

RESEARCH ARTICLE

Pathway-based analysis of anthocyanin diversity in diploid potato

Maria Angelica Parra-Galindo^{1†}, Johana Carolina Soto-Sedano², Teresa Mosquera-Vásquez¹, Federico Roda^{3*}

1 Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Sede Bogotá, Bogotá, Colombia, **2** Departamento de Biología, Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Bogotá, Colombia, **3** Max Planck Tandem Group, Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Bogotá, Colombia

† Current address: Fitomejoramiento, Medcolcanna S.A.S, Bogotá, Colombia

* frdad@unal.edu.co

OPEN ACCESS

Citation: Parra-Galindo MA, Soto-Sedano JC, Mosquera-Vásquez T, Roda F (2021) Pathway-based analysis of anthocyanin diversity in diploid potato. PLoS ONE 16(4): e0250861. <https://doi.org/10.1371/journal.pone.0250861>

Editor: Xiu-Qing Li, Agriculture and Agri-Food Canada, CANADA

Received: December 18, 2020

Accepted: April 14, 2021

Published: April 29, 2021

Copyright: © 2021 Parra-Galindo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information files](#).

Funding: This study received support from the following sources: Colciencias (now Minciencias, <https://minciencias.gov.co/>) grant number 110171250437, awarded to TMV; the International Development Research Centre (IDRC) and Global Affairs Canada through the Canadian International Food Security Research Fund (CIFSRF), which funded the project San Nariño number 108125-002, received by TMV; and Convenio 566 of 2014

Abstract

Anthocyanin biosynthesis is one of the most studied pathways in plants due to the important ecological role played by these compounds and the potential health benefits of anthocyanin consumption. Given the interest in identifying new genetic factors underlying anthocyanin content we studied a diverse collection of diploid potatoes by combining a genome-wide association study and pathway-based analyses. By using an expanded SNP dataset, we identified candidate genes that had not been associated with anthocyanin variation in potatoes, namely a Myb transcription factor, a Leucoanthocyanidin dioxygenase gene and a vacuolar membrane protein. Importantly, a genomic region in chromosome 10 harbored the SNPs with strongest associations with anthocyanin content in GWAS. Some of these SNPs were associated with multiple anthocyanin compounds and therefore could underline the existence of pleiotropic genes or anthocyanin biosynthetic clusters. We identified multiple anthocyanin homologs in this genomic region, including four transcription factors and five enzymes that could be governing anthocyanin variation. For instance, a SNP linked to the phenylalanine ammonia-lyase gene, encoding the first enzyme in the phenylpropanoid biosynthetic pathway, was associated with all of the five anthocyanins measured. Finally, we combined a pathway analysis and GWAS of other agronomic traits to identify pathways related to anthocyanin biosynthesis in potatoes. We found that methionine metabolism and the production of sugars and hydroxycinnamic acids are genetically correlated to anthocyanin biosynthesis. The results contribute to the understanding of anthocyanins regulation in potatoes and can be used in future breeding programs focused on nutraceutical food.

Introduction

Potato (*Solanum tuberosum* L.) is the main non-cereal food consumed worldwide [1] and the vegetable with the highest antioxidant contribution to human diet [2]. Within the *S. tuberosum* L. species, the Group Phureja is composed of diploid potatoes ($2n = 2x = 24$) with short-day

between Universidad Nacional de Colombia (<https://unal.edu.co/>) and Colciencias (now Minciencias, <https://minciencias.gov.co/>), received by FRF.

Competing interests: MAPG is a paid employee of Medcolcanna S.A.S, but began their employment after concluding the research included in this manuscript, so Medcolcanna S.A.S. did not provide any support for this study. This affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

adaptation and a lack of tuber dormancy that are widely grown by local farmers in the Andes mountains range of South America. Landraces from the Andes were the first domesticated potatoes and the main origin of cultivars, developed after the colonization of America and grown in most of the rest of the world today [3]. There is growing interest in recovering genetic variation for agronomic traits, one of these traits is the presence of bioactive compounds that are present in landraces and was lost during the improvement of cultivars [4, 5]. One of the bioactive compounds with increasing interest is reflected in the red and purple coloration in the skin and flesh of potato tubers [6–8], which result from the accumulation of anthocyanin pigments [9, 10]. Multiple potential health benefits have been described to the consumption of anthocyanin-pigmented potatoes, including the protection against several diseases, mainly because of their antioxidant capacity [11–13]. Phureja potatoes present a particularly broad variation in anthocyanin contents [14], with total anthocyanin values ranging from zero to 23 mg / 100 g fresh weight and from zero to 167.76 mg / 100 g dry weight [14, 15]. In fact, pigmentation is one of the main traits selected during the breeding of native Phureja landraces, producing an amazing diversity of coloration patterns, mostly associated with anthocyanin accumulation [16, 17].

Anthocyanins are synthesized in the cytosol through the phenylpropanoid pathway (Fig 1), which begins with the catalysis of the amino-acid phenylalanine by the enzyme phenylalanine-ammonia lyase (PAL). Then the chalcone synthase (CHS) catalyzes the condensation of three acetate units from malonyl-COA with p-coumaroyl-COA to yield tetrahydrochalcone. Chalcone isomerase (CHI) then catalyzes the tetrahydrochalcone to naringenin. Naringenin is hydrolyzed to dihydroflavonols by three enzymes, namely flavanone-3-hydroxylase (F3H), flavonoid-3'-hydroxylase (F3'H) and flavonoid-3',5'-hydroxylase (F3'5'H). The dihydroflavonols are reduced to three different leucoanthocyanidins by dihydroflavonol-4-reductase (DFR), and their glycosylation by leucoanthocyanidin dioxygenase/anthocyanidin synthase (LDOX/ANS) produces the basic structures of anthocyanins (anthocyanidins—aglycons) that determine the coloration in plant tissues [18–21]. Genes encoding anthocyanin biosynthetic enzymes are known as “structural genes” and they are conserved among different species [22]. However, tissue-specific expression of different structural genes are controlled by transcription factors (TF) known as “regulatory genes” [22]. While structural genes determine the ability to produce a set of compounds regulatory genes generally affect the intensity and pattern of anthocyanin biosynthesis, particularly through the MYB-bHLH-WD (MBW) complex [23–25].

In potatoes the synthesis of anthocyanins was initially described as being controlled by the *R* (red), *P* (purple) and *D/I* (developer or inhibitor) loci, which map to chromosomes 2, 11, and 10, respectively [27–30]. *R* and *P* loci govern red and violet pigmentation in tubers while *D/I* is responsible for the intensity of pigmentation [27–30]. Later it was found that *R* and *P* loci code for two enzymes of flavonoid biosynthesis, DFR [31] and F3'5'H [32], which are responsible for the creation of red and purple anthocyanin pigments respectively. The locus *I* corresponded to a regulatory gene encoding the TF R2R3 MYB, which has a high similarity with the product of the *Petunia hybrida* AN2 on chromosome 10 [33]. This gene governs the expression of multiple enzymes in the pathway, therefore affecting the level of multiple anthocyanin pigments [33]. Recently, a number of regulatory genes potentially controlling anthocyanin biosynthetic structural genes have been identified in potato tubers [20, 22, 34, 35] including three R2R3-MYB encoding genes (*StAN1*, *StMYBA1* and *StMYB113*) [20, 21, 36] two bHLH genes (*StJAF13* and *StbHLH1*) [21, 37] and one WD40 (*StWD40*) [38].

The pattern of anthocyanin composition across tissues and genotypes is controlled by multiple genes [18, 19]. Genome-wide association studies (GWAS) have been used to elucidate the complex genetic mechanisms that define anthocyanin content in potato tubers [15]. This

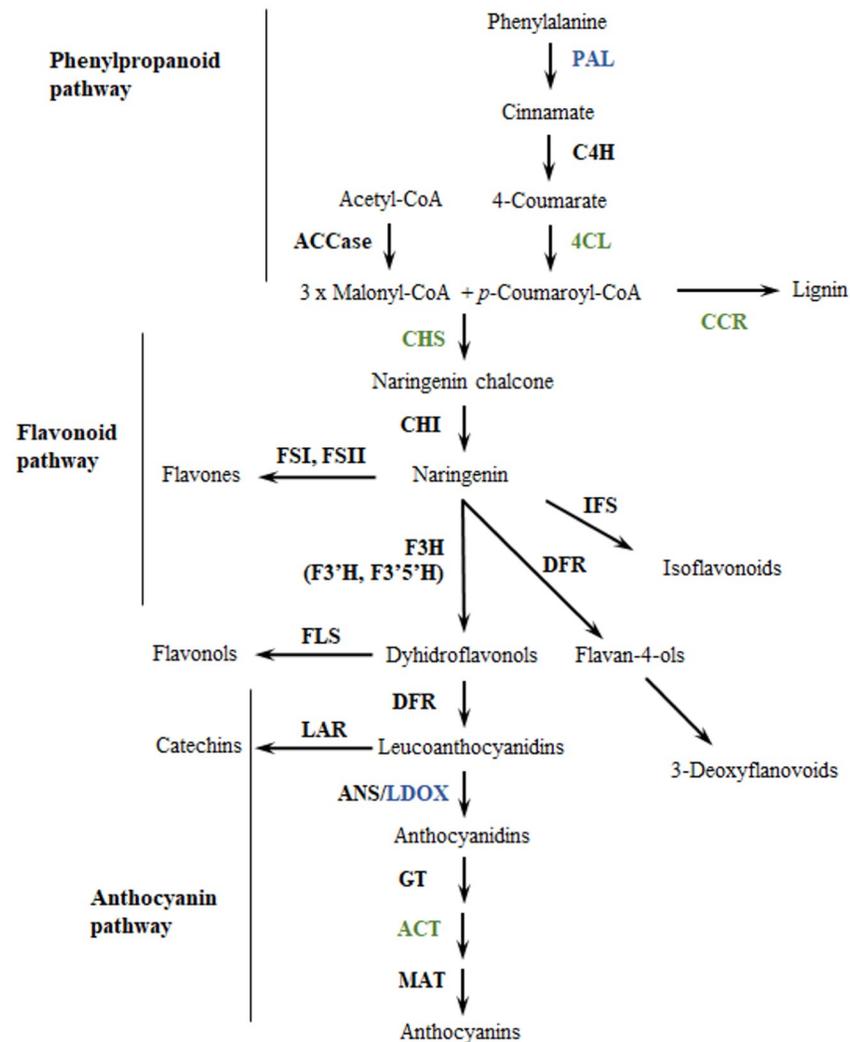


Fig 1. Enzymes from the phenylpropanoid pathway associated to anthocyanin variation in a population of *Solanum tuberosum* group Phureja. Enzymes in blue correspond to genes identified in GWAS and gene-set analysis. Enzymes in green correspond to genes identified in the gene-set analysis. Enzyme abbreviations: ACCase: acetyl-CoA carboxylase; PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; CCR: cinnamoyl CoA reductase; CHS: chalcone synthase; CHI: chalcone isomerase; FS: flavone synthase; IFS: isoflavone synthase; F3H (F3'H, F3'5'H): flavanone 3 β -hydroxylase; F3'H: flavonoid 3'-hydroxylase; F3'5'H: flavonoid 3',5'-hydroxylase; FLS: flavonol synthase; DFR: dihydroflavonol 4-reductase; LAR: leucoanthocyanidin reductase; ANS: anthocyanidin synthase; GT: glucosyltransferase; ACT: anthocyanin acyltransferase; MAT: malonyltransferase. Modified from Springob [26].

<https://doi.org/10.1371/journal.pone.0250861.g001>

methodology allows identifying quantitative trait locus (QTL) for a given trait and determining aspects of its genetic architecture, like the number of QTL and their respective contribution to the phenotype [27, 39, 40]. Pathway analysis can help exploiting the results of GWAS by using prior information on biological pathways and combining the genetic effects of many genes [41–43]. Different methods have been implemented to perform pathway-based analysis using data from GWAS [44]. Some studies use gene set enrichment analysis (GSEA) to examine whether a set of genes significantly associated with a trait of interest is enriched in specific pathways [45, 46]. Other analyses re-calculate associations in a predefined set of genes belonging to a biochemical route of interest [47]. A third approach analyses all genetic sequences

associated with a trait of interest, regardless of significance or magnitude, and uses the gene effect values to calculate an enrichment score for each pathway [48–50]. Finally, one can combine analyses of multiple molecular or morphological traits evaluated in the same population to determine how these traits interact genetically [51].

Previous genetic studies of anthocyanin pigmentation in potato tubers conducted using biparental populations of potato identified a small number of QTL which explained from 8% to 11% of the phenotypic variation [27, 29, 51, 52]. Therefore, we still ignore the genetic factors that contribute to this missing heritability as well as the genomic distribution and biochemical identity of anthocyanin determinants. We also know little about the evolution of anthocyanin pigmentation during the domestication of potatoes. Has the same trait evolved repeatedly under different genetic control or does it have a unique origin across cultivated potatoes? The answers to these questions are crucial to design breeding strategies to improve anthocyanin content. Diploid landraces represent an underexploited source of genetic diversity [53–55] and an excellent model to fill these knowledge gaps. Recently, an exploratory analysis of anthocyanin content in these landraces identified QTL explaining more than 30% of the phenotypic variation [15]. Therefore, the primary objective of this study is to use pathway analysis to exploit previous studies conducted in the same population of diploid potatoes [15] to identify new genes, genomic regions and biochemical routes important for anthocyanin production. We were thus able to discover new genetic factors that drove for the recurrent evolution of anthocyanin pigmentation during the domestication of potato landraces. This information provides functional links to bridge the knowledge gap between the genetic variants and the phenotypes.

Materials and methods

Genome-wide association analysis

The Working Collection of Potato Breeding Program of *Solanum tuberosum* Group Phureja from the Universidad Nacional de Colombia (CCC) was employed for the GWAS using information partially published in previous studies [15]. Briefly, potato tubers from an association panel consisting of 96 accessions was phenotyped through Ultra High-Performance Liquid Chromatographic (UHPLC) analysis for the detection of five different anthocyanidins compounds (cyanidin, peonidin, pelargonidin, delphinidin, and petunidin) [15].

In order to do functional analyses, we expanded the SNP matrix used to run the GWAS in previous studies [15]. The original Single Nucleotide Polymorphism (SNP) matrix obtained through genotyping by sequencing [56] was filtered by removing SNPs with a minor allele frequency (MAF) higher or equal to 0.05 and less than 10% of missing data. We thus obtained 47,298 SNP markers.

The GWAS was conducted for each anthocyanin compound using a compression mixed linear model (CMLM) [57] applied by the Genome Association and Prediction Integrated Tool (GAPIT) R package[®] [58]. The principal components in the CMLM were used in order to control for population structure [59]. The Benjamini & Hochberg corrected threshold probability based on individual tests was calculated to control false-discovery rates (FDRs) [60] using a threshold of 0.1 given the sample size of the association panel ($n = 96$). The linkage disequilibrium (LD) between pairs of SNP markers was calculated through squared allele-frequency correlations (R^2) by using TASSEL software [61].

Gene-set analysis

We used *prior* biological information about the biosynthesis of anthocyanins to pre-select a subset of candidate genes. We searched in the literature for structural and regulatory genes

involved in anthocyanin production in the Solanaceae family. We made use of a study that identified flavonoid orthologs in multiple Solanaceae species [62]. We also searched for genes associated with the term “anthocyanin” in the NCBI (<https://www.ncbi.nlm.nih.gov/>), KEGG (<https://www.genome.jp/kegg/>), Spud (<http://solanaceae.plantbiology.msu.edu/>), and BioCyc (<https://biocyc.org/>) genomic databases. In order to identify homologs of these anthocyanin genes in the potato genome we performed a BLASTx (v2.6.0) [63] of the sequences of from other plants against the potato reference genome DM—v4.03 [27, 64] and retrieved the best hits with a cutoff of 10^{-20} . We then retrieved SNPs located ± 100 Kb of these genes and ran again the GWAS with this subset of SNPs. We used a relatively large window of 100 Kb in order to get multiple SNPs associated with the genes. The p -values within each gene were recalculated with the new subset of SNP markers and inputted to corrected for multiple testing using the approach reported by Benjamini and Hochberg [60], based on procedure to control FDR at 0.1. We assigned the lowest p -value value among all SNPs mapped to a gene as the p -value of the gene [65, 66]. The goal of this analysis was to identify anthocyanin homologs that show the strongest association with anthocyanin variation. For this reason, we evaluate only SNPs linked to anthocyanin homologs using standard GWAS methods. Therefore, significant SNPs are those that pass the significance and FDR thresholds using this reduced dataset.

Pathway analysis

We used the PAST software [50] to conduct the pathway analysis using genomic annotation from two databases, PotatoCyc 4.0 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Stuberosum) and KEGG (<https://www.genome.jp/kegg/pathway>).

SNPs were assigned to genes based on LD information and a distance of 1,500 bp from the tagSNP [50, 67]. Statistical significance of a pathway was determined by taking 1,000 permutations of the gene effect values to generate a null distribution for the Enrichment Score (NES) [50]. Pathways with p -value < 0.05 were selected based on thresholds set for gene association and effect values of the genes.

Phylogenetic and population genetics analyses

We used the TASSEL software to conduct a phylogenetic tree of the populations using the Neighbor Joining algorithm and two inputs: (1) All SNPs used in GWAS and (2) significant SNPs from Chromosome 10. We also used Tassel to conduct principal components analysis (PCA) using the covariance option and the same sets of SNPs.

TASSEL was used to calculate the Tajima D statistic using the sliding window option (step = 10, window = 10). SNPs falling in the upper and lower 1% percentiles of the distribution were considered candidates for balancing and positive selection respectively (S11 Table).

Genomic architecture of variation for other traits

The potato population used in our study has also been rated for other agricultural and nutritional important traits like macronutrients [68], sugars [54], hydroxycinnamic acids (HCAs) [69], and resistance to late blight caused by the pathogen *Phytophthora infestans* [56]. We were interested in evaluating whether some of these traits show phenotypic or genomic correlations with anthocyanin variation. We ran a PCA using phenotypic data and calculated pairwise correlations between all traits using TASSEL. We used TASSEL to conduct GWAS for all traits using a mixed linear model (MLM), correcting for population structure with a PCA (covariance method, 5 components) and a Centered IBS method of Kinship estimation. Each variance component was estimated once (P3D).

Results and discussion

Genome-wide association in an extended SNP panel identifies new candidate genes

In this study we exploited previous research on the Work Collection of Potato Breeding Program from Colombia [15] to understand the genomics and evolution of anthocyanin variation. The population analyzed here is relatively small but genetically diverse [53, 70], thus representing a valuable resource to identify and manipulate traits that have been lost during the breeding of cultivars outside the original range of potatoes [4, 5]. We re-analyzed data from a GWAS of five anthocyanin compounds, namely cyanidin, peonidin, pelargonidin, delphinidin, and petunidin [15], using an expanded dataset of 47,298 SNP markers (S1 Table). LD decay is fast in this population (S3 Table) which makes it a good system to track causal genes. Therefore, by expanding the genotyping panel we were able to search for genes underlying the QTL detected previously and find new important variants. In total 22 SNPs were significantly associated with at least one compound at a genome-wide FDR of 0.1 (Table 1, S1 Fig and S2 Table). Sixteen of these significant SNPs were located in the coding region of annotated genes on the Chromosomes 1, 2, 6, 9, 10 and 11 (Table 1).

Our results confirmed the significant SNPs reported by Parra-Galindo [15] and identified new associations. The strongest association signals were detected again in two nearby defensive genes from Chromosome 10, namely within the *Chloroplast threonine deaminase 1* (PGSC0003DMG400017604) and the *STS14* (PGSC0003DMG400017597) genes. Chloroplast threonine deaminases are the first enzymes in the biosynthesis of the amino acid isoleucine but also mediate plant defenses against pathogens and herbivores [71]. STS14 belongs to a family of pathogenesis-related secretory proteins [72]. Although anthocyanin compounds are induced by biotic stress these two enzymes have not been associated to anthocyanin production previously. It is thus possible that causal mutations underlying these associations are positioned in other genes located in the vicinity, as we will explore in the next sections.

We highlight three newly detected associations from our expanded SNP set involving polymorphisms located within putative anthocyanin biosynthetic genes (Table 1); a SNP on Chromosome 11 linked to the *Leucoanthocyanidin dioxygenase* gene (*LDOX*; PGSC0003DMG400008650); a position on Chromosome 1 located in a R2R3-MYB transcription factor (*Myb12*; PGSC0003DMG400009033); and a SNP on chromosome 6 linked to the *Vacuolar membrane protein* (*PEP3*; PGSC0003DMG400025399). Firstly, *LDOX* enzyme, also known as anthocyanidin synthase, catalyzes the conversion of leucoanthocyanidins to anthocyanidins, the precursors of anthocyanins [73]. Previous studies [74, 75] found that *LDOX* genes are located on QTL for anthocyanin variation in Chromosomes 8 and 9. However, there are no reports of anthocyanin QTL in the genomic region of Chromosome 11 containing the *LDOX* gene reported in our study. Secondly, R2-R3 MYB transcription factors showing homology with *Petunia AN2* gene (Borevitz2000) regulate anthocyanin biosynthesis in dicotyledonous plants [76–80]. A number of these AN2 homologs regulate the expression of anthocyanin enzymes in tubers and flowers of potato [33, 35, 37], including three nearby genes from chromosome 10 named *StAN1*, *StAN2* and *StFLAN2*. The Myb12 TF detected in this study has not been linked to anthocyanin production in potatoes but its orthologs regulate anthocyanin production in other plants like *Arabidopsis thaliana* [81], apple [82], lily [83, 84], and grape [85]. Finally, the *Vacuolar membrane protein PEP3* is an interesting candidate because this gene is involved in vacuole organization [86], a process that is essential for anthocyanin biogenesis and accumulation.

By expanding the number of genetic markers evaluated in this potato population and analyzing the function of genes underneath the most associated markers from GWAS we were

Table 1. Summary of genome associations for anthocyanin content in a population of *Solanum tuberosum* group Phureja.

Trait	Chr	Position	R ² model	p-value	FDR_Adjusted	Effect	Gene_ID	Annotation gene
All anthocyanins	ch10	52004868	0,44	6.21E-09	2.9,E-04	0.0406	PGSC0003DMG400017604	Chloroplast threonine deaminase
Pelargonidin, peonidin, delphinidin	ch10	52261553	0,45	8.35E-08	0.002	-0.4556	PGSC0003DMG400017597	STS14 protein
Pelargonidin, delphinidin, peonidin	ch10	52261573	0,46	4.54E-08	0.002	-0.0506	PGSC0003DMG400017597	STS14 protein
Delphinidin, peonidin, pelargonidin	ch10	52004940	0,40	1.42E-07	0.0029	-0.364	PGSC0003DMG400017604	Chloroplast threonine deaminase
Peonidin, delphinidin, pelargonidin	ch10	54746624	0,35	2.07E-06	0.0196	-0.1967	PGSC0003DMG400011047	60S ribosomal protein L4/L1
Petunidin	ch06	8156658	0,19	2.77E-06	0.0262	0.4535	PGSC0003DMG400025399	Vacuolar membrane protein PEP3
Petunidin	ch06	8156696	0,19	2.77E-06	0.0262	-0.1985		
Petunidin	ch01	72364545	0,18	3.34E-06	0.0263	-0.2882	PGSC0003DMG402000051	HMG-I and HMG-Y
Pelargonidin	ch02	41058521	0,37	5.06E-06	0.0342	0.0631	PGSC0003DMG400012655	Nadph-cytochrome P450 oxydoreductase
Pelargonidin	ch02	41058534	0,37	5.06E-06	0.0342	-0.4182	PGSC0003DMG400012655	DUF292 domain containing protein
Pelargonidin	ch02	41058575	0,37	5.06E-06	0.0342	0.141	PGSC0003DMG400012655	DUF292 domain containing protein
Petunidin	ch01	24135132	0,19	7.37E-06	0.0498	0.0975		
Petunidin	ch01	24135105	0,19	9.94E-05	0.0588	0.0052		
Peonidin, delphinidin	ch10	54778485	0,33	7.67E-06	0.0605	0.1559	PGSC0003DMG400010985	Subtilase
Petunidin	ch11	3129502	0,18	1.29E-05	0.0679	-0.0546		
Pelargonidin	ch10	51318001	0,35	1.40E-05	0.0829	0.0855	PGSC0003DMG400019155	F-box family protein
Petunidin	ch01	53094258	0,19	2.28E-05	0.0982	0.1819	PGSC0003DMG400023276	60S ribosomal protein L27
Petunidin	ch01	53094259	0,19	2.28E-05	0.0982	0.3271	PGSC0003DMG400023276	60S ribosomal protein L27
Cyanidin	ch01	62416352	0,19	2.55E-05	0.0996	0.068	PGSC0003DMG400009033	Myb 12 transcription factor
Petunidin	ch11	33265607	0,19	1.31E-04	0.0996	-0.3339	PGSC0003DMG400008650	Leucoanthocyanidin dioxygenase LDOX
Cyanidin	ch09	59749919	0,19	2.54E-05	0.1008	-0.0838	PGSC0003DMG400020593	Acyl-CoA synthetase
Pelargonidin	ch10	54207412	0,19	3.29E-05	0.1011	0.0527	PGSC0003DMG400011082	PRA1 family protein

<https://doi.org/10.1371/journal.pone.0250861.t001>

able to identify potential determinants of anthocyanin variation in potatoes that were not detected in previous studies. Given that causal genes might be in the vicinity of significant SNPs we searched for anthocyanin homologs in broader QTL regions.

Analysis of anthocyanin homologs provides a deeper understanding of pathway regulation

The anthocyanin biosynthetic pathway is one of the most extensively studied pathways of plant specialized metabolism. Several genes have been reported to determine anthocyanin levels in potato cultivars and accessions [20, 21]. However, we know little about the causes of anthocyanin variation in potato landraces, which have a broader and largely untapped genetic diversity [5]. We made use of the extensive information available in the literature and genomic databases on anthocyanin genes in other plants to identify gene targets contributing to anthocyanin variation in this genetically diverse panel of potato landraces. In GWAS, it is a challenge to identify variants with moderate to weak effect sizes because the effect of many variants can be compounded by interactions with other loci [39]. In order to identify genes contributing to the missing heritability in anthocyanin accumulation we re-evaluated trait associations using only SNPs genetically linked to a set of anthocyanin homologs. Specifically, we tested for an

Table 2.

RefGen_DM -V4.03	Annotation gene	Compound	Chr	Position	p-value	FDR_Adjusted p-value
PGSC0003DMG400031365	Phenylalanine ammonia-lyase	All anthocyanins	Ch10	52,004,868	$8,14 \times 10^{-9}$	0,000555
PGSC0003DMG400019825	Cinnamoyl-coA reductase	Pelargonidin	Ch03	56,864,063	$2,17 \times 10^{-6}$	0,002551
PGSC0003DMG400034671	1-O-acylglucose:anthocyanin-O-acyltransferase	Cyanidin	Ch04	68,747,095	$2,17, \times 10^{-6}$	0,002551
PGSC0003DMG400010987	Wrky transcription factor	All anthocyanins	Ch10	54,746,624	$2,09 \times 10^{-6}$	0,002599
PGSC0003DMG400024129	Leucoanthocyanidin dioxygenase	Petunidin	Ch01	15,307,188	$4,22 \times 10^{-5}$	0,013962
PGSC0003DMG400003155	4-coumarate:CoA ligase	Petunidin	Ch03	47,051,715	$2,15, \times 10^{-4}$	0,028549
PGSC0003DMG400029620	Chalcone synthase	Petunidin	Ch09	58,297,746	$2,15 \times 10^{-4}$	0,043449
PGSC0003DMG400025373	Cinnamoyl-coA reductase	Cyanidin	Ch11	44,394,705	$9,24 \times 10^{-4}$	0,058597
PGSC0003DMG400006814	AN1-like transcription factor	Cyanidin	Ch01	66,008,979	$1,48 \times 10^{-3}$	0,098137
PGSC0003DMG401012339	BHLH transcription factor JAF13	Cyanidin	Ch08	54,831,215	$1,65 \times 10^{-3}$	0,099547

<https://doi.org/10.1371/journal.pone.0250861.t002>

association between the content of five anthocyanins in a subset of pre-selected candidate anthocyanin homologs retrieved from the literature and databases (S4 Table). The analysis involved recalculating trait associations among the five anthocyanins and the genotypes at SNP markers located within ± 100 Kb of the 108 *anthocyanin homologs* (S5 Table). Nineteen SNPs were statistically significant at 0.1 FDR for at least one of five anthocyanin compounds (S6 Table). These 19 significant SNP markers were located into 10 of the pre-selected candidate genes (Table 2). Nine of these SNPs were not detected in the initial genome-wide association study.

Seven structural genes showed significant associations in this analysis (Table 2). In Fig 1 we show their positions within the Flavonoid pathway. These were: *phenylalanine ammonia-lyase* (PAL); leucoanthocyanidin dioxygenase (LDOX); cinnamoyl-coA reductases (CCR1); 4-coumarate: CoA ligase (4CL1); 1-O-acylglucose:anthocyanin-O-acyltransferase enzyme (CtSCPLAT1); and chalcone synthase enzyme (CHS) [18]. CCR is considered a control point in regulating the overall carbon flux toward lignin [87] and downregulation of CCR activates the enzymes PAL, C4H and 4CL [88]. The 4-coumarate-CoA ligase is a key enzyme in the phenylpropanoid pathway participating in monolignol biosynthesis [89] while the CHS catalyzes the first committed step in the biosynthesis of anthocyanin pigments [18]. CHS enzymes located within anthocyanin QTL have been identified in tomato (TCHS1, TCHS2) [90] and potato [74]. Finally, the 1-O-acylglucose:anthocyanin-O-acyltransferase enzyme catalyzes the acylation reactions of anthocyanins one of the final steps in anthocyanin biosynthesis [91].

A *phenylalanine ammonia-lyase* homolog (PGSC0003DMG400031365) in chromosome 10 is particularly interesting because it is linked to the most significant SNPs that display associations to the five anthocyanins measured. The association of a PAL homolog to all anthocyanins makes sense biochemically because the PAL enzyme catalyzes the first reaction in the phenylpropanoid biosynthetic pathway [92], which leads to the biosynthesis of all anthocyanins (Fig 1). The role of PAL homologs in the regulation of anthocyanin production has been described in many plants [93–95]. For instance, in *A. thaliana*, mutations in the two isoforms of PAL gene cause a reduced production of all anthocyanins [96]. Within the potato genome there are 11 PAL genes but only the homolog identified in our study has been previously associated with variation in anthocyanin content. For instance, Liu and colleagues [97] found that changes in the expression of this gene are involved in anthocyanin biosynthesis under heat stress. Additionally, the upregulation of this gene is associated with a greater accumulation of anthocyanins in potato flowers [35].

Importantly, eight significant SNP markers were located near the PAL gene, in a region of 4 Mb in the extreme of Chromosome 10 (Fig 3). These SNPs were linked to anthocyanin

regulatory genes, namely *StFLAN2* (PGSC0003DMG400019217) and *WRKY 13* (PGSC0003DMG400010987) (Table 2). It was recently discovered that *StFLAN2* is the main regulator of floral anthocyanin production in potato [35], matching the function of its ortholog in petunia (*PhAN2*). On the other hand, *WRKY 13* is orthologous with *TRANSPARENT TESTA GLABRA 2* from Arabidopsis and *PhPH3* from petunia, which control the transcription of structural genes responsible for anthocyanin biosynthesis as well as ion pumps that determine the pH of vacuoles, where anthocyanins are stored [98]. Interestingly, the transcription factors identified in our study are members of three families known to control anthocyanin production through the formation of the so-called MBW complex [25]: a MYB TF (*StFLAN2*), a BHLH TF (*JAF13*) and a WRKY TF (*WRKY 13*). Furthermore, the orthologs of these genes interact to control anthocyanin production and storage in other plant species [36]. These results show that, by integrating previous information on pathways we were able to recover a more complete picture of the gene interactions that determine anthocyanin variation in potatoes.

Anthocyanin genes are clustered in Chromosome 10. Many of the QTL identified in studies of anthocyanin variation simultaneously govern variation in multiple anthocyanin compounds [33, 76, 99]. The co-localization of QTL often arises from the existence of pleiotropic genes governing the biosynthesis of multiple pigments [99] but can also result from genetic clustering of determinants of the different anthocyanins [100]. In this collection of potatoes, the levels of the five different anthocyanins are correlated across individuals [15]. Additionally, the most significant SNPs from GWAS govern variation in multiple anthocyanins and are located in a relatively small (4 Mb) region at the end of Chromosome 10 and (Tables 1 and 2, Fig 3). These results can be explained by the existence of pleiotropic genes and/or by the presence of anthocyanin biosynthetic clusters in Chromosome 10. In this context, we define a pleiotropic gene as a gene that governs simultaneously the production of multiple anthocyanins while a biosynthetic cluster is a physically clustered group of two or more genes that together determine the production of anthocyanins. We evaluated these two non-exclusive hypotheses by analyzing gene function, recombination and genetic variation in this 4 Mb genomic region.

We first looked at the distribution of anthocyanin homologs across chromosome 10 to see if these putative anthocyanin genes are clustered in the 4 MB region containing significant SNPs from GWAS. We found that this genomic region contains 10 out of the 28 putative anthocyanin genes located in Chromosome 10 and is among the regions of the genome with the highest density of anthocyanin homologs (Fig 3, S10 Table). These include the *PAL* gene, four putative 7-O-linked N-acetylglucosamine transferases, an oxidoreductase, a WRKY transcription factor, and at least three Myb transcription factors (S6 Table). Some of these genes are adjacent and seem to be the result of recent tandem duplications. These include *PAL* [97], 7-O-linked N-acetylglucosamine transferases and Myb transcription factors. In fact, previous studies have shown that this genomic region is very dynamic, with both transposon activity and copy number variation [35]. This genomic region contains the MYB TF *StFLAN2* regulating flower color as well as a close paralog also responsible for segregation of corolla anthocyanin production. More importantly, the genomic region also contains two additional R2R3 MYB TFs that determine anthocyanin production in potato tubers (*StAN1*) [33] and throughout the plant (*StAN2*) [36].

We calculated LD in the 4 Mb genomic region, as high LD can indicate the maintenance of a cluster of linked alleles that are inherited as a single haplotype [101]. LD decay is relatively fast in this genomic region (mean distance among markers with a $R^2 > 0.8$ for this genomic region = $11,935 \pm 5,007$, for the rest of the chromosome = $21,130 \pm 1,217$). This indicates that

recombination is not reduced at this site and that different significant SNPs from GWAs are located within different haplotypes (mean distance between SNPs = $7,422 \pm 1,965$) (S7 Table).

It is likely that the genetic basis of anthocyanin variation differs across individuals from our panel. To evaluate whether potatoes with high anthocyanin content shared alleles at the loci governing anthocyanin variation, we analyzed the genomic variation of significant SNPs from GWAS using phylogenetic and multivariate analyses. We found that the genotypes at significant SNPs from Chromosome 10 do not fully separate plants with high and low anthocyanin content (Fig 2, S2 Fig). This suggests that anthocyanin variation has multiple phylogenetic and genetic origins in the population. Finally, we searched for footprints of natural selection in this genomic region using Tajima's D statistic. Values of Tajima's D are high (i.e., upper 99 the percentile of genome wide distribution) at the edge of the putative anthocyanin cluster (Fig 3, S1 Table). This indicates that alleles are kept at intermediate frequencies, a genomic pattern that can result from balancing selection.

Pleiotropy and genetic clustering are common genomic patterns in specialized metabolism [102–105]. Both patterns can produce correlations between the concentrations of different metabolites. Pleiotropy usually involves enzymes acting upstream in the biosynthetic pathway or regulatory genes governing the expression of key enzymes [102, 106, 107]. We found evidence of pleiotropy since the most significant SNPs from GWAS are associated to the content of all anthocyanins. Additionally, these SNPs are linked to putatively pleiotropic genes. For instance, the *PAL* gene catalyzes the first step in the phenylpropanoid pathway and therefore could be a pleiotropic gene whose expression or/and sequence affects the production of anthocyanins located downstream in the biosynthetic pathway. Intriguingly, *PAL* is located nearby three pleiotropic transcription factors that govern the production of multiple anthocyanins *StAN1*, *StAN2* *StfLAN2*. Another candidate from Chromosome 10 that could pleiotropically control multiple anthocyanins is the *WRKY* 13 TF [25]. It should be mentioned that given the fast LD decay of the population it is not likely that the GWAS signal comes from any of these genes exclusively.

The clustering of genes from the same route has been reported in many specialized metabolism pathways and it is proposed as a mechanism to synchronize gene expression or to maintain favorable allelic combinations in the face of recombination [103, 104]. Despite the tandem duplication of many enzymes and TFs involved in anthocyanin metabolism some studies suggest that this pathway is particularly reticent to clustering [108, 109]. Surprisingly, we found evidence of clustering of anthocyanin determinants, as there is a concentration of structural and regulatory genes in the region of Chromosome 10 containing significant SNPs from GWAS. Many of these genes are located at very short distances from each other and present polymorphic tandem duplications and deletions in the potato lineage [64, 110], including the *PAL* gene [97] as well as the 7-O-GTs [62] and MYB transcription factors [37, 64, 110]. Finally, the genes from this putative cluster show coordinated expression patterns [97], which suggest that they are under a common genetic control. This genomic region is remarkably dynamic in potatoes as well as in other Solanaceae like petunia [111] and tomato [112], perhaps due to high transposon density. This suggests that natural selection as well as domestication have shaped this region concentrating anthocyanin determinants.

Phylogenetic analyses of the genomic region containing this group of anthocyanin genes suggest that multiple alleles or allelic combinations were involved in the creation of potato landraces with high anthocyanin content. Interestingly, a study of historical European samples shows a drastic reduction of genetic diversity and negative values of Tajima's D in this genomic region [113]. The authors associate this pattern to selection in gibberellin genes during the adaptation of potatoes to European temperate weathers after their introduction from South America [113]. This is in contrast with Elevated Tajima's D found in our study, which

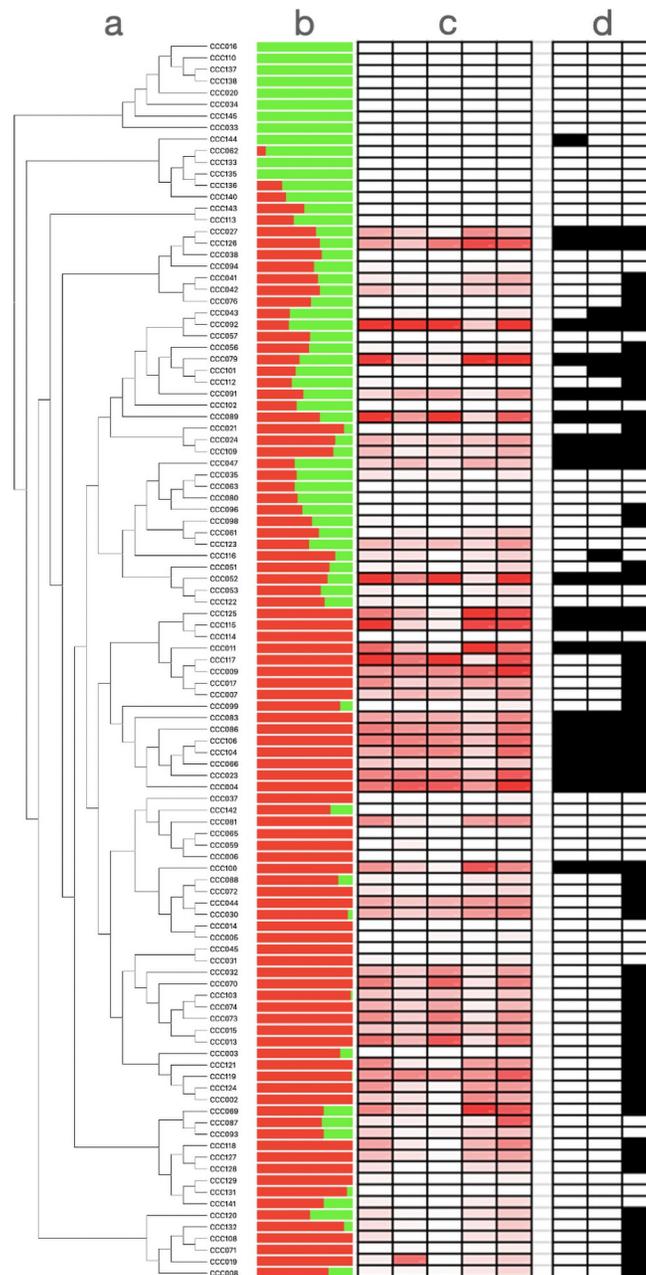


Fig 2. Genetic and metabolic variation shown in *Solanum tuberosum* group Phureja. (a) Neighbor joining phylogeny conducted with all SNPs used in GWAS. Names on the top of the branches correspond to the names of potato accessions from the Colombian Core Collection (CCC). (b) Results of Admixture with a $K = 2$. (c) Anthocyanin content: Darker colorations are associated with darker concentrations of the compound. 1: Delphinidin; 2: Cyanidin; 3: Petunidin; 4: Pelargonidin; 5: Peonidin. (d) Genotypes at significant SNPs from Chromosome 10. The allele associated to a darker coloration (in homozygous or heterozygous state) is in black. 1: pos 52,004,868; 2: pos 52,261,553; 3: pos 54,746,624.

<https://doi.org/10.1371/journal.pone.0250861.g002>

indicates that South American varieties maintained high genetic diversity in this genomic region. We postulate that this genetic diversity could have resulted from selection for diverse patterns of tuber coloration during the domestication and improvement of potato landraces in the Andes.

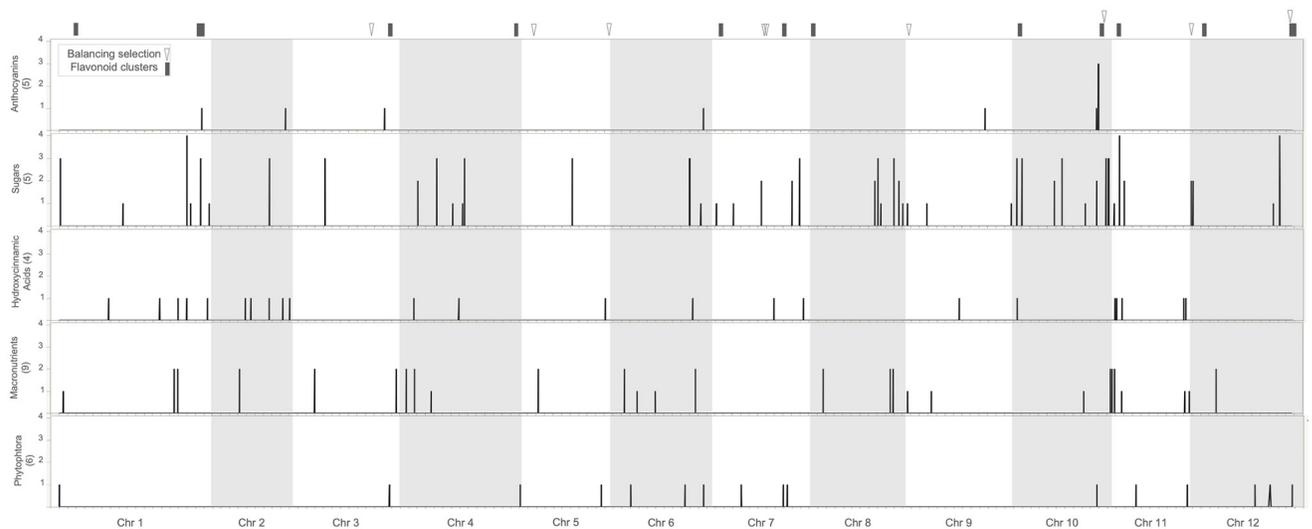


Fig 3. Genomic architecture of variation in anthocyanins and other agronomic traits. Each horizontal panel corresponds to a set of traits, with the number of traits in the set indicated in parenthesis. In the X axis we present the position of the SNPs across the genome and in the Y axis we show the number of traits significant for a particular SNP. In the top we indicate the location of 1 Mb intervals showing clusters (>3 genes) of flavonoid genes as well as footprints of balancing selection according to Tajima's D statistic.

<https://doi.org/10.1371/journal.pone.0250861.g003>

Amino acid metabolism and sugar metabolism are associated to anthocyanin variation according to pathway analysis

Anthocyanin variation is a complex trait determined by interactions among genes that influence the expression of each specific compound. Pathway analyses can help identifying genes with small effects in the phenotype by using previous functional annotation to identify pathways that are enriched in genes with high GWAS associations [114]. We used the PAST software to conduct pathway analyses [50] by using gene annotation from KEGG and PotatoCyc databases. SNPs used in GWAS were assigned to 8,833 genes based on LD information. The genes were associated with 111 PotatoCyc pathways, and 104 sot-KEGG pathways. We thus identified 22 significantly enriched pathways in the GWAS of anthocyanin content in potato tubers (p -value < 0.05, Table 3, S8 Table).

We found an enrichment of genes involved in biosynthesis of methionine and sugars and the degradation of spermidine and spermine. An association between the biosynthesis of the amino acid methionine and anthocyanin has been previously reported in many plants [115–117]. For instance, Dancs and colleagues [118] found that over-expressing a gene involved in methionine synthesis induced a decrease of the expression of *PAL*, which caused a reduction in the amounts of anthocyanin pigments in mutant potato tubers. On the other hand, the catabolism of the sugar glucose via the pentose phosphate pathway (PPP) also has been associated with anthocyanin production in fruits [119]. Stimulating PPP activity in fruits induces an increase in anthocyanin content since some products of the PPP are essential precursors for the production of anthocyanins [120, 121]. The enzyme Glucose-6-phosphate dehydrogenase (G6PDH) plays a particularly important role in this crosstalk between the primary and secondary metabolism and shows a correlation with the levels of mRNAs encoding *PAL* and *CHS* [122]. Finally, the hormone ethylene, which regulates anthocyanin biosynthesis during senescence and stress [122], is derived from spermidine and spermine [123, 124]. Pathway analysis thus revealed plausible links between anthocyanin production and other metabolic and

Table 3. Summary of the pathway-based analysis for pathways with p -value < 0.03 to five anthocyanin compounds in *Solanum tuberosum* group Phureja.

Trait	Data Base	ID	PW Name	p -value	NES	Genes ^a
Pelargonidin	Potatocyc	PWY-6441	Spermine and spermidine degradation III	0,0117	0,67	5
Pelargonidin	Potatocyc	PWY-6596	Adenosine nucleotides degradation I	0,0143	0,71	9
Pelargonidin	KEGG-sot	sot00260	Glycine, serine and threonine metabolism	0,0144	0,40	24
Pelargonidin	KEGG-sot	sot01230	Biosynthesis of amino acids	0,0177	0,27	67
Cyanidin	Potatocyc	PWY-2261	ascorbate glutathione cycle	0,0078	0,67	7
Cyanidin	Potatocyc	PWY-702	L-methionine biosynthesis II	0,0111	0,6	8
Cyanidin	Potatocyc	LEUSYN-PWY	L-leucine biosynthesis	0,0212	0,67	11
Cyanidin	Potatocyc	PWY-6441	Spermine and spermidine degradation III	0,0201	0,63	9
Cyanidin	KEGG-sot	sot00030	Pentose phosphate pathway	0,0105	0,58	10
Cyanidin	KEGG-sot	sot00520	Amino sugar and nucleotide sugar metabolism	0,0111	0,41	24
Cyanidin	KEGG-sot	sot00900	Terpenoid backbone biosynthesis	0,0235	0,41	20
Peonidin	Potatocyc	PWY-6441	Spermine and spermidine degradation III	0,0201	0,62	9
Peonidin	KEGG-sot	sot00030	Pentose phosphate pathway	0,0227	0,56	10
Delphindin	Potatocyc	PWY-6441	spermine and spermidine degradation III	0,0041	0,77	8
Delphindin	Potatocyc	PWY-702	L-methionine biosynthesis II	0,0211	0,59	8
Delphindin	KEGG-sot	sot00030	Pentose phosphate pathway	0,0206	0,58	9
Petunidin	Potatocyc	PWY-7184	Pyrimidine deoxyribonucleotides biosynthesis I	0,0088	0,75	8
Petunidin	Potatocyc	PWY-702	L-methionine biosynthesis II	0,0159	0,61	8
Petunidin	Potatocyc	GLUCOSE	Glucose and glucose-1-phosphate degradation	0,0250	0,63	6
Petunidin	KEGG-sot	sot00030	Pentose phosphate pathway	0,0233	0,53	10

PW pathway, NES normalized enrichment score.

^aThe number of genes that were mapped to a pathway and contributed to the enrichment score calculation.

<https://doi.org/10.1371/journal.pone.0250861.t003>

signaling pathways. Interestingly, according to the literature the PAL enzyme is important to establish these physiological tradeoffs, which could explain its strong association to anthocyanin variation in our study. Given that tradeoffs between the primary and secondary metabolism can impact agronomical attributes we evaluated if anthocyanin content is genetically correlated to other important traits in our potato collection.

Anthocyanin content is correlated to other agronomic traits

We analyzed phenotypic variation for multiple agronomically important traits to identify pathways associated with anthocyanin production. We first evaluated pairwise correlations (S9 Table) between the levels of anthocyanins, macronutrients, sugars, hydroxycinnamic acids (HCAs), and resistance to late blight. We found that the levels of all anthocyanins are positively correlated, which is consistent with the co-localization of significant SNPs for the different compounds and with the linkage between these SNPs and the *PAL* gene. We also found significant correlations between the levels of anthocyanins with tuber content of HCAs (chlorogenic acid, crypto-chlorogenic acid and neo-chlorogenic acid). This correlation is not surprising given that HCAs are chemically conjugated to anthocyanins and anthocyanin-linked HCAs are frequently reported in red skin or flesh of potato tubers [125, 126].

We also analyzed the genomic location of significant SNPs identified with GWAS for the different traits (Fig 3). We found that the putative anthocyanin cluster at chromosome 10 also contains SNPs associated with resistance to *P. infestans* and sugars content. We also identified positions in chromosomes 1, 2, and 6 governing simultaneously anthocyanins and other traits

(Fig 3). The co-localization of QTL for anthocyanins and sugars supports our results of pathway analysis linking the PPP to anthocyanin variation and is consistent with the biochemistry of anthocyanins, which are sugar-decorated [127]. The colocalization of QTL for anthocyanin variation and resistance to late blight is consistent with recent studies showing that anthocyanins have played a role in mounting defenses against Oomycete infection since the divergence of land plant lineages [128]. Overall, these results suggest that breeding strategies aimed at increasing anthocyanin content will likely cause changes in other important traits. This highlights the importance of maintaining genetic diversity to evaluate combinations of genetic variants that produce the most favorable phenotype.

Conclusions

In natural populations, genome wide association studies allow to explore the genetic architecture of complex traits. Here we used accessions of diploid potatoes to identify structural and regulatory genes associated with five anthocyanins. Among these genes, we highlight a *PAL* gene on Chromosome 10 associated with the five-anthocyanin compounds. This gene is contained in a region on chromosome 10 that also harbors other significant SNPs as well as multiple anthocyanin homologs. These results highlight the value of using a diverse collection of native landraces: On one hand genes like *PAL* which are pleiotropic and show evidence of recurrent selection are excellent targets for breeding programs because they have repeatedly tested by selection and produce big changes in the phenotype. The short distance between this gene and multiple MYB TFs associated with anthocyanin regulation in potato, proves that loci identified in QTL mapping can contain multiple causal genes. On the other hand, varieties that do not contain selected variants at these loci can be used to identify novel anthocyanin determinants that could help improve the concentration or expression patterns of anthocyanins during tuber development.

Given that potatoes with high anthocyanin content have multiple origins, we wanted to evaluate if selection in the same alleles or haplotypes at this cluster was involved in the repeated breeding of potatoes with high anthocyanin content. We found that most potatoes with high anthocyanin content share the same genotypes at this cluster, suggesting that there was recurrent selection on the same alleles. However, according to phylogenetic analyses the accumulation of anthocyanin seems to also have involved other alleles. Accordingly, this region has high diversity, consistent with balancing artificial selection to breed varieties with diverse colors.

Finally, we integrated data from multiple traits and used a pathway analysis to find candidate pathways that might be underlying anthocyanin accumulation in potato tubers. The results of this analysis revealed a putative relation between anthocyanin regulation in diploid potato and the biosynthesis of methionine, sugars and hydroxycinnamic acids. The knowledge gained with this complementary analysis has improved the understanding of differences in anthocyanin accumulation and can help identify strategies for increasing anthocyanin production through physiological manipulation, genomic selection, or metabolic engineering.

Supporting information

S1 Fig. Genome wide association study of five anthocyanin compounds in diploid potato. Manhattan Plots of association results from a compression mixed linear model of each anthocyanidin. Negative log₁₀-transformed P-values (y-axis) from a GWAS are plotted against physical position (DM_v4.04_pseudomolecules) on each of 12 chromosomes. (TIF)

S2 Fig. Multivariate analysis of genetic variation in the association panel. Principal components analysis of all SNPs used in the GWAS. The color of each dot indicates the total anthocyanin content (arithmetic sum of the five anthocyanins) for each plant.

(PDF)

S1 Table. Data set of 47,298 SNP markers array. For each polymorphic site we present the diploid genotype of the 96 accessions used.

(XLSX)

S2 Table. Results from the genome-wide association study of 5 anthocyanin compounds in diploid potatoes. Analyses were run with GAPIT using an array of 47,298 SNP markers.

(TXT)

S3 Table. Linkage disequilibrium. Calculated with Tassel for the 47,298 SNP markers evaluated in 96 potato genotypes.

(TXT)

S4 Table. Genomic information for the subset of 108 candidate genes. Putative anthocyanin genes retrieved from the literature.

(XLSX)

S5 Table. GWAS results obtained for SNP markers located nearby anthocyanin genes.

GWAS was run using only SNPs mapped within ± 100 Kb of 108 anthocyanin homologs.

(XLSX)

S6 Table. Statistically significant results of gene-set analysis. Genome-wide association of 5 anthocyanin compounds in SNPs located nearby a pre-selected set of 108 candidate genes.

(XLSX)

S7 Table. LD data. For each marker we calculated the maximum distance in bp to markers associated with a minimum correlation (r^2) of 0.8.

(XLSX)

S8 Table. Association effect and p-values for tagSNPs and genes in the pathways with p-value < 0.05. Results from the PAST software using GWAS results and gene annotation from the PotatoCyc and KEGG databases.

(XLSX)

S9 Table. Trait correlations in mapping populations. Pairwise correlation between traits previously evaluated in this population, namely anthocyanin, macronutrients, sugars, Hydroxycinnamic acids (HCAs), and resistance to late blight.

(TXT)

S10 Table. Distribution of anthocyanin genes in the potato genome. Number and density of anthocyanin homologs across non-overlapping 1 Mb intervals.

(TXT)

S11 Table. Tajima's D across the diploid potato genome. TASSEL was used to calculate Tajima's D statistic as well as genetic diversity (π and θ) using a sliding window (step = 10, window = 10). SNPs falling in the upper and lower 1% percentiles of the distribution were considered candidates for balancing and positive selection respectively.

(TXT)

Acknowledgments

We are grateful to Johan-Sebastian Urquijo for his collaboration in the filtering of genotyping matrices. We are grateful with Adam Thrash and Marilyn Warburton for their helpful advice on using the Past software.

Author Contributions

Conceptualization: Maria Angelica Parra-Galindo, Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez, Federico Roda.

Data curation: Maria Angelica Parra-Galindo.

Formal analysis: Maria Angelica Parra-Galindo, Federico Roda.

Funding acquisition: Federico Roda.

Investigation: Maria Angelica Parra-Galindo, Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez, Federico Roda.

Methodology: Maria Angelica Parra-Galindo, Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez, Federico Roda.

Resources: Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez.

Supervision: Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez, Federico Roda.

Writing – original draft: Maria Angelica Parra-Galindo, Federico Roda.

Writing – review & editing: Maria Angelica Parra-Galindo, Johana Carolina Soto-Sedano, Teresa Mosquera-Vásquez, Federico Roda.

References

1. Wijesinha-Bettoni R, Mouillé B. The Contribution of Potatoes to Global Food Security, Nutrition and Healthy Diets. *American Journal of Potato Research*. Springer; 2019. pp. 139–149. <https://doi.org/10.1007/s12230-018-09697-1>
2. Chun OK, Kim DO, Smith N, Schroeder D, Han JT, Chang YL. Daily consumption of phenolics and total antioxidant capacity from fruit and vegetables in the American diet. *J Sci Food Agric*. 2005; 85: 1715–1724. <https://doi.org/10.1002/jsfa.2176>
3. Pearsall DM. Plant domestication and the shift to agriculture in the Andes. *The handbook of South American archaeology*. Springer; 2008. pp. 105–120.
4. Fernie AR, Yan J. De novo domestication: an alternative route toward new crops for the future. *Mol Plant*. 2019; 12: 615–631. <https://doi.org/10.1016/j.molp.2019.03.016> PMID: 30999078
5. Stokstad E. The new potato. *Science*. 2019; 363: 574–577. <https://doi.org/10.1126/science.363.6427.574> PMID: 30733400
6. Puértolas E, Cregenzán O, Luengo E, Álvarez I, Raso J. Pulsed-electric-field-assisted extraction of anthocyanins from purple-fleshed potato. *Food Chem*. 2013; 136: 1330–1336. <https://doi.org/10.1016/j.foodchem.2012.09.080> PMID: 23194531
7. Wei Q, Wang QY, Feng ZH, Wang B, Zhang YF, Yang Q. Increased accumulation of anthocyanins in transgenic potato tubers by overexpressing the 3GT gene. *Plant Biotechnol Rep*. 2012; 6: 69–75. <https://doi.org/10.1007/s11816-011-0201-4>
8. Camire ME. Potatoes and Human Health. *Advances in potato chemistry and technology: Second Edition*. Taylor & Francis; 2016. pp. 685–704. <https://doi.org/10.1016/B978-0-12-800002-1.00023-6>
9. Lewis CE, Walker JRL, Lancaster JE, Sutton KH. Determination of anthocyanins, flavonoids and phenolic acids in potatoes. I: Coloured cultivars of *Solanum tuberosum* L. *J Sci Food Agric*. 1998; 77: 45–57. [https://doi.org/10.1002/\(SICI\)1097-0010\(199805\)77:1<45::AID-JSFA1>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0010(199805)77:1<45::AID-JSFA1>3.0.CO;2-S)
10. Eichhorn S, Winterhalter P. Anthocyanins from pigmented potato (*Solanum tuberosum* L.) varieties. *Food Research International*. 2005. pp. 943–948. <https://doi.org/10.1016/j.foodres.2005.03.011>

11. Tsuda T, Horio F, Osawa T. The role of anthocyanins as an antioxidant under oxidative stress in rats. *BioFactors*. 2000; pp. 133–139. <https://doi.org/10.1002/biof.5520130122> PMID: 11237172
12. Brown CR. Antioxidants in potato. *Am J Potato Res*. 2005; 82: 163–172. <https://doi.org/10.1007/BF02853654>
13. Han KH, Sekikawa M, Shimada K ichiro, Hashimoto M, Hashimoto N, Noda T, et al. Anthocyanin-rich purple potato flake extract has antioxidant capacity and improves antioxidant potential in rats. *Br J Nutr*. 2006; 96: 1125–1133. <https://doi.org/10.1017/bjn20061928> PMID: 17181888
14. Brown CR, Culley D, Bonierbale M, Amorós W, Amorós W. Anthocyanin, carotenoid content, and antioxidant values in native South American potato cultivars. *HortScience*. 2007; 42: 1733–1736. <https://doi.org/10.21273/hortsci.42.7.1733>
15. Parra-Galindo MA, Piñeros-Niño C, Soto-Sedano JC, Mosquera-Vasquez T. Chromosomes I and X harbor consistent genetic factors associated with the anthocyanin variation in potato. *Agronomy*. 2019; 9: 11–13. <https://doi.org/10.3390/agronomy9070366>
16. CIP. Catalog of ancestral potato varieties from Chugay, La Libertad—Peru. Catalog of ancestral potato varieties from Chugay, La Libertad—Peru. 2015. <https://doi.org/10.4160/9789290604679>
17. Gambardella M. Catálogo de nuevas variedades de papa. 2012. Available: <http://cipotato.org/wp-content/uploads/2013/08/005909.pdf>.
18. Liu Y, Tikunov Y, Schouten RE, Marcelis LFM, Visser RGF, Bovy A. Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. *Front Chem*. 2018; 6. <https://doi.org/10.3389/fchem.2018.00052> PMID: 29594099
19. Li Z, Vickrey TL, McNally MG, Sato SJ, Clemente TE, Mower JP, et al. Assessing anthocyanin biosynthesis in Solanaceae as a model pathway for secondary metabolism. *Genes*. 2019; 10. <https://doi.org/10.3390/genes10080559> PMID: 31349565
20. Tengku N, Dongdong W, Xiaohui M, Yue C, Qin C. Analysis of key genes involved in potato anthocyanin biosynthesis based on genomics and transcriptomics data. *Front Plant Sci*. 2019; 10: 1–12.
21. Strygina K V., Kochetov A V., Khlestkina EK. Genetic control of anthocyanin pigmentation of potato tissues. *BMC Genet*. 2019; 20. <https://doi.org/10.1186/s12863-019-0728-x> PMID: 30885125
22. Zhang H, Yang B, Liu J, Guo D, Hou J, Chen S, et al. Analysis of structural genes and key transcription factors related to anthocyanin biosynthesis in potato tubers. *Sci Hortic (Amsterdam)*. 2017; 225: 310–316. <https://doi.org/10.1016/j.scienta.2017.07.018>
23. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci*. 2015; 20: 176–185. <https://doi.org/10.1016/j.tplants.2014.12.001> PMID: 25577424
24. Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, et al. A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots. *Plant Cell*. 2014; 26: 962–980. <https://doi.org/10.1105/tpc.113.122069> PMID: 24642943
25. Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, et al. Advances in the MYB-bHLH-WD repeat (MBW) pigment regulatory model: addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. *Plant Cell Physiol*. 2017; 58: 1431–1441. <https://doi.org/10.1093/pcp/pcx075> PMID: 28575507
26. Springob K, Nakajima JI, Yamazaki M, Saito K. Recent advances in the biosynthesis and accumulation of anthocyanins. *Nat Prod Rep*. 2003; 20: 288–303. <https://doi.org/10.1039/b109542k> PMID: 12828368
27. De Jong H. Inheritance of anthocyanin pigmentation in the cultivated potato: A critical review. *Am Potato J*. 1991; 68: 585–593. <https://doi.org/10.1007/bf02853712>
28. Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, et al. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011; 475: 189–195. <https://doi.org/10.1038/nature10158> PMID: 21743474
29. van Eck HJ, Jacobs JME, van Dijk J, Stiekema WJ, Jacobsen E. Identification and mapping of three flower colour loci of potato (*S. tuberosum* L.) by RFLP analysis. *Theor Appl Genet*. 1993; 86: 295–300. <https://doi.org/10.1007/BF00222091> PMID: 24193472
30. van Eck HJ, Jacobs JME, Van Den Berg P, Stiekema WJ, Jacobsen E. The inheritance of anthocyanin pigmentation in potato (*Solanum tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. *Heredity (Edinb)*. 1994; 73: 410–421. <https://doi.org/10.1038/hdy.1994.189>
31. Zhang Y, Cheng S, De Jong D, Griffiths H, Halitschke R, De Jong W. The potato R locus codes for dihydroflavonol 4-reductase. *Theor Appl Genet*. 2009; 119: 931–937. <https://doi.org/10.1007/s00122-009-1100-8> PMID: 19588118
32. Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, De Jong WS. The potato P locus codes for flavonoid 3',5'-hydroxylase. *Theor Appl Genet*. 2005; 110: 269–275. <https://doi.org/10.1007/s00122-004-1829-z> PMID: 15565378

33. Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, et al. The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theor Appl Genet*. 2009; 120: 45–57. <https://doi.org/10.1007/s00122-009-1158-3> PMID: 19779693
34. Villano C, Esposito S, D'Amelia V, Garramone R, Alioto D, Zoina A, et al. WRKY genes family study reveals tissue-specific and stress-responsive TFs in wild potato species. *Sci Rep*. 2020; 10: 1–12.
35. Bargmann BORR, Holt SH, Pratt V, Veilleux RE, Tech V, Laimbeer PPE, et al. Characterization of the f locus responsible for floral anthocyanin production in potato. *G3 Genes|Genomes|Genetics*. 2020; 10: 3871–3879. <https://doi.org/10.1534/g3.120.401684> PMID: 32855168
36. Liu Y, Lin-Wang K, Espley R V., Wang L, Yang H, Yu B, et al. Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *J Exp Bot*. 2016; 67: 2159–2176. <https://doi.org/10.1093/jxb/erw014> PMID: 26884602
37. D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, et al. Subfunctionalization of duplicate MYB genes in *Solanum commersonii* generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. *Plant Cell Environ*. 2018; 41: 1038–1051. <https://doi.org/10.1111/pce.12966> PMID: 28386931
38. Payyavula RS, Singh RK, Navarre DA. Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. *J Exp Bot*. 2013; 64: 5115–5131. <https://doi.org/10.1093/jxb/ert303> PMID: 24098049
39. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*. 2013. pp. 1–9.
40. Gupta PK, Kulwal PL, Jaiswal V. Association mapping in plants in the post-GWAS genomics era. *Advances in Genetics*. Elsevier; 2019. pp. 75–154. <https://doi.org/10.1016/bs.adgen.2018.12.001> PMID: 31200809
41. Sun Y V. Integration of biological networks and pathways with genetic association studies. *Hum Genet*. 2012; 131: 1677–1686. <https://doi.org/10.1007/s00439-012-1198-7> PMID: 22777728
42. Jin L, Zuo XY, Su WY, Zhao XL, Yuan MQ, Han LZ, et al. Pathway-based analysis tools for complex diseases: A review. *Genomics, Proteomics and Bioinformatics*. Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China; 2014. pp. 210–220. <https://doi.org/10.1016/j.gpb.2014.10.002> PMID: 25462153
43. White MJ, Yaspan BL, Veatch OJ, Goddard P, Risse-Adams OS, Contreras MG. Strategies for Pathway Analysis Using GWAS and WGS Data. *Curr Protoc Hum Genet*. 2019; 100: 1–17. <https://doi.org/10.1002/cphg.79> PMID: 30387919
44. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 2010; 11: 843–854. <https://doi.org/10.1038/nrg2884> PMID: 21085203
45. Medina I, Montaner D, Bonifaci N, Pujana MA, Carbonell J, Tarraga J, et al. Gene set-based analysis of polymorphisms: Finding pathways or biological processes associated to traits in genome-wide association studies. *Nucleic Acids Res*. 2009; 37: 340–344. <https://doi.org/10.1093/nar/gkp481> PMID: 19502494
46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102: 15545–15550. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517
47. Tang JD, Perkins A, Williams WP, Warburton ML. Using genome-wide associations to identify metabolic pathways involved in maize aflatoxin accumulation resistance. *BMC Genomics*. 2015; 16: 1–12.
48. Mooney M a., Nigg JT, McWeeney SK, Wilmot B. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet*. 2014; 30: 390–400. <https://doi.org/10.1016/j.tig.2014.07.004> PMID: 25154796
49. Zhao H, Nyholt DR, Yang Y, Wang J, Yang Y. Improving the detection of pathways in genome-wide association studies by combined effects of SNPs from Linkage Disequilibrium blocks. *Sci Rep*. 2017; 7: 1–8.
50. Thrash A, Tang JD, Deornellis M, Peterson DG, Warburton ML. PAST: The pathway association studies tool to infer biological meaning from GWAS datasets. *Plants*. 2020; 9: 1–9. <https://doi.org/10.3390/plants9010058> PMID: 31906457
51. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet*. 2012; 44: 1066–1071. <https://doi.org/10.1038/ng.2376> PMID: 22902788

52. Dodds KS, Long DH. The inheritance of colour in diploid potatoes—I. Types of anthocyanidins and their genetic loci. *J Genet.* 1955; 53: 136–149. <https://doi.org/10.1007/BF02981517>
53. Juyó D, Sarmiento F, Álvarez M, Brochero H, Gebhardt C, Mosquera T. Genetic diversity and population structure in diploid potatoes of *Solanum tuberosum* group phureja. *Crop Sci.* 2015; 55: 760–769. <https://doi.org/10.2135/cropsci2014.07.0524>
54. Duarte-Delgado D, Núñez-López CE, Narváez-Cuenca CE, Restrepo-Sánchez LP, Melo SE, Sarmiento F, et al. Natural variation of sucrose, glucose and fructose contents in Colombian genotypes of *Solanum tuberosum* Group Phureja at harvest. *J Sci Food Agric.* 2016; 96: 4288–4294. <https://doi.org/10.1002/jsfa.7783> PMID: 27133474
55. Mosquera Vásquez T, Del Castillo S, Gálvez DC, Rodríguez LE, Vásquez TM, Del Castillo S, et al. Breeding Differently: Participatory Selection and Scaling Up Innovations in Colombia. *Potato Res.* 2017; 60: 361–381. <https://doi.org/10.1007/s11540-018-9389-9>
56. Rojas DKJ, Sedano JCS, Ballvora A, León J, Vásquez TM, Juyo-Rojas D, et al. Novel organ-specific genetic factors for quantitative resistance to late blight in potato. *PLoS One.* 2019; 14: 1–15. <https://doi.org/10.1371/journal.pone.0213818> PMID: 31310605
57. Zhang Z, Ersoz E, Lai C, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet.* 2010; 42: 355–360. <https://doi.org/10.1038/ng.546> PMID: 20208535
58. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. *Plant Genome.* 2016; 9: 0. <https://doi.org/10.3835/plantgenome2015.11.0120> PMID: 27898829
59. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38: 904–9. <https://doi.org/10.1038/ng1847> PMID: 16862161
60. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995; 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
61. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007; 23: 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
62. Scossa F, Roda F, Tohge T, Georgiev MI, Fernie AR. The hot and the colorful: understanding the metabolism, genetics and evolution of consumer preferred metabolic traits in pepper and related species. *CRC Crit Rev Plant Sci.* 2019; 38: 339–381. <https://doi.org/10.1080/07352689.2019.1682791>
63. Altschul S. Basic Local Alignment Search Tool. *J Mol Biol.* 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
64. Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, et al. Genome Reduction Uncovers a Large Dispensable Genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *Plant Cell.* 2016; 28: 388–405. <https://doi.org/10.1105/tpc.15.00538> PMID: 26772996
65. O’dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, et al. The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics.* 2009; 25: 2762–2763. <https://doi.org/10.1093/bioinformatics/btp448> PMID: 19620097
66. Wang K, Li M, Bucan M. Pathway-Based Approaches for Analysis of Genomewide Association Studies. *Am J Hum Genet.* 2007; 81: 1278–1283. <https://doi.org/10.1086/522374> PMID: 17966091
67. Li H, Thrash A, Tang JD, He L, Yan J, Warburton ML. Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J.* 2019; 853–863. <https://doi.org/10.1111/tpj.14282> PMID: 30742331
68. Narváez-Cuenca CE, Peña C, Restrepo-Sánchez LP, Kushalappa A, Mosquera T. Macronutrient contents of potato genotype collections in the *Solanum tuberosum* Group Phureja. *J Food Compos Anal.* 2018; 66: 179–184. <https://doi.org/10.1016/j.jfca.2017.12.019>
69. Ji L, Yogendra KN, Mosa KA, Kushalappa AC, Piñeros-Niño C, Mosquera T, et al. Hydroxycinnamic acid functional ingredients and their biosynthetic genes in tubers of *Solanum tuberosum* Group Phureja. *Cogent Food Agric.* 2016; 2. <https://doi.org/10.1080/23311932.2016.1138595>
70. Berdugo-Cely J, Valbuena RI, Sánchez-Betancourt E, Barrero LS, Yockteng R. Genetic diversity and association mapping in the Colombian central collection of *Solanum tuberosum* L. Andigenum group using SNPs markers. *PLoS One.* 2017; 12. <https://doi.org/10.1371/journal.pone.0173039> PMID: 28257509
71. Gonzales-Vigil E, Bianchetti CM, Phillips GN, Howe GA. Adaptive evolution of threonine deaminase in plant defense against insect herbivores. *Proc Natl Acad Sci U S A.* 2011; 108: 5897–5902. <https://doi.org/10.1073/pnas.1016157108> PMID: 21436043

72. Kombrink E, Schroder M, Hahlbrock K. Several "pathogenesis-related" proteins in potato are 1,3- β -glucanases and chitinases. *Proc Natl Acad Sci*. 1988; 85: 782–786. <https://doi.org/10.1073/pnas.85.3.782> PMID: 16578829
73. Pelletier K, Murrell JR, Shirley BW. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis*. *Plant Physiol*. 1997; 113: 1437–1445. <https://doi.org/10.1104/pp.113.4.1437> PMID: 9112784
74. Zhang Y, Jung CS, De Jong WS. Genetic analysis of pigmented tuber flesh in potato. *Theor Appl Genet*. 2009; 119: 143–150. <https://doi.org/10.1007/s00122-009-1024-3> PMID: 19363602
75. De Jong WS, Eannetta NT, De Jong DM, Bodis M. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. *Theor Appl Genet*. 2004; 108: 423–432. <https://doi.org/10.1007/s00122-003-1455-1> PMID: 14523517
76. Hoballah ME, Gübitz T, Stuurman J, Broger L, Barone M, Mandel T, et al. Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell*. 2007; 19: 779–90. <https://doi.org/10.1105/tpc.106.048694> PMID: 17337627
77. De Vetten N, Quattrocchio F, Mol J, Koes R. The an11 locus controlling flower pigmentation in *petunia* encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev*. 1997; 11: 1422–1434. <https://doi.org/10.1101/gad.11.11.1422> PMID: 9192870
78. Borovsky Y, Oren-Shamir M, Ovadia R, De Jong W, Paran I. The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of *Petunia*. *Theor Appl Genet*. 2004; 109: 23–29. <https://doi.org/10.1007/s00122-004-1625-9> PMID: 14997303
79. Yamagishi M, Shimoyamada Y, Nakatsuka T, Masuda K. Two R2R3-MYB genes, homologs of *petunia* AN2, regulate anthocyanin biosyntheses in flower tepals, tepal spots and leaves of asiatic hybrid Lily. *Plant Cell Physiol*. 2010; 51: 463–474. <https://doi.org/10.1093/pcp/pcq011> PMID: 20118109
80. Zong Y, Zhu X, Liu Z, Xi X, Li G, Cao D, et al. Functional MYB transcription factor encoding gene AN2 is associated with anthocyanin biosynthesis in *Lycium ruthenicum* Murray. *BMC Plant Biol*. 2019; 19: 1–9.
81. Mehrtens F, Kranz H, Bednarek P, Weisshaar B. The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol*. 2005; 138: 1083–1096. <https://doi.org/10.1104/pp.104.058032> PMID: 15923334
82. Wang N, Xu H, Jiang S, Zhang Z, Lu N, Qiu H, et al. MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant J*. 2017; 90: 276–292. <https://doi.org/10.1111/tpj.13487> PMID: 28107780
83. De Jong WS, De Jong DM, De Jong H, Kalazich J, Bodis M. The MicroRNA828/MYB12 Module mediates bicolor pattern development in Asiatic Hybrid Lily (*Lilium* spp.) Flowers. *Theor Appl Genet*. 2003; 107: 1375–1383.
84. Yamagishi M, Uchiyama H, Handa T. Floral pigmentation pattern in Oriental hybrid lily (*Lilium* spp.) cultivar 'Dizzy' is caused by transcriptional regulation of anthocyanin biosynthesis genes. *J Plant Physiol*. 2018; 228: 85–91. <https://doi.org/10.1016/j.jplph.2018.05.008> PMID: 29879604
85. De Jong WS, De Jong DM, De Jong H, Kalazich J, Bodis M. Post-veraison sunlight exposure induces MYB-mediated transcriptional regulation of anthocyanin and flavonol synthesis in berry skins of *Vitis vinifera*. *Theor Appl Genet*. 2003; 107: 1375–1383.
86. Lukowitz W, Mayer U, Jürgens G. Cytokinesis in the *Arabidopsis* embryo involves the syntaxin-related KNOLLE gene product. *Cell*. 1996; 84: 61–71. [https://doi.org/10.1016/s0092-8674\(00\)80993-9](https://doi.org/10.1016/s0092-8674(00)80993-9) PMID: 8548827
87. Mir M, Jimmy D, Sierra B, Ruel K, Pollet B, Johanne C Do, et al. Redirection of the phenylpropanoid pathway to feruloyl malate in *Arabidopsis* mutants deficient for cinnamoyl-CoA reductase 1. 2008; 943–956. <https://doi.org/10.1007/s00425-007-0669-x> PMID: 18046574
88. Wang Z, Ge Q, Wang Z. Concerning the role of cinnamoyl coa reductase gene in phenolic acids biosynthesis in *Salvia miltiorrhiza* 1. *Russ. J. Plant Physiol*. 2017; 64: 553–559. <https://doi.org/10.1134/S1021443717040197>
89. Lavhale SG, Kalunke RM, Giri AP. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta*. Springer; 2018. pp. 1063–1078. <https://doi.org/10.1007/s00425-018-2965-z> PMID: 30078075
90. O'Neill SD, Tong Y, Spörlein B, Forkmann G, Yoder JI. Molecular genetic analysis of chalcone synthase in *Lycopersicon esculentum* and an anthocyanin-deficient mutant. *MGG Mol Gen Genet*. 1990; 224: 279–288. <https://doi.org/10.1007/BF00271562> PMID: 1980524
91. Sasaki N, Nishizaki Y, Ozeki Y, Miyahara T. The role of acyl-glucose in anthocyanin modifications. *Molecules*. 2014. pp. 18747–18766. <https://doi.org/10.3390/molecules191118747> PMID: 25405291

92. Saito K, Yonekura-sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, et al. The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiol Biochem*. 2013; 1–14. <https://doi.org/10.1016/j.plaphy.2013.02.001> PMID: 23473981
93. Faragher J, Chalmers D. Regulation of anthocyanin synthesis in apple skin. Iii. Involvement of phenylalanine ammonia-lyase. *Funct Plant Biol*. 1977; 4: 133. <https://doi.org/10.1071/pp9770133>
94. Reddy VS, Goud KV, Sharma R, Reddy AR. Ultraviolet-B-responsive anthocyanin production in a rice cultivar is associated with a specific phase of phenylalanine ammonia lyase biosynthesis. *Plant Physiol*. 1994; 105: 1059–1066. <https://doi.org/10.1104/pp.105.4.1059> PMID: 12232265
95. Cheng GW, Breen PJ. Activity of Phenylalanine Ammonia-Lyase (PAL) and Concentrations of Anthocyanins and Phenolics in Developing Strawberry Fruit. *J Am Soc Hortic Sci*. 2019; 116: 865–869. <https://doi.org/10.21273/jashs.116.5.865>
96. Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou YH, et al. Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiol*. 2010; 153: 1526–1538. <https://doi.org/10.1104/pp.110.157370> PMID: 20566705
97. Liu Y, Lin-Wang K, Espley R V., Wang L, Li Y, Liu Z, et al. StMYB44 negatively regulates anthocyanin biosynthesis at high temperatures in tuber flesh of potato. *J Exp Bot*. 2019; 70: 3809–3824. <https://doi.org/10.1093/jxb/erz194> PMID: 31020330
98. Gonzalez A, Brown M, Hatlestad G, Akhavan N, Smith T, Hembd A, et al. TTG2 controls the developmental regulation of seed coat tannins in Arabidopsis by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev Biol*. 2016; 419: 54–63. <https://doi.org/10.1016/j.ydbio.2016.03.031> PMID: 27046632
99. Hopkins R, Rausher MD. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature*. 2011; 469: 411–414. <https://doi.org/10.1038/nature09641> PMID: 21217687
100. Iorizzo M, Cavagnaro PF, Bostan H, Zhao Y, Zhang J, Simon PW. A cluster of MYB transcription factors regulates anthocyanin biosynthesis in carrot (*Daucus carota* L.) root and petiole. *Front Plant Sci*. 2019; 9. <https://doi.org/10.3389/fpls.2018.01927> PMID: 30693006
101. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annual Review of Genetics*; 2013. pp. 97–120. <https://doi.org/10.1146/annurev-genet-111212-133526> PMID: 24274750
102. Chae L, Kim T, Nilo-Poyanco R, Rhee SY. Genomic signatures of specialized metabolism in plants. *Science (80-)*. 2014; 344: 510–513. <https://doi.org/10.1126/science.1252076> PMID: 24786077
103. Nützmann H-WW, Osbourn A. Gene clustering in plant specialized metabolism. *Current Opinion in Biotechnology Elsevier*; 2014 pp. 91–99. <https://doi.org/10.1016/j.copbio.2013.10.009> PMID: 24679264
104. Nützmann H-W, Scazzocchio C, Osbourn A. Metabolic gene clusters in Eukaryotes. *Annu Rev Genet*. 2018; 52. <https://doi.org/10.1146/annurev-genet-120417-031237> PMID: 30183405
105. Nützmann HW, Huang A, Osbourn A. Plant metabolic clusters—from genetics to genomics. *New Phytol*. 2016; 211: 771–789. <https://doi.org/10.1111/nph.13981> PMID: 27112429
106. Wurtzel ET, Kutchan TM. Plant metabolism, the diverse chemistry set of the future. *Science (80-)*. 2016; 353: 1232–1236. <https://doi.org/10.1126/science.aad2062> PMID: 27634523
107. Weng JK, Philippe RN, Noel JP. The rise of chemodiversity in plants. *Science*. 2012; 336: 1667–1670. <https://doi.org/10.1126/science.1217411> PMID: 22745420
108. Kliebenstein DJ, Osbourn A. Making new molecules—evolution of pathways for novel metabolites in plants. *Curr Opin Plant Biol*. 2012; 15: 415–423. <https://doi.org/10.1016/j.pbi.2012.05.005> PMID: 22683039
109. Shi M, Xie D. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat Biotechnol*. 2014; 8: 47–60. <https://doi.org/10.2174/1872208307666131218123538> PMID: 24354533
110. Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, et al. Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci U S A*. 2017; 114: E9999–E10008. <https://doi.org/10.1073/pnas.1714380114> PMID: 29087343
111. Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, et al. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat Plants*. 2016; 2. <https://doi.org/10.1038/nplants.2016.74> PMID: 27255838
112. Kiferle C, Fantini E, Bassolino L, Povero G, Spelt C, Buti S, et al. Tomato R2R3-MYB proteins SIANT1 and SIANT2: Same protein activity, different roles. *PLoS One*. 2015; 10: 1–20. <https://doi.org/10.1371/journal.pone.0136365> PMID: 26308527

113. Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Luis Fernández-Alonso J, et al. The origins and adaptation of European potatoes reconstructed from historical genomes. *Nat Ecol Evol.* 2019; 3. <https://doi.org/10.1038/s41559-019-0921-3> PMID: 31235927
114. Cirillo E, Parnell LD, Evelo CT. A review of pathway-based analysis tools that visualize genetic variants. *Front Genet.* 2017; 8: 1–11.
115. Ravanel S, Gakière B, Job D, Douce R. The specific features of methionine biosynthesis and metabolism in plants. *Proc Natl Acad Sci U S A.* 1998; 95: 7805–7812. <https://doi.org/10.1073/pnas.95.13.7805> PMID: 9636232
116. Deikman J, Hammer PE. Induction of anthocyanin accumulation by cytokinins in *Arabidopsis thaliana*. *Plant Physiol.* 1995; 108: 47–57. <https://doi.org/10.1104/pp.108.1.47> PMID: 12228453
117. Pelletier MK, Murrell JR, Shirley BW. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis*: Further evidence for differential regulation of “early” and “late” genes. *Plant Physiol.* 1997; 113: 1437–1445. <https://doi.org/10.1104/pp.113.4.1437> PMID: 9112784
118. Dancs G, Kondrák M, Bánfalvi Z. The effects of enhanced methionine synthesis on amino acid and anthocyanin content of potato tubers. *BMC Plant Biol.* 2008; 8: 1–10.
119. Gianfagna TJ, Berkowitz GA. Glucose catabolism and anthocyanin production in apple fruit. *Phytochemistry.* 1986; 25: 607–609. [https://doi.org/10.1016/0031-9422\(86\)88007-4](https://doi.org/10.1016/0031-9422(86)88007-4)
120. Ju ZG, Yuan YB, Liou CL, Xin SH. Relationships among phenylalanine ammonia-lyase activity, simple phenol concentrations and anthocyanin accumulation in apple. *Sci Hortic (Amsterdam).* 1995; 61: 215–226. [https://doi.org/10.1016/0304-4238\(94\)00739-3](https://doi.org/10.1016/0304-4238(94)00739-3)
121. Jia HJ, Araki A, Okamoto G. Influence of fruit bagging on aroma volatiles and skin coloration of “Hakuho” peach (*Prunus persica* Batsch). *Postharvest Biol Technol.* 2005; 35: 61–68. <https://doi.org/10.1016/j.postharvbio.2004.06.004>
122. Logemann E, Tavernaro A, Schulz W, Somssich IE, Hahlbrock K. UV light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. *Proc Natl Acad Sci U S A.* 2000; 97: 1903–1907. <https://doi.org/10.1073/pnas.97.4.1903> PMID: 10677554
123. Reddy GN, Arteca RN, Dai Y, Flores HE, Negm FB, Pell EJ. Changes in ethylene and polyamines in relation to mRNA levels of the large and small subunits of ribulose biphosphate carboxylase/oxygenase in ozone-stressed potato foliage. *Plant Cell Environ.* 1993; 16: 819–826. <https://doi.org/10.1111/j.1365-3040.1993.tb00503.x>
124. El-Kereamy A, Chervin C, Roustan JP, Cheynier V, Souquet JM, Moutounet M, et al. Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. *Physiol Plant.* 2003; 119: 175–182. <https://doi.org/10.1034/j.1399-3054.2003.00165.x>
125. Andre CM, Oufir M, Guignard C, Hoffmann L, Hausman JF, Evers D, et al. Antioxidant profiling of native Andean potato tubers (*Solanum tuberosum* L.) reveals cultivars with high levels of β -carotene, α -tocopherol, chlorogenic acid, and petanin. *J Agric Food Chem.* 2007; 55: 10839–10849. <https://doi.org/10.1021/jf0726583> PMID: 18044831
126. Ieri F, Innocenti M, Andrenelli L, Vecchio V, Mulinacci N. Rapid HPLC / DAD / MS method to determine phenolic acids, glycoalkaloids and anthocyanins in pigmented potatoes (*Solanum tuberosum* L.) and correlations with variety and geographical origin. *Food Chem.* 2011; 125: 750–759. <https://doi.org/10.1016/j.foodchem.2010.09.009>
127. Tohge T, Zhang Y, Peterek S, Matros A, Rallapalli G, Tandrón YA, et al. Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant J.* 2015; 83: 686–704. <https://doi.org/10.1111/tpj.12920> PMID: 26108615
128. Carella P, Gogleva A, Hoey DJ, Bridgen AJ, Stolze SC, Nakagami H, et al. Conserved biochemical defenses underpin host responses to Oomycete infection in an early-divergent land plant lineage. *Curr Biol.* 2019; 29: 2282–2294.e5. <https://doi.org/10.1016/j.cub.2019.05.078> PMID: 31303485