeweg et al., 2004; Valinas et al., 2015), while CSE is an enzyme competing with HCT for the common substrate, Caffeoyl shikimic acid, to produce CA during lignin pathway (Valinas et al., 2015). In contrast to PAL, 4CL, and CSE, the two enzymes, HQT, and HCT are supposed to catalyze the committed steps for CGA biosynthesis, although their roles were not consistent in different studies (Valinas et al., 2017; Niggeweg et al., 2004; Payyavula et al., 2015; Finally, we also analyzed levels of CA and FA in tuber peel and flesh Valinas et al., 2015). Interestingly, the CGA content of tuber peel was co-related with the expression levels of *HQT* and *HCT*, but not with *PAL* (Fig. 6). While the CGA content of tuber flesh was co-related with the expression level of all the three enzymes. In addition, expression of *CSE* and *4CL* were negatively co-related with the CGA contents of tuber peel and flesh (Fig. 6). Since CSE is potentially involved in producing the by-product CA, this result is consistent with our expectations. However, for 4CL, this result indicates this upstream enzyme may also participate in the downstream reaction to direct metabolic flux toward by-products.

FA was not detected in either tuber peel or flesh, while the CA content of the tuber flesh was very low, even in the line that has the highest CGA level (Fig. 6). Although this result is inconsistent with previous findings that CA and FA are co-related with the CGA level in tuber peel and flesh (Valinas et al., 2015), a similar result was also observed in purple potato flesh (Albishi et al., 2013; Ru et al., 2019; Valinas et al., 2017), and color potato could be explained by the fact that both CA and FA were in bound fraction, not in a free status, as in the colored potato (Ru et al., 2019).



**Fig. 6. Relationships between three phenolic acids and five biosynthetic enzymes in diploid potatoes.** A and H, The levels of CGA, CA, and FA of tuber peel (A) or flesh (H) in three diploid lines from the *S. stenotomum* group. FA level of tuber flesh or peel is extremely low in all three lines. Data are presented as means  $\pm$  SD ( $n = 3$  biological replicates). (B-F) and (H-M), The expression levels of *HQT*, *HCT*, *PAL*, *CSE*, and *4CL* in tuber flesh (B-F) or peel (H-M) of three varieties. The expression level of the candidate gene in C107 was set as 1, and was used to normalize against the results from other samples. Data are presented as means  $\pm$  SD ( $n = 3$  biological replicates). G and N, Pearson correlation coefficients were calculated using the value of CGA level and expression of the biosynthetic enzyme in tuber peel (G) or flesh (N) of the three varieties ( $n = 3$  biological replicates). The data was presented using a heatmap. The dark color indicates a high correlation rate. The CGA content of tuber peel was co-related with the expression levels of *HQT* and *HCT*, not with *PAL*, while the CGA content of tuber flesh was co-related with the expression level of all three enzymes. In addition, expressions of *CSE* and *4CL* were negatively co-related with the CGA contents of tuber peel and flesh.

#### 4. Conclusion

CGA levels of tuber flesh in diploid potatoes are associated with flesh color and antioxidant activity and are mainly influenced by genetic variation of the plants and the environment. During the domestication of diploid potatoes, a decreased tuber CGA level was potentially selected for, likely due to the customer preference of white flesh potatoes and the selected metabolic flux toward side products, such as kynurenic acid. We also provide the SNPs and candidate genes associated with the CGA level of tuber flesh using re-sequencing genome data of diploid lines. This study provides a basis for the future breeding of high-quality potatoes by manipulation of the tuber CGA level.

#### 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.

#### Acknowledgments

We thank Prof. Jianjian Qi (Inner Mongolia University) and Qun Lian (Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences) for experimental and bioinformatics support. This work was supported by the National Natural Science Foundation of China (31972433), Yunnan Science Fund (202005AE160015 and 2019FJ004 to Y.S.), and YNNU Innovation Fund Project (yjs2018151 to H.Q.Y.). This work was also supported by Shenzhen Municipal Governments (AGIS-ZDKY202001).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fochms.2021.100039>.

## References

- Albishi, T., John, J. A., Al-Khalifa, A. S., & Shahidi, F. (2013). Phenolic content and antioxidant activities of selected potato varieties and their processing by-products. *Journal of Functional Foods*, 5(2), 590–600. <https://doi.org/10.1016/j.jff.2012.11.019>
- Deusser, H., Guignard, C., Hoffmann, L., & Evers, D. (2012). Polyphenol and glycoalkaloid contents in potato cultivars grown in Luxembourg. *Food Chemistry*, 135(4), 2814–2824. <https://doi.org/10.1016/j.foodchem.2012.07.028>
- Devaux, A., Goffart, J. P., Petsakos, A., Kromann, P., & Hareau, G. (2020). Global food security, contributions from sustainable potato agri-food systems. *The Potato Crop*. [https://doi.org/10.1007/978-3-030-28683-5\\_1](https://doi.org/10.1007/978-3-030-28683-5_1)
- Friedman, M. (1997a). Chemistry, biochemistry, and dietary role of potato polyphenols. *Journal of Agricultural and Food Chemistry*, 45(5), 1523–1540. <https://doi.org/10.1021/jf960900s>
- Friedman, M. (1997b). Potato polyphenols: Role in the plant and in the diet. *ACS Symposium Series*, 61–93. <https://doi.org/10.1021/bk-1997-0662.ch005>
- Glosek-Sobieraj, M., Cwalina-Ambroziak, B., Waśkiewicz, A., Hamouz, K., & Perczak, A. (2019). The effect of biostimulants on the health status and content of chlorogenic acids in potato tubers (*Solanum Tuberosum* L.) with colored flesh. *Gesunde Pflanzen*, 71(1), 45–60. <https://doi.org/10.1007/s10343-018-00441-7>
- Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., ... Buell, C. R. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences of the United States of America*, 114(46), E9999–E10008. <https://doi.org/10.1073/pnas.1714380114>
- Harrison, H. F., Mitchell, T. R., Peterson, J. K., Wechter, W. P., Majetic, G. F., & Snook, M. E. (2008). Contents of caffeoylquinic acid compounds in the storage roots of sixteen sweetpotato genotypes and their potential biological activity. *Journal of the American Society for Horticultural Science*, 133(4), 492–500. <https://doi.org/10.21273/JASHS.133.4.492>
- Howles, P. A., Sewalt, V.J.H., Paiva, N. L., Elkind, Y., Bate, N. J., Lamb, C., & Dixon, R. A. (1996). Overexpression of L-phenylalanine ammonia-lyase in transgenic tobacco plants reveals control points for flux into phenylpropanoid biosynthesis. *Plant Physiology*, 112(4), 1617–1624. <https://doi.org/10.1104/pp.112.4.1617>
- Im, H. W., Suh, B. S., Lee, S. U., Kozukue, N., Ohnisi-Kameyama, M., Levin, C. E., & Friedman, M. (2008). Analysis of phenolic compounds by high-performance liquid chromatography and liquid chromatography/mass spectrometry in potato plant flowers, leaves, stems, and tubers and in home-processed potatoes. *Journal of Agricultural & Food Chemistry*, 56(9), 3341–3349. <https://doi.org/10.1021/jf073476b>
- Jansky, S. H. (2010). Potato Flavor. *American Journal of Potato Research*, 87(2), 209–217. <https://doi.org/10.1007/s12230-010-9127-6>
- Jansky, S. H., Charkowski, A. O., Douches, D. S., Gusmini, G., Richael, C., Bethke, P. C., ... Jiang, J. (2016). Reinventing potato as a diploid inbred line-based crop. *Crop Science*, 56(4), 1412–1422. <https://doi.org/10.2135/cropsci2015.12.0740>
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., ... Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354. <https://doi.org/10.1038/ng.548>
- Li, Y., Colleoni, C., Zhang, J., Liang, Q., Hu, Y., Ruess, H., ... Huang, B. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. *Molecular Plant*, 11(3), 473–484. <https://doi.org/10.1016/j.molp.2018.01.009>
- Lindhout, P., Meijer, D., Schotte, T., Hutten, R. C. B., Visser, R. G. F., & Eck, H. J. V. (2011). Towards F<sub>1</sub> hybrid seed potato breeding. *Potato Research*, 54(4), 301–312. <https://doi.org/10.1007/s11540-011-9196-z>
- Maldonado, A. F. S., Mudge, E., Gänzle, M. G., & Schieber, A. (2014). Extraction and fractionation of phenolic acids and glycoalkaloids from potato peels using acidified water/ethanol-based solvents. *Food Research International*, 65, 27–34. <https://doi.org/10.1016/j.foodres.2014.06.018>
- Meyer, R. S., Whitaker, B. D., Little, D. P., Wu, S. B., Kennelly, E. J., Long, C. L., & Litt, A. (2015). Parallel reductions in phenolic constituents resulting from the domestication of eggplant. *Phytochemistry*, 115, 194–206. <https://doi.org/10.1016/j.phytochem.2015.02.006>
- Navarre, D. A., Pillai, S. S., Shakya, R., & Holden, M. J. (2011). HPLC profiling of phenolics in diverse potato genotypes. *Food Chemistry*, 127(1), 34–41. <https://doi.org/10.1016/j.foodchem.2010.12.080>
- Niggeweg, R., Michael, A. J., & Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology*, 22(6), 746–754. <https://doi.org/10.1038/nbt966>
- Nzaramba, N. M., Bamberg, J. B., Scheuring, D. C., & Miller, J. C. (2006). Antioxidant activity in solanum species as influenced by seed type and growing location. *American Journal of Potato Research*, 83, 127. <https://doi.org/10.1007/BF02869613>
- Onakpoya, I. J., Spencer, E. A., Thompson, M. J., & Heneghan, C. J. (2015). The effect of chlorogenic acid on blood pressure: A systematic review and meta-analysis of randomized clinical trials. *Journal of Human Hypertension*, 29(2), 77–81. <https://doi.org/10.1038/jhh.2014.46>
- Orsák, M., Hamouz, K., Lachman, J., & Kasal, P. (2019). Chlorogenic acid content in potato tubers with colored flesh as affected by a genotype, location and long-term storage. *Plant, Soil and Environment*, 65(No. 7), 355–360. <https://doi.org/10.17221/195/2019-PSE>
- Payyavula, R. S., Shakya, R., Sengoda, V. G., Munyaneza, J. E., Swamy, P., & Navarre, D. A. (2015). Synthesis and regulation of chlorogenic acid in potato: Rerouting phenylpropanoid flux in HQT-silenced lines. *Plant Biotechnology Journal*, 13(4), 551–564. <https://doi.org/10.1111/pbi.12280>
- Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1-2), 209–228. <https://doi.org/10.1007/s10681-007-9449-8>
- Reddivari, L., Hale, A. L., & Miller, J. C. (2007). Genotype, location, and year influence antioxidant activity, carotenoid content, phenolic content, and composition in specialty potatoes. *Journal of Agricultural and Food Chemistry*, 55(20), 8073–8079. <https://doi.org/10.1021/jf071543w>
- Riciputi, Y., Diaz-De-Cerio, E., Akyol, H., Capanoglu, E., Cerretani, L., Caboni, M. F., & Verardo, V. (2018). Establishment of ultrasound-assisted extraction of phenolic compounds from industrial potato by-products using response surface methodology. *Food Chemistry*, 269, 258–263. <https://doi.org/10.1016/j.foodchem.2018.06.154>
- Ru, W., Pang, Y., Gan, Y., Liu, Q., & Bao, J. (2019). Phenolic compounds and antioxidant activities of potato cultivars with white, yellow, red and purple flesh. *Antioxidants*, 8(10), 419. <https://doi.org/10.3390/antiox8100419>
- Sands, D. C., Morris, C. E., Dratz, E. A., & Pilgeram, A. L. (2009). Elevating optimal human nutrition to a central goal of plant breeding and production of plant-based foods. *Plant Science*, 177(5), 377–389. <https://doi.org/10.1016/j.plantsci.2009.07.011>
- Shin, J. H., Blay, S., Mcnenny, B., & Graham, J. (2006). LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software*, 16(1), 1–9. <https://doi.org/10.18637/jss.v016.c03>
- Sottys-Kalina, D., Murawska, Z., Strzelczyk-Żyta, D., Wasilewicz-Flis, I., & Marczewski, W. (2019). Phytotoxic potential of cultivated and wild potato species (*Solanum* sp.): Role of glycoalkaloids, phenolics and flavonoids in phytotoxicity against mustard (*Sinapis alba* L.). *Acta Physiologiae Plantarum*, 41(5), 1–9. <https://doi.org/10.1007/s11738-019-2848-3>
- Song, W., Derito, C. M., Liu, M. K., He, X., Dong, M., & Liu, R. (2010). Cellular antioxidant activity of common vegetables. *Journal of Agricultural and Food Chemistry*, 58(11), 6621–6629. <https://doi.org/10.1021/jf9035832>
- Stokstad, E. (2019). The new potato. *Science*, 363(6427), 574–577. <https://doi.org/10.1126/science.363.6427.574>
- Tajik, N., Tajik, M., Mack, I., & Enck, P. (2017). The potential effects of chlorogenic acid, the main phenolic components in coffee, on health: A comprehensive review of the literature. *European Journal of Nutrition*, 56(7), 2215–2244. <https://doi.org/10.1007/s00394-017-1379-1>
- Tiemann, D., Zhu, G., Resende, M. F. R., Lin, T., Nguyen, C., Bies, D., ... Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science*, 355(6323), 391–394. <https://doi.org/10.1126/science.aal1556>
- Valinas, M. A., Lanteri, M. L., Have, A. T., & Andreu, A. B. (2015). Chlorogenic acid biosynthesis appears linked with suberin production in potato tuber (*Solanum tuberosum*). *Journal of Agricultural and Food Chemistry*, 63(19), 4902–4913. <https://doi.org/10.1021/jf505777p>
- Valinas, M. A., Lanteri, M. L., Have, A. T., & Andreu, A. B. (2017). Chlorogenic acid, anthocyanin and flavan-3-ol biosynthesis in flesh and skin of Andean potato tubers (*Solanum tuberosum* subsp. andigena). *Food Chemistry*, 229, 837–846. <https://doi.org/10.1016/j.foodchem.2017.02.150>
- Wojciechowska, E., Weinert, C. H., Egert, B., Trierweiler, B., Schmidt-Heydt, M., Horneburg, B., ... Geisen, R. (2014). Chlorogenic acid, a metabolite identified by untargeted metabolome analysis in resistant tomatoes, inhibits the colonization by *Alternaria alternata* by inhibiting alternariol biosynthesis. *European Journal of Plant Pathology*, 139(4), 735–747. <https://doi.org/10.1007/s10658-014-0428-3>
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yao, K., & Brisson, L. N. (1995). Creation of a metabolic sink for tryptophan alters the phenylpropanoid pathway and the susceptibility of potato to phytophthora infestans. *Plant Cell*, 7(11), 1787–1799. <https://doi.org/10.1105/tpc.7.11.1787>