An Acad Bras Cienc (2022) 94(3): e20210672 DOI 10.1590/0001-3765202220210672
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

CROP SCIENCE

Development of anthocyanin markers: gene mapping, genomic analysis and genetic diversity studies in *Ipomoea* species

CARLA M. ARIZIO, SABRINA M. COSTA-TÁRTARA, IGNACIO M. ZUNINO & MARÍA M. MANIFESTO

Abstract: The anthocyanins are pigments responsible for a wide range of colours in plants, from blue, red and purple, play essential biological roles as well as their genes are evolutionarily conserved. Purple sweet potatoes have anthocyanins as the predominant colour, even though they are present in orange roots masked by carotenoids. Several studies have focused on molecular aspects of anthocyanin genes, mainly in wild *Ipomoea* species, although the structure and segregation analysis of those genes in sweet potato hexaploid species are still unknown. Based on an "exon-primed intron-crossing" (EPIC) approach, fourteen pairs of primers were designed, on five structural anthocyanin genes as candidates. The strategy exploits the Intron Length Polymorphism (ILP) from Candidate Genes (CG), resulting in 93% of successful markers giving scorable and reproducible alleles. The results allowed to define partial structure and sequence of the introns and exons from the selected CG, and to determine patterns of sequence variation. The evaluation of marker dosage and allelic segregations in an *Ipomoea batatas* (L.) Lam mapping population identified several alleles for linkage analysis. The study validated the utility of ILP-CG markers for genetic diversity and conservation applicability and a successful amplification gradient across wild *Ipomoea* species validated their transferability.

Key words: anthocyanin, candidate genes, ILP-CG markers, *Ipomoea batatas*, genetic diversity, wild *Ipomoeas*.

INTRODUCTION

Anthocyanins play important roles in plants conferring accessory pigments for light-harvesting and preventing photo-oxidative damage during photosynthesis. The accumulation of anthocyanin provides orange, yellow, and red colours in flowers, fruits, and other plant organs that attract pollinators and animals for seed dispersal. Both in experimental models, as in *in vitro* tests, clinical and epidemiological investigations, they have shown antioxidant properties and inhibition of various forms of cancer, cardiovascular diseases and neurodegenerative diseases, among others, recognising as a significant ingredient in the human diet with an impact on health (Lila et al. 2016, Lila 2004). Gould (2004) lists the role anthocyanins and related compounds play in protecting against different leaf stressors (light and heavy metals). Furthermore, Shirley (1998) describes the entomophilic role they play in flowers and how they deter pathogens and seed predators. The role of anthocyanin in underground organs is unclear; it may have similar protective role in these reproductive organs (Mano et al. 2007). Several authors (Kano et al. 2005, Suda et al. 2003, Matsui et al. 2002,

Yoshimoto et al. 1999) have focused attention on the multiple physiological functions of the anthocyanins contained in the flesh of purple sweet potatoes (*Ipomoea batatas* (L.) Lam). Mainly in its potent antioxidant activity, antimutagenicity, antihyperglycemic effect, and hepatoprotective and antihypertensive effects.

Anthocyanins are responsible for red and purple skin and flesh colour in sweet potato storage roots (Shi et al. 1992). Sweet potato is a perennial dicotyledonous plant belonging to the family Convolvulaceae, which can give satisfactory yields under adverse climatic and soil conditions and low or non-use of external inputs. Globally, the high yields and wide adaptability made it an important food crop. Sweet potato is a hexaploid (2n = 6 × = 90) outbreeding species with a large and complex genome (2205 Mb) and a high degree of heterozygosis (Peiyong et al. 2016). Although the importance of sweet potato as a source of anthocyanin, relatively little research has focused on inheritance studies whereas that the molecular genetics aspects of anthocyanin genes of this species are more studied to date (Yang et al. 2020, Qin et al. 2020, Amoanimaa-Dede et al. 2020, Tanaka et al. 2012). There are few previous studies focused on the Mendelian genetics of colour in sweet potato. Hernández et al. (1965) found that cooper, rose, pink or purple colours were incompletely dominant over white or cream skin colour. When a parent with roots of white skin colour was crossed with a parent with anthocyanin pigments, 64% of the seedling roots had skin pigmented and transgressive segregation was observed (Arizio et al. 2009a). These studies remained relegated because of some sweet potato complex genetic and reproductive characteristics: polyploid nature, high heterozygosity, self-incompatibility, and a high cross-incompatibility. Consequently, the development of appropriate breeding and selection methodologies using molecular markers were also set aside.

The main genes that encode the anthocyanin biosynthesis pathway enzymes have been isolated in different species (Saito et al. 2013, Petroni & Tonelli 2011). Its regulation has also been characterised in several model species, such as *Petunia hybrida* (Zhang et al. 2019). However, the nucleotide variation or gene expression level and its relationship with anthocyanin pigments production are still unknown in most crops. However, several authors have advanced in characterising different phenotypes and their relationship with genotype. In Solanaceae, an *anthocyaninless* mutant was observed. A premature stop codon is responsible for that phenotype (De Jong et al. 2004). A 390 bp insertion was identified upstream from the putative transcription start site in the pink allele in the anthocyanidin synthase (ANS) gene in *Allium cepa* (Kim et al. 2005). A large insertion in the third exon on F3H gene synthesis of a truncated transcript, resulting in the production of pink flowers rather than purple in *Ipomoea purpurea* (Zufall & Rausher 2003). Tanaka et al. (2004) identified 2 Kb deletion in the intergenic region of the dihydroflavonol 4-reductase gene (DFR-B) in cultivars with high anthocyanin content in *I. batatas* storage roots. Not only SNPs (Single Nucleotide Polymorphisms) and Indels (Insertions and deletion) mutation were responsible for changes in anthocyanin content, but also the presence of different transposable elements inserted within genes. Inagaki et al. (1999), identified the mutable allele a^3 flecked as the DFR-B gene (structural gene of anthocyanin biosynthesis) carrying the 6.4 Kb transposable element Tpn1. In *I. purpurea*, the mutant allele a^f in Chalcone sintetase gen (CHS) is caused by integrating transposable elements, named Tip10 (Habu et al. 1998).

Candidate Genes (CGs) are suspected of having a functional role in the phenotype of interest (Pflieger et al. 2001). DNA sequence variation located within a CG or physically close to it would be associated with trait variation. Exploit CG polymorphism is a valuable alternative to investigate allelic variation involved in sweet potato root pigmentation and as a source to develop new molecular markers for breeding selection. Although different types of molecular markers have been developed for sweet potato (Sasai et al. 2019, Miano et al. 2008, Mcharo et al. 2005), no marker exists for anthocyanin biosynthesis gene. Only a few studies report the use of CG as a strategy for developing functional molecular markers (Arizio et al. 2014, Miller et al. 1999). Mapping CG and identifying their significant allelic associations with a particular trait gives primary insight into the molecular mechanisms and provides robust functional markers (Pflieger et al. 2001). The current availability of plant genomics resources [e.g. Expressed Sequence Tag (EST) databases] and gene function analysis in model organisms are sources for designing those markers which are valuable molecular tools for evolutionary biologists, conservation managers, and agricultural researchers (Reeves et al. 2012).

Even though transcribed regions have been less polymorphic due to DNA sequence conservation within and among plant species (Varshney et al. 2005), this can be overcome, for genetics mapping purposes by exploiting intron polymorphisms. The detection of Intron Length Polymorphisms (ILPs) and Intron Single Nucleotide Polymorphisms (ISNPs) can be done through exon-primed intron-crossing PCR (EPIC-PCR) approach (Wang et al. 2005, Bierne et al. 2000). The proximity of introns to exons makes them especially well-suited for linkage disequilibrium studies that promise to add a powerful new dimension to the understanding and crop improvement.

By exploiting intron length polymorphism (ILP-CG), the present study reports the first set of polymorphic functional markers for anthocyanin genes in *I. batatas* and demonstrated their transferability in six wild *Ipomoea* species. We identified important structural features in the intron regions of CG, like mobile elements and sequence repeats. Also, we provide information about genomic constitution and allelic dosage of CG and their partial structure that might be useful for future studies related to the improvement of *I. batatas* and its wild relatives.

MATERIALS AND METHODS

Plant material

Forty-one accessions of *I. batatas* from INTA *in vitro* collection (Table I) were used to assess the ILP-CG marker's polymorphism. The accessions were chosen under different genetic diversity criteria, geographic origin and storage root colour (Manifesto et al. 2010). We use seven wild *Ipomoea* species (15 accessions) representing two subgenera to test the marker's transferability (Table II). Segregation rates and allelic dosage were evaluated in a double pseudo-testcross mapping population (113 F1 individuals) among two highly heterozygous parents: "30BG" (a clone with white flesh and skin root colour) and "45BG" (a clone with yellow flesh with orange dots and purple skin root colour). Genomic DNA of all plant material was extracted from lyophilised young leaves using a modified CTAB procedure (Doyle & Doyle 1990), and the quality was tested.

Table I. Set of 41 *I. batatas* accessions used for allelic diversity analysis.

	INTA Code (CIP Code)	Root colour (Skin/flesh)		INTA Code (CIP Code)	Root colour (Skin/flesh)
1	2 (CIP3338)	Purple/yellow cream	22	218 (DLP4727)	Cream/white
2	4	Purple/pale yellow	23	225	Cream/cream
3	7 (CIP31)	Cream/cream	24	227 (DLP4709)	White/white
4	15	Purple/dark cream	25	243 (DLP4879)	White/cream
5	20	Cream/pale yellow	26	246 (DLP4793)	Cream/dark cream
6	30 (CIP17)	Cream/cream	27	264 (DLP4803)	Orange/pale orange
7	31 (CIP16)	Cream/cream	28	281	Pink/cream
8	45	Purple/yellow with orange spots	29	324 (DLP4883)	Purple red/cream
9	59 (DLP3983)	White/white	30	358 (CIP9)	Purple/yellow with orange spots
10	83 (DLP4013)	Purple red/orange	31	6	Purple red/dark yellow
11	85 (DLP4017)	Red/yellow cream	32	18	Purple red/dark orange
12	93 (DLP4025)	Purple red/purple	33	35	Cooper/intermediate orange
13	100 (DLP4042)	Cream/cream	34	48	Cooper/cream
14	108 (DLP4053)	White/cream	35	306 (Beauregard)	Cooper/intermediate orange
15	127 (DLP4080)	White/cream	36	336	Red/pale yellow
16	129 (DLP4083)	Cream/white	37	338	Red/ dark yellow
17	134 (DLP4092)	White/cream	38	339	Red/ dark orange
18	148 (DLP4115)	Purple red/cream	39	340	Red/cream
19	166 (DLP4144)	Cream/yellow with orange spots	40	341	Dark Purple/white
20	171	Purple red/cream with purple spots	41	353 (USDA ID531154)	No data
21	178 (DLP4059)	Purple red/purple			

Development, amplification and sequencing validation of Intron Length Polymorphism markers for the Anthocyanins Candidate Genes (ILP-CG)

Fourteen primer pairs were designed to anneal in five structural anthocyanin biosynthesis genes, generating amplicons ranging from 200-600bp (Table III): Chalcone isomerase (CHI), Chalcone synthase (CHS), Flavanone 3-hydroxylase (F3H), Dihydroflavonol 4-reductase (DFR), Anthocyanidin synthase (ANS) and UDP-glucose flavonoid 3-oxy-glucosyltransferase (UF3GT) (Durbin et al. 2000). The protocol for PCR amplification and purification products were according to Arizio et al. (2014).

Table II. Set of wild *Ipomoeas* species used for transferability analysis.

Subgenus	Section	Series	Species	ID
Eriospermum	Eriospermum	Batatas	*Ipomoea grandifolia* (Dammer) O'Donell	CIP460440 CIP460448 CIP460457 CIP460788
			Ipomoea x leucantha Jacq. (pro) sp.	CIP460204 CIP460710
			Ipomoea cordatotriloba	CIP460712 ND
		Jalapae	*Ipomoea amnicola* Morong	CIP470235 CIP470368 ND
Quamoclit	Calonyction	-	*Ipomoea alba* L.	ND
	Tricolores	-	*Ipomoea aristolochiifolia* (Kunth) G. Don	BBC17798 ND
		-	*Ipomoea parasítica* Kunth G. Don	CIP470232

ND: no data.

Identity of the sequences (ABI3130XL, Applied Biosystems) was confirmed through multiple alignments using BLAST (Basic Local Alignment Search Tool). Different types of polymorphism among sequences (SNP, InDels, Simple Sequence Repeat -SSR- and transposons) were also determined using ClustalW (Thompson et al. 1994), BioEdit 7.0.5.3 (Hall 1999), BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and GDR SSR (Genome Database for Rosaceae; http://www.rosaceae.org)

Genetic diversity and segregation ratio analysis

We performed genetic diversity studies in *Ipomoea* accessions to evaluate the discrimination power of the ILP-CG markers. Each band was scored as an individual dominant locus like presence (1) or absence (0) (Rodzen et al. 2004, Zhang et al. 2000) to obtain a binary matrix. Similarity among accessions was determined using Jaccard's coefficient, and a cluster analysis was performed to reveal relationships among them using the UPGMA method (Sneath & Sokal 1973). Pearson's product-moment calculated correlations between arrays through Mantel's test (10,000 permutations) (Mantel 1967). Every analysis was executed using the NTSYSpc 2.11W program (Rohlf 1997). The binary matrix from the F1 population was used for ILP-CG segregation ratio analysis. Both polymorphic and biparental fragments (presence in both parents but segregating in the population) were taken into account. F1 data was analyzed as a double pseudo-testcross (Grattapaglia & Sederoff 1994). The allelic dosage and segregation analysis details are available in Arizio et al. (2014).

RESULTS

ILP-CG markers development and allelic diversity assessment in *I. batatas*

We designed 14 pairs of primers to amplify different introns for five anthocyanin biosynthetic genes of sweet potato using reference sequences of *I. batatas*, *I. nil, Arabidopsis thaliana, Nicotiana tabacum* and *Solanum lycopersicum*, available at NCBI (Figure 1, Table III). PCR products

Table III. List of primers of ILP-GC markers developed for anthocyanins biosynthesis pathway in sweet potato: locus name, corresponding gene, Genbank accession number, primer sequences, and annealing temperature.

Locus name	Gene function	Accession Nº *Ipomoea batatas*	Primer sequences	Annealing temperature
Ib_CHI1	*Chalcone isomerase*	gi:19223827	F: CCGAAGTCAAAGTGGAGAGC R: TTTTCCCGTTCCACTTAACG	56ºC
Ib_CHI3			F: GGATTCCGTCCAATTTTTCA R: GGTGGGAACATTTGGTCACT	58ºC
Ib_CHI5			F: CGATGCAGAGAGCAAAGCTA R: CTTTGCTTCAGGGGAAACAC	52 ºC
Ib_F3H1	Flavanone 3-hydroxylase	gi:119394508	F: GGGATCTTTCAGGTGGTTGA R: AATAGTCCCTTGCCCTCACC	56ºC
Ib_F3H3			F: CCTAAAACGCCACACAGACC R: AGTTCACCACCGCTTGATG	58ºC
Ib_DFR1	Dihydroflavonol 4-reductase	gi:40056989	F: AACGAGGCTACCATGTTCACGC R: TTCACACCCTGCAATGGCT	56ºC
Ib_DFR2			F: ATGTGGCCACCCCTATGGA R: ACACAGGCTTTTGTTGTGGTTGGA	56ºC
Ib_DFR3			F: CAACGTCCAACCACAACAAAAGCCTGTG R: TGTTGGGGTGATGAATGGGCCAACCACT	59ºC
Ib_DFR4			F: CACTAGTGGTTGGCCCATT R: CCTTCTGCTTTGGGATGCTCAT	56ºC
Ib_DFR5			F: CCATCCATGGTTTAGCGGAGATG R: CATTTGCCACAAGCGGGTTAGCT	56ºC
Ib_ANS1	Anthocyanidin synthase	gi:4512586 gi:32441920	F: ATGCTAGTGGGCAGCTTGAGT R: GCTGATTGAGTTGGCTGTTTGGGT	58ºC
Ib_ANS2			F: ACAGCCATCAGCTAACAGCT R: TGGAAGAGTCACTACAGCCATCT	56ºC
Ib_ANS3			F: AGGTGTCATGAGCACATCCT R: ACCGTTGTGGTTAATGCCCTCAG	56ºC
Ib_UF3GT	UDP-glucose flavonoid 3-oxy-glucosyltransferase	AB038248.1	F: GACGGATTCGTTTCTGTGGT R: TTTCAGCTTCTGGTCCAACC	56ºC

amplified in both parents ("30BG" and "45BG") were sequenced in order to confirm the identity of the target regions. Only one pair of primer (Ib-CHI3) amplified alleles without sequence homology. Four pairs of primers amplified a single band, whereas ten pairs amplified from two to four bands in parental genotypes.

Chalcone isomerase (CHI) (EC 5.5.1.6)

The alignment of *P. hybrida* (X14590.1) genomic DNA and cDNA sequences from *I. batatas* (AB037396.1) showed homology (blast e-value: $3e^{-26}$) with the last three exons out of four of *P. hybrida*. Three pairs of primers (Ib-CHI1, Ib-CHI3 and Ib-CHI5) were designed to cover intron one, two and three, respectively. The identity of the exon region was confirmed for Ib-CHI1

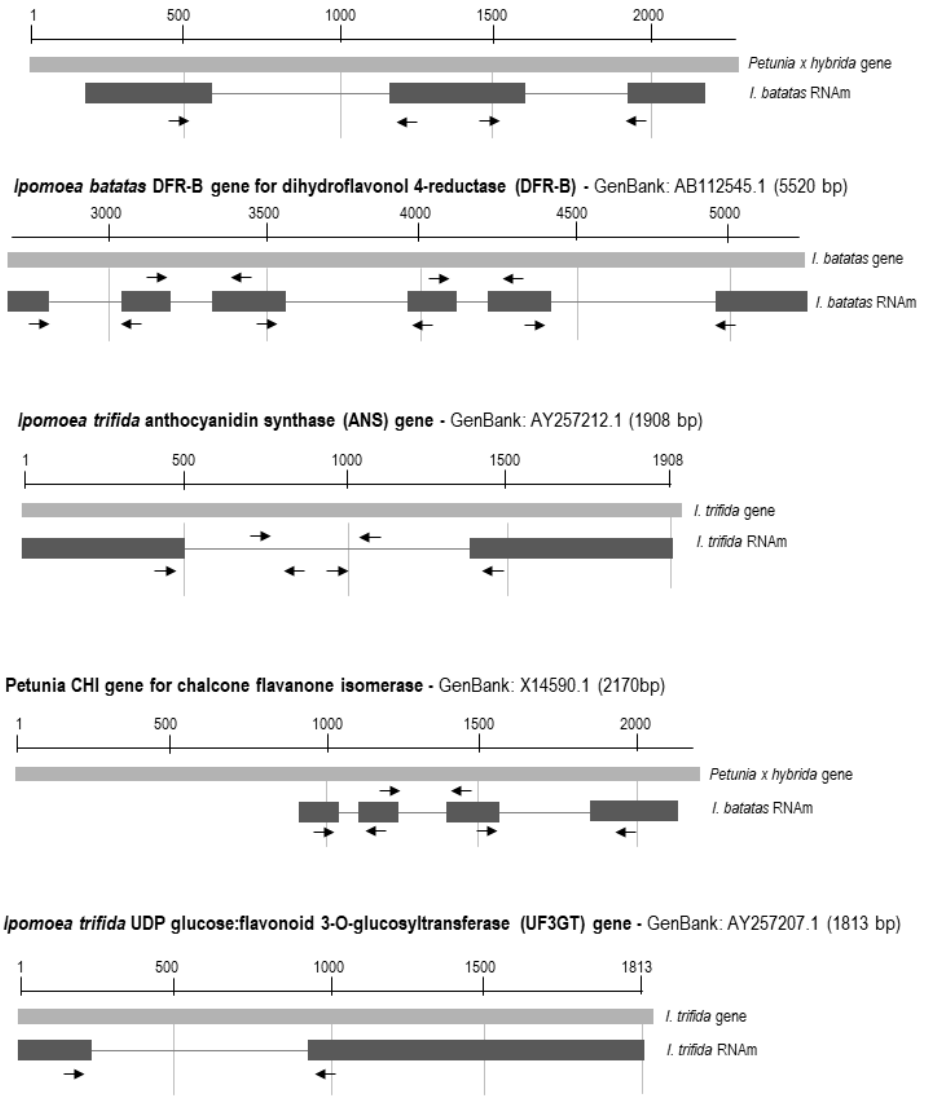

Figure 1. Scheme of the five structural anthocyanin biosynthesis genes. The grey line indicates the gene structure, exons as a black box, and intron as a thin line. Black arrows in opposite directions represent a pair of primers.

and Ib-CHI5 (blast e-value: $5e^{-91}$ and $3e^{-16}$ with AB037396.1, respectively), whereas no sequence homology was detected in the intron region with the GenBank database. No sequence homology was detected for Ib-CHI3 bands since only the intronic region was sequenced. Ib-CHI1 primer amplified one monomorphic band, Ib-CHI3 amplified three bands (two monomorphic and one polymorphic, present in "30BG" parent), whereas Ib-CHI5 amplified one monomorphic and one polymorphic band (present in "30BG").

Flavanone 3-hydroxylase (F3H) (EC 1.14.11.9)

The alignment of the genomic sequences of the F3H gene from *P. hybrida* (AF022142.1) (2320 bp) with the cDNA from *I. batatas* (EF108572.1) showed high homology (blast e-value: $3e^{-115}$) with the three exons of *P. hybrida*. Two pairs of primers were designed for amplified two intron sequences (Ib-F3H1 and Ib-F3H3). Both primers amplified reproducible bands. Ib-F3H1 amplified one polymorphic (312 pb) band in "45BG" and one biparental band. The polymorphic band was sequenced and corresponding with the target sequence, covering the first exon (partially) and first intron. Ib-F3H3 amplified four bands [two

polymorphic in "45BG" (650 pb) and "30BG" and two biparental], covering completely the intron two and partially exon two and three.

Dihydroflavonol 4-reductase (DFR) (EC 1.1.1.219)

The complete DFR gene sequence of *I. batatas* is available at the GenBank database (AB112545.1). The DFR gene presented six exons. Five pairs of primers were designed to assess polymorphism in the introns. They amplified the expected sequences and showed homology only with one member of the DFR gene family: DFR-B. Ib-DFR1 amplified two monomorphic bands (700 bp and 380 bp). The 700 bp band showed homology only with 50 bp of 5` intron one region of DFR-B gene; whereas the rest of alleles sequences displayed homology with different intergenic or intron gene relationship with anthocyanin biosynthesis [*I. trifida* UF3GT gene (AY257207.1) blast e-value: $8e^{-31}$, *I. nil* InNhx pseudogene for Na+/H+ exchanger protein (AB055063.1) blast e-value: $8e^{-31}$, *I. nil* EFP gene for chalcone-flavanone isomerase family protein (AB545802.1) blast e-value:$3e^{-24}$]. Also presented homology with transposable element type *En/Spm*, Tpn101 (AB072271). The 380 bp band presented high homology in all sequences with the DFR-B gene (blast e-value: $6e^{-121}$). Four bands were amplified with Ib-DFR2 (two polymorphic, one biparental and one monomorphic), from which the two-polymorphic (IbDFR2 30BG *b* and *c*) were sequenced. Both bands correspond with intron two and 3' exon three of the DFR-B gene. The band (c) contained the dMELS1, a deleted mobile element-like sequence (53 bp) also detected in DFR-B *I. batatas* by Tanaka et al. (2004) (AB112545.1: 3238 to 3276bp). However, the complete mobile element MELS1 (56 bp) was described first in *I. nil* (AB006793) (Inagaki et al. 1999) and *I. purpurea* (Hisatomi et al. 1997) (Figure 2).

Ib-DFR3 amplified two monomorphic bands (600 bp and 700 bp), and Ib-DFR4 only one monomorphic band (260 bp). All bands presented homology with the target region of *I. batatas* DFR-B gen (blast e-value: $2e^{-99}$ and $1e^{-102}$). Finally, three monomorphic bands were amplified with Ib-DFR5 (450 bp, 400 bp y 370 bp), whose identities were confirmed. All bands present the deleted mobile elements-like sequence dMELS5 (MITEs family). MELS5 (215 bp) was described for the first time in *I. nil* by Inagaki et al. (1999), while dMELDS5 (91 bp) was described in *I. batatas* by Tanaka et al. (2004). Several nucleotide changes were observed within the different bands. The one of 370 bp shows 5 pb insertion and 27 bp deletion along the sequence.

Antocyanidin synthase (ANS) (EC:1.14.11.19)

The sequence of the *I. trifida* gene for Antocyanidin synthase presents high similarity with *I. batatas* mRNA (AB023787.1). Three pairs of primers

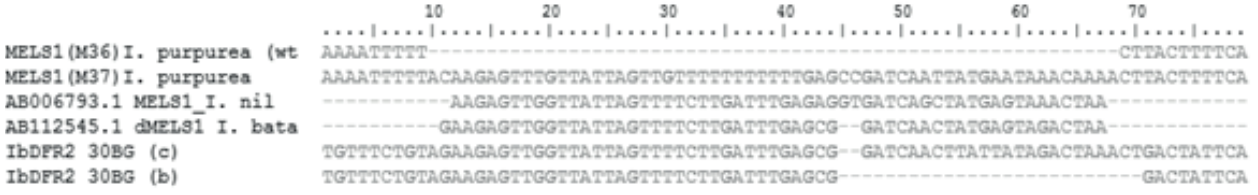

Figure 2. Sequence alignment of two bands corresponding to intron two and 3´ exon three of the DFR-B genes amplified in *I. batatas* (Ib DFR2) to the region of the transposable elements described in *I. nil* (AB006793.1: 9709-9764) and *I. purpurea*. The sequence of dMELS1 shows some base substitutions. Band (b) presented the same dMELS1 and an additional deletion (33 pb). Band (c) also presented dMELS1 with some base substitutions.

(Ib-ANS1, Ib-ANS2 and Ib-ANS3) were designed to cover the intron region, overlapping each other. All of them amplified reproducible and scorable bands. Ib-ANS1 and Ib-ANS2 amplified one, and two monomorphic bands, respectively, meanwhile Ib-ANS3 amplified one polymorphic and two biparental bands. The bands' identity was confirmed by high homology with mRNA sequence from NCBI, except for the second monomorphic band of Ib-ANS2, corresponding to the intron region. The overlapped sequence presents complete homology each other as expected. It was possible to define a complete intron (870 bp) and the flanking exons partially. Different numbers of repeats in the SSR detected were observed in the sequences amplified with Ib-ANS3. The polymorphic bands amplified with Ib-ANS3 have 34 bp insertions without sequence homology and one SSR repeat less (AAAAAT) than the other sequences amplified.

UDP glucose: flavonoid 3-O-glucosyltransferase (UF3TG)

UF3GT gene presents one intron, with differences in length among *Ipomoea* species (746 bp in *I. trifida*, 408 bp in *I. nil* and 77 bp in *I. purpurea)*. The mRNA sequence of UF3GT from *I. batatas* (AB038248.1) has a high similarity with the exon region of several *Ipomoea* species. We designed only one pair of primers to cover the intron region. Ib-UF3GT amplified one monomorphic band (930 bp) and one biparental band (900 bp). The biparental band was sequenced and presented high homology only with the exon and 5' intron region.

Homology validation of ILP-CG markers

To validate PCR products' sequence, we picked up five primer pairs, Ib-CHI1, Ib-DFR1, Ib-DFR2, Ib-DFR3 and Ib-DFR4. The five candidate gene markers were amplified in five wild *Ipomoea* species (*Ipomoea aristolochiifolia, I. alba, I. parasitica, I. amnicola, I. grandifolia*) and also in different accessions of *I. batatas*. The results showed that the sequences were consistent with the selected anthocyanin genes. As expected, the sequences of genes revealed InDels and several point mutations, such as single-base insertions, deletions or translocations, and polymorphism in introns length was observed. Overall, multiple sequence alignment showed that they were homologous and comprised conserved exon regions and non-conserved or variable intron regions. Exon multiple sequence alignment (1940 bp) analysis for amplified CHI gene (Figure 3) from six *Ipomoea* species identifies the presence of 27 substitution mutations resulting in single nucleotide polymorphisms (SNPs). Also, one InDel was detected (3 nucleotides) in *I. aristolochiaefolia*.

The sequence amplified with Ib-DFR1 in *I. parasitica* presents complete homology (blast e-value:$2e^{-110}$) with *I. tricolor* DFR-B gene (Sequence ID: AB267077.1) and partial homology with *I. batatas* DFR gene (3' of intron 1 and 5' of exon 2). *I. grandifolia* presents partial homology with the expected gene; meanwhile, *I. alba* and *I. aristolochiaefolia* do not present homology at all. For Ib-DFR2 (Figure 4), 11 sequences (3746 bp) were analyzed. Wild *Ipomoea* species present better homology with *I. tricolor* (blast e-value:$2e^{-100}$) in the DFR exon region than *I. batatas*, except *I. alba*, which presented homology with *I. nil*. The Ib-DFR2 amplified region presents eighteen SNPs and high polymorphism in intron two with a variable-length (201 bp to 269 bp) and many gaps and SNPs. Also, one SSR motif $(TG)_8$ was detected in *I. alba*.

For Ib-DFR3, ten sequences (5216 bp) were analyzed, five sequences of *I. batatas* (2678 bp) and five of wild *Ipomoea* species (2538 bp). All sequences presented homology in the exon region with *I. batatas* except *I. aristolochiaefolia*, which presented better homology with *I.*

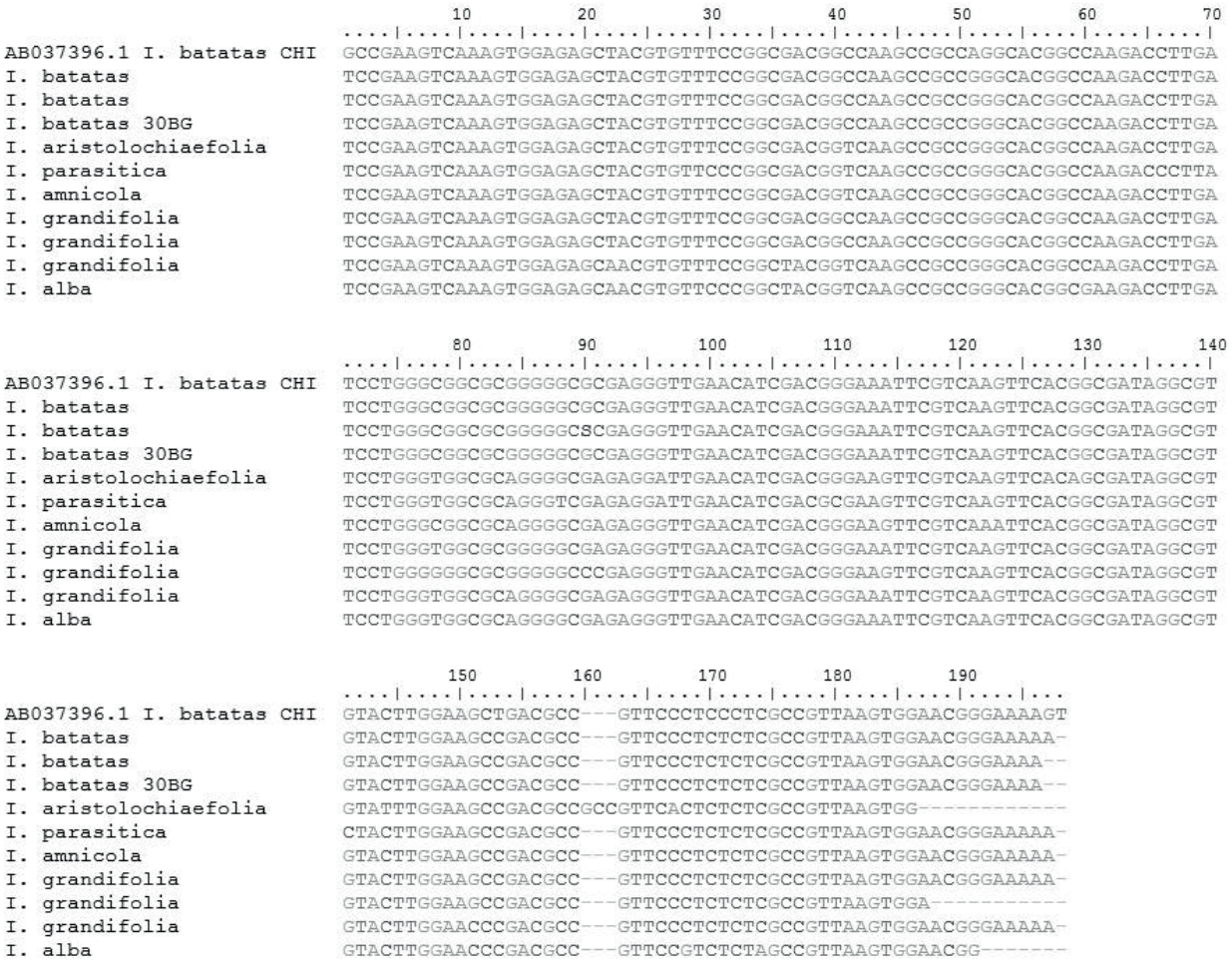

Figure 3. Sequence alignment of different genomic sequences amplified with Ib-CHI-1 (*I. batatas* and five wild *Ipomoea* species) and mRNA of *I. batatas* (AB037396.1). The sequence corresponds to exon region.

purpurea. The intron three presented 418 bp in *I. batatas*; meanwhile, its length in other wild *Ipomoea* species was within 468 bp and 484 bp. The variable-length is due to the presence of many gaps, InDels and SNPs.

Allelic diversity of anthocyanin genes in sweet potato germplasm and cross-species transferability

Thirteen pairs of primers were used to evaluate the genetic variability of 41 sweet potato accessions from INTA *in vitro* collection, and their transferability were tested in 15 wild *Ipomoea* accessions. Ninety-four bands were detected across ILP-CG loci of *Ipomoea* accessions with an average of 7.23 ± 3.06 alleles per polymorphic locus, ranging from 2 (Ib-CHI1) to 13 (Ib-F3H3). Thirty-two alleles were exclusive of *I. batatas*, and forty-four were exclusive of wild *Ipomoea* species, while eighteen bands were shared between both *Ipomoea* groups (Table IV). Two pairs of primers (Ib-DFR1 and Ib-CHI1) were monomorphic, whereas eleven pairs were polymorphic among sweet potato accessions (Ib-ANS1, Ib-ANS2, Ib-ANS3, Ib-F3H1, Ib-F3H3 Ib-CHI3, Ib-UF3GT, Ib-DFR2, Ib-DFR3, Ib-DFR4 and Ib-DFR5). Primers pair Ib-ANS3 and Ib-F3H3 were the most polymorphic markers with 6 and 7 bands, respectively. Fifty bands (92.6%) were polymorphic among the 41 sweet

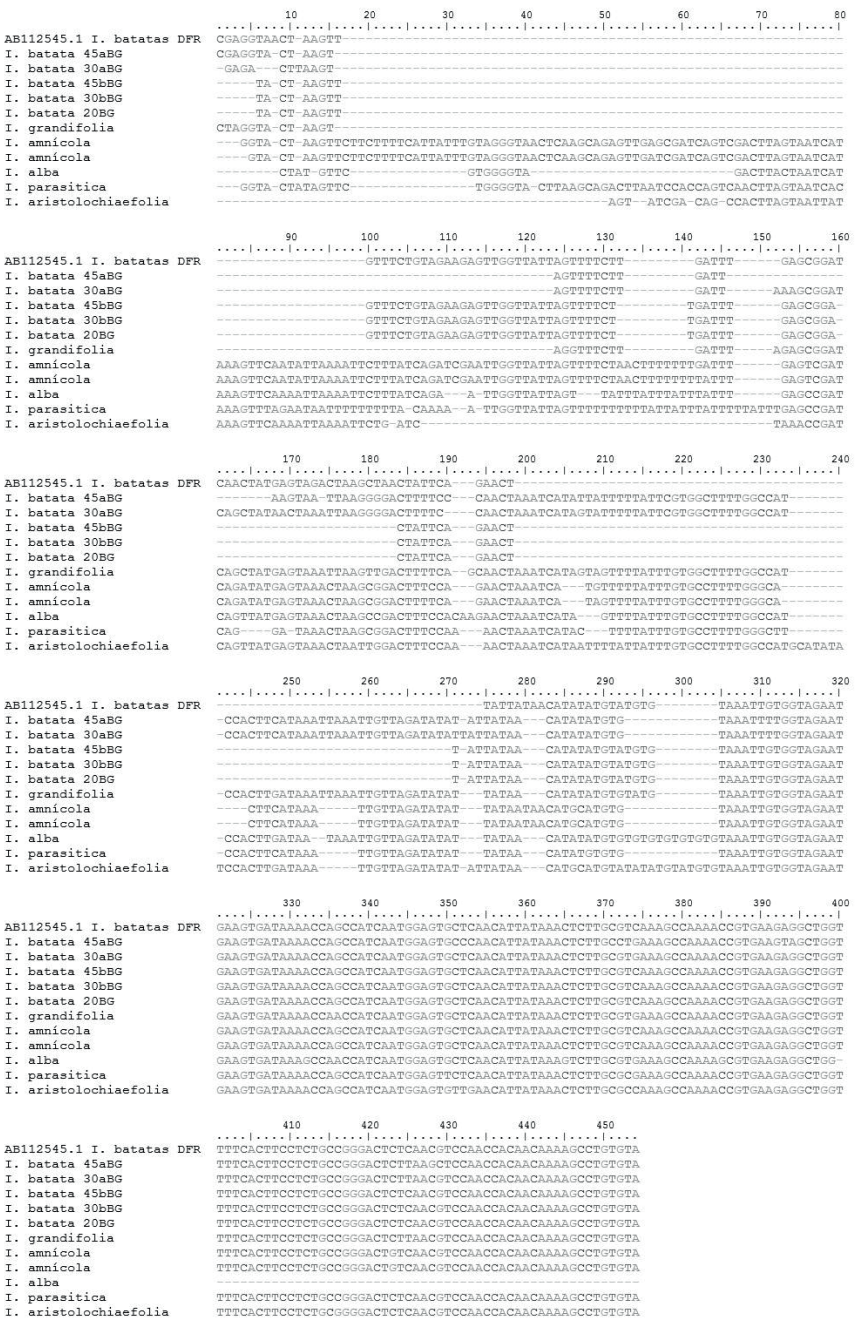

Figure 4. Sequence alignment of different genomic sequences amplified with Ib-DFR-2 (*I. batatas* and five wild *Ipomoea* species) and mRNA of *I. batatas* (AB112545.1). The sequence corresponds to intron 2 and exon 3 regions (288 pb).

potato accessions, while only four bands were monomorphic. Transferability of sweet potato ILP-CG markers were evaluated in terms of successful amplification on seven wild *Ipomoea* species: eleven accessions belong to subgenus *Eriospermum* (*I. grandifolia, I. amnicola, I. cordatotriloba* and *I. leucantha*) and four to subgenus *Quamoclit* (*I. alba, I. parasitica* and *I. aristolochiaefolia*). A range of transferability was observed between wild species. *I. alba* was the species with less transferability (5 n. a.) meanwhile *I. cordatotriloba* amplified all loci like *I. batatas*. Only one locus does not amplify in *I. grandifolia* and *I. aristolochiaefolia,* two loci do not amplify in *I. leucantha,* three loci for *I. amnicola* and four loci for *I. parasitica.*

Table IV. Amplified alleles of ILP-GC markers in *I. batatas* and wild relatives. Sizes are indicated in base pairs (bp).

Primer	Ib-CHI1	Ib-CHI3	Ib-DFR1	Ib-DFR2	Ib-DFR3	Ib-DFR4	Ib-DFR5	Ib-F3H1	Ib-F3H3	Ib-ANS1	Ib-ANS2	Ib-ANS3	Ib-UF3GT
I. batatas	*192*	*1231* *1000* *935* 871	391	341 338 308 303 278	722 714 652 622 585	270 268 265 263	441 431 414	459 435 430 377 367	1300 1160 1058 1000 970 861 700	456 396 331	349 346 338 333 167	517 508 489 482 475 468	930 900
I. parasitica	192	886	264	423	652	257	441 614 500 477	n. a.	768	456 396	n. a.	n. a.	n. a.
I. alba	192	n. a.	331	406	746 585	290	630 575	n. a.	262	456	n. a.	n. a.	n. a.
I. grandifolia	192	n. a.	700 676	338	622 672	265 281 257	441	373 377 367	427 700	456	346	557	900
I. leucantha	192	910 886	676	338	672	281	n. a.	367	1058 700 544 580	456	n. a.	557	1300 920
I. amnicola	192	935 952	403 398	432	700 622 686	251	441 471	n. a.	427	n. a.	346	482	n. a.
I. aristolochiaefolia	195	886	576	376	590 585	290 298	441	304	280	456	346	482	n. a.
I. cordatotriloba	192	886	676	338	672	281	441 431	367	700	456	167	534	1300 920

n. a.: no amplification.

Average genetic similarity based on ILP-CG data across sweet potato accessions and wild *Ipomoea* species was 0.378 ± 0.222. The average similarity among *I. batatas* was 0.562 ± 0.097; meanwhile, among wild *Ipomoea* species was 0.236 ± 0.073. Fifty-two out of 56 accessions (92.8%) were differentiated with 94 polymorphic bands. Cluster analysis shows a clear separation among cultivated species from the wild ones (Figure 5). One cluster grouped 41 accessions of *I. batatas* without any pattern concerning skin and flesh colour. Only two accessions were indistinguishable using these markers. Wild *Ipomoea* species grouped in different clusters. *I. parasitica* belonging to the subgenus *Quamoclit* branched out individually from *I. aristolochiaefolia* and *I. alba* meanwhile *I. grandifolia, I. Leucantha, I. cordatotriloba, I. amnicola* belonging to the subgenus *Eriospermum* joined together in two groups. Dendrogram resulted in a good representation of genetic relationships between accessions. Correlation among similarity matrix and it's

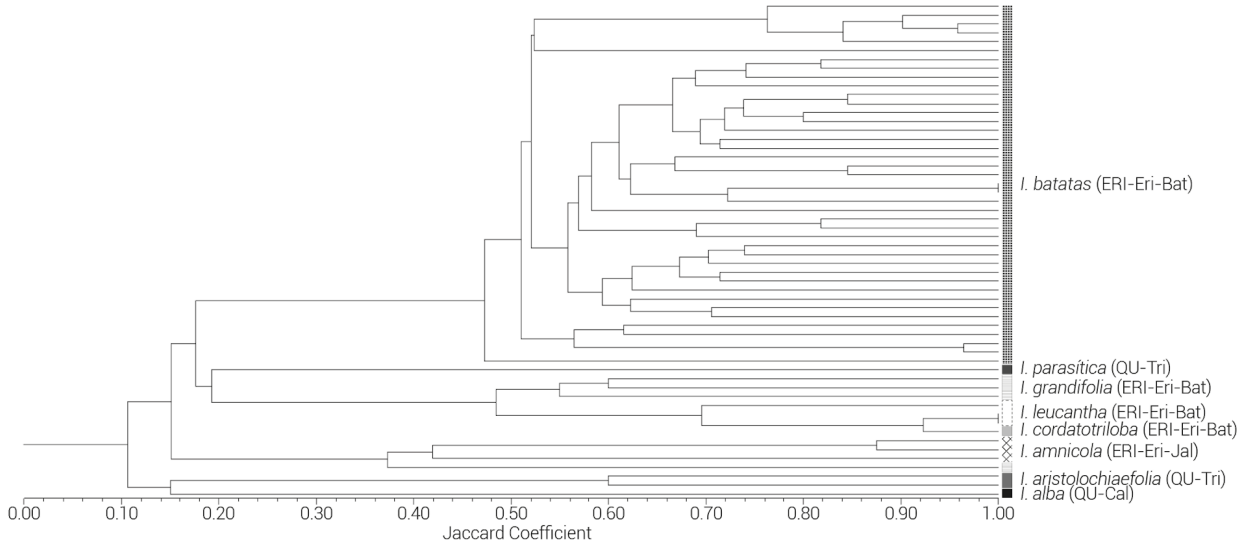

Figure 5. Dendrogram of 41 accessions of sweet potato and 15 wild *Ipomoea* species. Species are followed by abbreviations referring to the classification of Austin and coworkers. Subgenera: ERI=*Eriospermum*, QU=*Quamoclit*; Sections: Tri=*Tricolores*, Cal=*Calonyction*, Eri=*Eriospermum*; Series: Jal=*Jalapae*, Bat=*Batatas*.

cophenetic was high (r = 0.952) and significant (Mantel's test, P < 0.0001).

Segregation analysis of ILP-CG alleles

Considering the segregation model described for an F1-pseudo-test-cross population (Kriegner et al. 2003, Kumagai et al. 1990), we estimated the allele dosage using polymorphic and biparental alleles between parental genotypes ("30BG" and "45BG") and 113 F1 individuals. Seven out of 14 ILP-CG markers corresponding to Ib-DFR2, Ib-F3H1, Ib-F3H3, Ib-CHI3 Ib-CHI5, Ib-UF3GT and Ib-ANS3 showed polymorphic alleles between parents. The allelic dose of eight polymorphic and eight biparental bands and segregation ratio were analyzed (Table V). We identified seven polymorphic simplex bands (1:1) and one polymorphic double band (4:1) meanwhile three polymorphic bands showed distortion in segregation ratio. Four out of eight biparental bands fitted into the 3:1 segregation ratio for double-simplex markers; one fitted into the 9:1 for double-simplex biparental alleles; and three biparental alleles fitted into the 14:1 for double-double biparental alleles. A total of five alleles were monomorphic in 4 loci (Ib-DFR2, Ib-F3H1, Ib-CHI5, Ib-UF3GT), which means no segregation was observed and also their dose is probably high for both parents (AAAaaa /AAAAaa/AAAAAa /AAAAAA).

DISCUSSION

We reported new gene markers in sweet potato related to anthocyanin biosynthesis. The CG strategy, combined with the EPIC-PCR technique, revealed polymorphism in all functionally important genes selected as candidate loci. The structure and sequence of five anthocyanin genes were partially defined. We developed, amplified and characterized gene-specific molecular markers CHI, DFR, F3H, ANS and UF3GT genes from *Ipomoea* species. The markers were validated in their use to evaluate genetic diversity in germplasm collections and genetic mapping. The successful marker transferability in wild *Ipomoea* species extends gene markers usefulness.

The ILP-CG molecular markers are codominant, locus-specific, and more

Table V. Segregation ratio and putative allelic dosage in *Ipomoea batatas* with ILP-GC markers.

Locus	Number of alleles	Observed frequency[a]	Allele type	Segregation ratio	χ^{2b}	Potential allelic dosage
Ib-DFR2	1	51/61	Polymorphic	1:1	0.8928 NS	Aaaaaa x aaaaaa
	2	50/63	Polymorphic	1:1	1.4956 NS	Aaaaaa x aaaaaa
	3	83/30	Biparental	3:1	0.1450 NS	Aaaaaa x Aaaaaa
Ib-F3H1	1	99/14	Polymorphic	4:1	4.0907 NS	AAaaaa x aaaaaa
Ib-F3H3	1	52/54	Polymorphic	1:1	0.0377 NS	Aaaaaa x aaaaaa
	2	56/50	Polymorphic	1:1	0.3396 NS	Aaaaaa x aaaaaa
	3	99/7	Biparental	14:1	0.000 NS	AAaaaa x AAaaaa
	4	98/6	Biparental	14:1	0.13 NS	AAaaaa x AAaaaa
Ib-CHI3	1	51/51	Polymorphic	1:1	0.000 NS	Aaaaaa x aaaaaa
	2	91/10	Biparental	9:1	0.004 NS	AAaaaa x Aaaaaa/AAaaaa x AAaaaa
	3	51/51	Biparental	3:1	- *	------ x aaaaaa (distortioned)
Ib-CHI5	1	46/55	Polymorphic	1:1	0.8019 NS	Aaaaaa x aaaaaa
Ib-UF3GT	1	78/31	Biparental	3:1	0.688 NS	Aaaaaa x Aaaaaa
Ib-ANS3	1	62/59	Polymorphic	1:1	1.522 NS	Aaaaaa x aaaaaa
	2	108/3	Biparental	14:1	2.80 NS	AAaaaa x AAaaaa
	3	84/27	Biparental	3:1	0.027 NS	Aaaaaa x Aaaaaa

[a]Presence/absence of bands. [b]*p*-value (NS: not significant; *: significant).

reproducible. The strategy for molecular primer design ILP-CG markers was successful. Ninety-three per cent (93%) of the markers effectively amplified the target region, only one pair of primer amplified alleles without sequence homology. The same strategy was also successful for carotenoid genes (79%) (Arizio et al. 2014). Similar results were obtained with the EST-SSR strategy (84.6%) (Wang et al. 2011). Among the successful primer pairs, all amplified PCR products were of the expected sizes in *I. batatas*, although some of them resulted in larger or smaller PCR products than expected, particularly in wild *Ipomoea* species. These length variations are attributed mainly to polymorphism in introns (insertions/deletions, repeat number variations or transposons). We determined the existences of SNPs, InDels, SSR and transposable elements within DNA sequences of CG. Introns have less evolutionary constraint than exons and should, be more likely to present polymorphism (Feltus et al. 2006). Moreover, nuclear introns represent a largely untapped source of genetic variation for population genetics and phylogeography. By examining the fourteen ILP-CG marker intron sequences, we found that only one of the ILP were due to SSR motif variation. Since there was no overlapping between ILPs and SSRs, new Sequence-Tagged Sites (STS) markers could be developed from ILP as a complement for SSR

markers. According to Hu et al. (2004a, b) some polyploids species showed low efficiency of PCR amplification due to the complexity of the genome and the high percentage of repetitive DNA sequences. The successful results obtained with the ILP-CG approach reflected in the high efficiency of PCR amplification and homology validation prove that it is appropriate to develop molecular markers in polyploid.

ILP-CG markers can serve as anchor loci for linking sweet potato genetic maps and localizing QTL for different segregating characteristics like differential anthocyanin accumulation. An essential step for the analysis in polyploidy genetics studies is to estimate a dominant band copy number (also referred as a marker dosage) based on the assumptions of known ploidy and random pairing of chromosomes (George 2009). Constructing genetic maps in an autopolyploid demand organised loci along individual chromosomes, which also must be assigned to homologous groups (HGs). For linkage analysis, simplex markers were used to construct framework linkage maps, but double-simplex markers (Aaaaaa x Aaaaaa) together with duplex and triplex markers (AAAAAA × AAAaaa) were employed to identify HGs. The best approach to develop a mapping population for species with self-incompatibility and a high level of heterozygosity is through a pseudo-test-cross population (Grattapaglia & Sederoff 1994). We detected "simplex" and "double-simplex" (biparental) (3:1) alleles together with other biparental alleles with different segregation ratio. Two linkage maps are developed when a pseudo test cross approach is used. The biparental alleles or bridging markers (Isobe et al. 2009) allows merging one integrated linkage map. Multiplex markers (higher dose markers) provide connections between linkage groups (Ripol et al. 1999) identifying and merging homologous co-segregation groups. We detected allele high-dose within the ILP-CG developed. Over the last two decades, seven molecular marker linkage maps have been developed for sweet potato using different molecular markers and, in general, using pseudo-test-cross population (Kim et al. 2017, Shirasawa et al. 2017, Monden et al. 2015, Zhao et al. 2013, Li et al. 2010, Cervantes-Flores et al. 2008, Kriegner et al. 2003, Mwanga et al. 2002, Ukoskit & Thompson 1997). Meanwhile, Shirasawa et al. (2017) report the first high-density SNP linkage map using a population derived from a single parent's self-pollination. They successfully constructed a genetic linkage map with many double-simplex SNPs (biparental) and simplex alleles. However, they cannot identify HGs through the classical approach because no anchoring markers were available. Monden et al. (2015) analyze segregation ratios of molecular markers in sweet potato obtained from different studies. The simplex markers were between 43% (Zhao et al. 2013) and 88% using retrotransposons markers. The simplex alleles detected by ILP-CG markers is less (34%) than the report in published linkages maps, but we also amplified biparental and potential multiplex segregations alleles, all of them, corresponding to CG. In polyploid organisms, the study of allelic transmission and genetic linkage is considerably complicated due to the range of meiotic configurations (Gallais 2003, Sybenga 1975). In *I. batatas*, the nature of the inheritance is discussed, although an autopolyploid hypothesis is supported (Ukoskit & Thompson 1997, Shiotani 1987, Nishiyama et al. 1975). The sweet potato genome constitution postulated by Shiotani & Kawase (1989) is "B1B1B1B1B2B2" suggesting certain homeology between the B1 and the B2 genomes, based on the frequency of tetravalent and hexavalent formations observed in cytogenetic studies. More recently, Mollinari et al. (2019) showed polysomic inheritance and demonstrated the

autopolyploid origin of sweet potato. Our data support the polysomic inheritance proposed for sweet potato as an autopolyploid species.

We evaluate the usefulness of gene-specific ILP-CG markers for genetic diversity analysis in *Ipomoea* species. Thirteen ILP-CG loci out of fourteen primers designed amplified scorable alleles, from which eleven loci were polymorphic and useful for differentiating sweet potato genotypes belonging to a germplasm collection. Although ILP-CG markers are expected to be less polymorphic than neutral markers -like SSR- because they are designed on conserved genes, the allelic diversity detected was enough to discriminate thirty-nine sweet potato genotypes. The results can be explained by the polyploid nature and the reproductive system (sporophytic self-incompatible) of the species, which increases the chances of allelic diversity. Arizio et al. (2009b) and Monteros-Altamirano et al. (2020) detected a similar number of bands (57 and 89, respectively) with only a few SSR markers (7 and 8, respectively). The results confirmed the usefulness of ILP-CG markers to assess the genetic diversity of *I. batatas* for genetic conservation, increasing the availability of molecular markers for genes involved in pigment biosynthesis (Arizio et al. 2014). The polymorphism in conserved nuclear genes was also used in plant phylogenetic analysis (Koch et al. 2001, Galloway et al. 1998, Gout & Clegg 1991) as nuclear gene encoding β-amylase in the series Batatas (Rajapakse et al. 2004). Carotenoid biosynthesis genes have been used to analyze evidence of geographical subdivision and linkage disequilibrium in carrot (Clotault et al. 2010). Undoubtedly, ILP-CG markers have a wide range of uses.

ILP-CG markers amplified in fifteen wild *Ipomoea* species belonging to seven species showed successful transferability. A high degree of ILP locus conservation with transferability ranging from 61.54 to 100% was observed and greater than reported by Almeida et al. (2014) (55%) and by Jayaswall et al. (2019) (58.7%). All markers amplified at least eight *Ipomoea* species, indicating different levels of transferability. *I. cordatotriloba* are the most closed to *I. batatas* based on phylogenetics analysis with ITS and waxy sequences (Miller et al. 1999) and amplified all loci presenting similar and equal alleles through different loci. The other *Ipomoeas* have different degrees of positive amplification. The success in transferability indicates high conservation of primer binding sites in genomic DNA over a long evolutionary history. Although we detected several polymorphisms in intron regions, sequence analysis showed a high degree of similarity between exons of wild *Ipomoea* species. Considering that exon-intron structures are highly conservative (Xi et al. 2011, Yang et al. 2007), the ILP-CG markers can be developed using predicted intron positions in other plant species. In this study, 15 introns predicted from the genome of model species (Arabidopsis, Lycopersicum, and Citrus sp.) appear in sweet potato as expected. These characteristics make ILP-CG markers more transferable among species than others, which was verified in Magnoliophyta. EPIC-PCR approach to develop primers in rice amplified successfully orthologous sequences in other monocot species showing high transferability as well as in some dicot species (Wang et al. 2005).

Cluster analysis among *Ipomoea* species shows groups according to the *Ipomoea* genome constitution ("A" or "B") based on selfing abilities, interspecific crossing capabilities, as well as morphological and cytological analysis (Oracion et al. 1990, Nishiyama et al. 1975, Martin & Jones 1973, Jones 1965). *I. batatas* accessions that possess B-genome and self-incompatibility behaviour joined in one group whereas wild *Ipomoea* species which have A-genome and

self-compatibility behaviour joined in three different groups clustering *Eriospermum* and *Quamoclit,* two from the three subgenera recognised for *Ipomoea* genus (Austin & Bianchini 1998, Austin 1997, Austin & Huáman 1996).

ILP-CG markers enrich the current genomic resources for sweet potato and related wild species. The functional markers developed for anthocyanins genes in *Ipomoea* species provide a new molecular tool for linkage and comparative genetic mapping, genetic diversity and germplasm conservation. They also may be useful for marker-assisted selection in breeding programs.

Acknowledgments

We wish to thank PhD Acuña Cintia for her critical reading of this paper. This work was supported by Instituto Nacional de Tecnología Agropecuaria.

REFERENCES

ALMEIDA NF, LEITÃO ST, CAMINERO C, TORRES AM, RUBIALES D & VAZ PATO MC. 2014. Transferability of molecular markers from major legumes to *Lathyrus* spp. for their application in mapping and diversity studies. Mol Biol Rep 41: 269-283.

AMOANIMAA-DEDE H, CHUNTAO S, AKWASI Y, CHUNHUA C, SHAOXIA Y, HONGBO Z & MIAO C. 2020. Flesh Color Diversity of Sweet Potato: An Overview of the Composition, Functions, Biosynthesis, and Gene Regulation of the Major Pigments. Phyton 89(4): 805-833.

ARIZIO CM, COSTA TÁRTARA SM & MANIFESTO MM. 2014. Carotenoids gene markers for sweet potato (*Ipomoea batatas* L. Lam): applications in genetic mapping, diversity evaluation and cross-species transference. Mol Genet Genomics 289(2): 237-251.

ARIZIO CM, MANIFESTO MMM & MARTI H. 2009a. Análisis de caracteres relacionados con el color de la raíz engrosada en un cruzamiento de dos clones de *Ipomoea batatas* L. (Lam.). Hortic Argen 28: 5-13.

ARIZIO CM, HOMPANERA N, SUAREZ E & MANIFESTO M. 2009b. Genotypic identification and diversity evaluation of a sweet potato (*Ipomoea batatas* (L). Lam) collection using microsatellites. Plant Genet Resour 7(2): 135-138.

AUSTIN DF. 1997. Nomenclator Ipomoeeae. https://cals.arizona.edu/herbarium/sites/cals.arizona.edu.herbarium/files/old_site/assoc/people/daustin/nomen-1.htm.

AUSTIN DF & BIANCHINI RS. 1998. Additions and corrections in American Ipomoea (Convolvulaceae). Taxon 47: 833-838.

AUSTIN DF & HUÁMAN Z. 1996. A synopsis of Ipomoea (Convolvulaceae) in the Americas. Taxon 45: 3-38.

BIERNE N, LEHNERT SA & BEDIER E. 2000. Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. Mol Ecol 9: 233-235.

CERVANTES-FLORES JC, YENCHO CG, KRIEGNER A, PECOTA KV, FAULK MA, MWANGA ROM & SOSINSKI BR. 2008. Development of a genetic linkage map and identification of homologous linkage groups in sweet potato using multiple-dose AFLP markers. Mol Breeding 21: 511-532.

CLOTAULT J, GEOFFRIAU E, LIONNETON E, BRIARD M & PELTIER D. 2010. Carotenoid biosynthesis genes provide evidence of geographical subdivision and extensive linkage disequilibrium in the carrot. Theor Appl Genet 121: 659-672.

DE JONG WS, EANNETTA NT, DE JONG DM & BODIS M. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor Appl Genet 108: 423-432.

DOYLE JJ & DOYLE JL. 1990. Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

DURBIN ML, MCCAIG B & CLEGG MT. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol Biol 42: 79-92.

FELTUS FA, SINGH HP, LOHITHASWA HC, SCHULZE SR, SILVA TD & PATERSON AH. 2006. A Comparative Genomics Strategy for Targeted Discovery of Single-Nucleotide Polymorphisms and Conserved-Noncoding Sequences in Orphan Crops. Plant Physiol 140(4): 1183-1191.

GALLAIS A. 2003. Quantitative genetics and breeding methods in autopolyploid plants. INRA Editions, Paris, France, p. 30-31.

GALLOWAY GL, MALMBERG RL & PRICE RA. 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. Mol Biol Evol 15: 1312-1320.

GEORGE AW. 2009. Estimation of copy number in polyploidy plants: the good, the bad, and the ugly. Theor Appl Genet 119: 483-496.

GOULD KS. 2004. Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. J Biomed Biotechnol 5: 314-320.

GOUT BS & CLEGG TM. 1991. Molecular evolution of alcohol dehydrogenase in members of the grass family. PNAS 88: 2060-2064.

GRATTAPAGLIA D & SEDEROFF R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121-1137.

HABU Y, HISATOMI Y & IIDA S. 1998. Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. The Plant J 16(3): 371-376.

HALL TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.

HERNÁNDEZ TP, HERNÁNDEZ T, CONSTANTIN R & MILLER J. 1965. Inheritance of and method of rating flesh color in *Ipomoea batatas*. Proc Amer Soc Hort Sci 1965(87): 387-390.

HISATOMI Y, YONEDA Y, KASAHARA K, INAGAKI Y & IIDA S. 1997. DNA rearrangements at the region of the dihydroflavonol 4-reductase gene for flower pigmentation and incomplete dominance in morning glory carrying the mutable flaked mutation. Theor Appl Genet 95: 509-515.

HU J, NAKATANI M, GARCIA LALUSIN A & FUJIMURA T. 2004a. New microsatellite markers developed from reported *Ipomoea trifida* sequences and their application to sweet potato and its related wild species. Sci Hortic 102: 375-386.

HU J, NAKATANI M, MIZUMO K & FUJIMURA T. 2004b. Development and characterization of microsatellite markers in sweet potato. Breed 54: 177-188.

INAGAKI Y, JOHZUKA-HISATOMI Y, MORI T, TAKAHASHI S, HAYAKAWA Y, PEYACHOKNAGUL S, OZEKI Y & IILA S. 1999. Genomic organization of the genes enconding dihydroflavonol 4-reductase for flower pigmentation in the japanese and common morning glories. Gene 265: 1-10.

ISOBE S, KÖLLIKER R, HISANO H, SASAMOTO S, WADA T, KLIMENKO I, OKUMURA K & TABATA S. 2009. Construction of a consensus linkage map for red clover (Trifolium pratense L.). BMC Plant Biol 9: 57.

JAYASWALL K, SHARMA H, BHANDAWAT A, SAGAR R, YADAR K, SHARMA V, MAHAJAN V, ROY J & SINGH M. 2019. Development of intron length polymorphic (ILP) markers in onion (*Allium cepa* L.), and their cross-species transferability in garlic (*A. sativum* L.) and wild relatives. Genet Resour Crop Evol 66: 1379-1388.

JONES A. 1965. Cytological observations and fertility measurements of sweet potato (*Ipomoea batatas (L.*) Lam.) PNAS 86: 527-537.

KANO M, TAKAYANAGI T, HARADA K, MAKINO K & ISHIKAWA F. 2005. Antioxidative activity of anthocyanins from purple sweet potato, *Ipomoea batatas* cultivar Ayamurasaki. Biosci Biotechnol Biochem 69: 979-988.

KIM JH, CHUNG IK & KIM KM. 2017. Construction of a genetic map using EST-SSR markers and QTL analysis of major agronomic characters in hexaploid sweet potato (*Ipomoea batatas* (L.) Lam). PLoS ONE 12(10): e0185073.

KIM S, JONES R, YOO K & PIKE L. 2005. The L locus, one of complementary genes required for anthocyanin production in onions (*Allium cepa*), encodes anthocyanidin synthase. Theor Appl Genet 111: 120-127.

KOCH M, HAUBOLD B & MITCHELL-OLDS T. 2001. Molecular systematics of the Brassicaceae: Evidence from coding plastidic matk and nuclear CHS sequences. American J Bot 88(2): 534-544.

KRIEGNER A, CERVANTES JC, BURG K, MWANGA ROM & ZHANG D. 2003. A genetic linkage map of sweet potato [*Ipomoea batatas (L.*) Lam.] based on AFLP markers. Mol Breed 11: 169-185.

KUMAGAI T, UMEMURA Y, BABA T & IWANAGA M. 1990. The inheritance of beta-amylase null in storage roots of sweet potato, *Ipomoea batatas (L.)* Lam. Theor Appl Genet 79: 369-376.

LI AX, LIU QC, WANG QM, ZHANG LM, ZHAI H & LIU SZ. 2010. Construction of molecular linkage maps using SRAP markers in sweet potato. Act Agron 36(8): 1286-1295.

LILA MA. 2004. Anthocyanins and Human Health: An In Vitro Investigative Approach. J Biomed Biotechnol 5: 306-313.

LILA MA, BURTON-FREEMAN B, MARY GRACE & KALT W. 2016. Unraveling Anthocyanin Bioavailability for Human Health. Annu Rev Food Sci Technol 7(1): 375-393.

MANIFESTO MM, COSTA TARTARA SM, ARIZIO CM, ALVAREZ MA & HOMPANERA NR. 2010. Analysis of the morphological attributes of a sweet potato collection. Ann Appl Biol 157(2): 273-281.

MANO H, OGASAWARA F, SATO K, HIGO H & MINOBE Y. 2007. Isolation of a Regulatory Gene of Anthocyanin Biosynthesis in Tuberous Roots of Purple-Fleshed sweet potato. Plant Physiol 143(3): 1252-1268.

MANTEL NA. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27: 209-220.

MARTIN FW & JONES A. 1973. The species of *Ipomoea* closely related to the sweet potato. Econ Bot 26: 201-215.

MATSUI T, EBUCHI S, KOBAYASHI M, FUKUI K, SUGITA K, TERAHARA N & MATSUMOTO K. 2002. Anti-hyperglycemic effect of diacylated anthocyanin derived from *Ipomoea batatas* cultivar Ayamurasaki can be achieved through the alpha-glucosidase inhibitory action. J Agric Food Chem 50: 7244-7248.

MCHARO M, LABONTE DR, MWANGA ROM & KREIGNER A. 2005. Associating molecular markers with virus resistance to classify sweet potato genotypes. J Am Soc Hort Sci 130: 355-359.

MIANO DW, LABONTE DR & CLARK CA. 2008. Identification of molecular markers associated with sweet potato resistance to sweet potato virus disease in Kenya. Euphytica 160: 15-24.

MILLER R, RAUSHER M & MANOS P. 1999. Phylogenetic Systematics of *Ipomoea* (Convolvulaceae) Based on ITS and Waxy Sequences. Syst Bot 24: 209-227.

MOLLINARI M, OLUKOLU B, PEREIRA G, KHAN A, GEMENET DC, YENCHO C & ZENG ZB. 2019. Unraveling the Hexaploid Sweet potato Inheritance Using Ultra-Dense Multilocus Mapping. G3: Genes Genom Genet 10(1): 281-292.

MONDEN Y, HARAM T, OKADA K, JOHANA O, KOBAYASHI A, TABUCHI H, ONAGA S & TAHARA M. 2015. Construction of a linkage map based on retrotransposon insertion polymorphisms in sweet potato via high-throughput sequencing. Breed Sci 65: 145-153.

MONTEROS-ALTAMIRANO A, PAREDES D, BUITRÓN-BUSTAMANTE J, TAPIA C & PEÑA G. 2020. Genetic diversity of sweet potatoes [*Ipomoea batatas (L.)* Lam.] in Ecuador. Genet Resour Crop Evol 68: 307-320.

MWANGA ROM, KRIEGNER A, CERVANTES-FLORES JC, ZHANG DP, MOYER JW & YENCHO CG. 2002. Resistance to sweet potato chlorotic stunt virus and sweet potato feathery mottle virus is mediated by two separate recessive genes in sweet potato. J Amer Soc Hort Sci 127: 798-806.

NISHIYAMA I, NIYAZAKI T & SAKAMOTO S. 1975. Evolutionary autopolyploidy in sweet potato [*Ipomoea batatas (L.*) Lam] and its progenitors. Euphytica 24: 197-208.

ORACION MZ, NIWA K & SHIOTANI I. 1990. Cytological analysis of tetraploid hybrids between sweet potato and diploid *Ipomoea trifida* (H.B.K.) Don. Theor Appl Genet 80: 617-624.

PEIYONG M, XIAOFENG B, ZHAODONG J, XIAODING G & YIZHI X. 2016. De novo sequencing and comprehensive analysis of the mutant transcriptome from purple sweet potato (*Ipomoea batatas L.*). Gene 575(2): 641-649.

PETRONI K & TONELLI C. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci 181: 219-229.

PFLIEGER S, LEFEBVRE V & CAUSSE M. 2001. The candidate gene approach in plant genetics: a review. Mol Breed 7: 275-291.

PHA SNEATH & SOKAL RR. 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco, p. 230-234.

QIN Z, HOU F, LI A, DONG S, HUABF C, WANG Q & ZHANG L. 2020. Comparative analysis of full-length transcriptomes based on hybrid population reveals regulatory mechanisms of anthocyanin biosynthesis in sweet potato (*Ipomoea batatas (L.*) Lam). BMC Plant Biol 20: 299.

RAJAPAKSE S, NILMALGODA SD, MOLNAR M, BALLARD RE, AUSTIN DF & BOHAC JR. 2004. Phylogenetic relationships of the sweet potato in Ipomoea series Batatas (Convolvulaceae) based on nuclear β-amylase gene sequences. Mol Phylogenet Evol 30(3): 623-632.

REEVES PA, PANELLA LW & RICHARDS CM. 2012. Retention of agronomically important variation in germplasm core collections: implications for allele mining. Theor Appl Genet 124: 1155-1171.

RIPOL MI, CHURCHILL GA, DA SILVA JAG & SORRELLS M. 1999. Statistical aspects of genetic mapping in autopolyploids. Gene 235: 31-41.

RODZEN JA, FAMULA TR & MAY B. 2004. Estimation of parentage and relatedness in the polyploidy white sturgeon (*Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci. Aquaculture 232: 165-182.

ROHLF FJ. 1997. NTSYSpc. Numerical Taxonomy and Multivariate Analysis System. Vol. 2.0. Exeter Software, Setauket, New York, p. 33.

SAITO K, YONEKURA-SAKAKIBARA K, NAKABAYASHI R, HIGASHI Y, YAMAZAKI M, TOHGE T & FERNIE AR. 2013. The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. Plant Physiol Biochem 72: 21-34.

SASAI R ET AL. 2019. Development of molecular markers associated with resistance to Meloidogyne incognita by performing quantitative trait locus analysis and genome-wide association study in sweet potato. DNA Res 26(5): 399-409. doi:10.1093/dnares/dsz018.

SHIOTANI I. 1987. Genomic structure and the gene flow in sweetpotato and related species. In: Gregory P (Ed), Exploration, maintenance and utilization of sweetpotato genetic resources. Rep 1st sweetpotato Planning Conf. CIP, Lima, Peru, p. 61-73.

SHIOTANI I & KAWASE T. 1989. Genomic Structure of the sweetpotato and Hexaploids in *Ipomoea trifida* (H.B.K.) DON. Japan J Breed 39(1): 57-66.

SHIRASAWA K ET AL. 2017. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). Sci Rep 7: 44207.

SHIRLEY BW. 1998. Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. Seed Sci Res 8: 415-422.

SHI Z, BASSA IA, GABRIEL SL & FRANCIS FJ. 1992. Anthocyanin pigments of sweet potatoes - *Ipomoea batatas*. J Food Sci 57(3): 755-758.

SUDA I, OKI T, MASUDA M, KOBAYASHI M, NISHIBA Y & FURUTA S. 2003. Physiological functionality of purple-fleshed sweet potatoes containing anthocyanins and their utilization in foods. JARQ 37: 167-173.

SYBENGA A. 1975. The quantitative analysis of chromosome pairing and chiasma formation based on the relative frequencies of MI configurations. VII. Autotetraploids. Chromosoma 50: 211-222.

TANAKA M, NAKATANI M, NAKAZAWA Y & TAKAHATA Y. 2004. Structural characterization of the dihydroflavonol 4-reductase B (DFR-B) gene in the sweetpotato. DNA Seq 15(4): 277-282.

TANAKA M, TAKAHATA Y, KURATA R, NAKAYAMA H & YOSHINAGA M. 2012. Structural and functional characterization of IbMYB1 genes in recent Japanese purple-fleshed sweet potato cultivars. Mol Breed 29: 565-574.

THOMPSON JD, HIGGINS DG & GIBSON TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

UKOSKIT K & THOMPSON PG. 1997. Autopolyploidy versus allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweet potato. J Amer Soc Hort Sci 122: 822-828.

VARSHNEY RK, GRANER A & SORRELLS ME. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1): 48-55.

WANG X, ZHAO X, ZHU J & WU W. 2005. Genome-wide Investigation of Intron Length Polymorphisms and Their Potential as Molecular Markers in Rice (*Oryza sativa L.*). DNA Res 12: 417-427.

WANG Z, LI J, LUO Z, HUANG L, CHEN X, FANG B, CHEN J & ZHANG X. 2011. Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). BMC Plant Biol 11: 139.

XI C, ZHANG C & WU W. 2011. Investigation and utilization of intron length polymorphisms in conifers. New For 41(3): 379-388.

YANG L, JIN G, ZHAO X, ZHENG Y, XU Z & WU W. 2007. PIP: a database of potential intron polymorphism markers. Bioinformatics 23(16): 2174-2177.

YANG Y ET AL. 2020. Transcript profiling for regulation of sweet potato skin color in Sushu8 and its mutant Zhengshu20. Plant Physiol Biochem 148: 1-9.

YOSHIMOTO M, OKUNO S, YOSHINAGA M, YAMAKAWA O, YAMAGUCHI M & YAMADA J. 1999. Antimutagenicity of sweet potato (*Ipomoea batatas*) roots. Biosci Biotechnol Biochem 63: 537-541.

ZHANG DP, CARBAJULCA D, OJEDA L, ROSSEL G, MILLA S, HERRERA C & GHISLAIN M. 2000. Microsatellite Analysis of Genetic Diversity in Sweet potato Varieties from Latin America. CIP Program Report 1999-2000, p. 295-301.

ZHANG H ET AL. 2019. Identification and functional analysis of three new anthocyanin R2R3-MYB genes in Petunia. Plant Direct 3(1): e00114.

ZHAO N, YU X, JIE Q, LI H, LI H, HU J, ZHAI H, HE S & LIU Q. 2013. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweet potato. Mol Breed 32: 807-820.

ZUFALL RA & RAUSHER MD. 2003. The genetic basis of a flower color polymorphism in the common morning glory (*Ipomoea purpurea*). J Hered 94(6): 442-448.

Abbreviations

ANS: Anthocyanidin synthase
BLAST: Basic Local Alignment Search Tool
bp: base pair
CG: Candidate Genes
CHI: Chalcone Isomerase
CHS: Chalcone sintetase
DFR: Dihydroflavonol 4-reductase
DFR-B: Dihydroflavonol 4-reductase
dMEL: deleted Mobile Element
DNA: Deoxyribonucleic Acid
EPIC: Exon-Primed Intron-Crossing
EST: Expressed Sequence Tag
F3H: Flavanone 3-hydroxylase

GDR: Genome Database for Rosaceae
HG: Homologous Group
ILP: Intron Length Polymorphism
InDel: Insertion and deletion
INTA: Instituto Nacional de Tecnología Agropecuaria
ISNP: Intron Single Nucleotide Polymorphisms
ITS: Internal transcribed spacer
PCR: Polymerase Chain Reaction
SNP: Single Nucleotide Polymorphism
SSR: Simple Sequence Repeat
STS: Sequence-Tagged Sites
UF3GT: UDP-glucose flavonoid 3-oxy-glucosyltransferase
UPGMA: Unweighted Pair Group Method using Arithmetic averages

How to cite
ARIZIO CM, COSTA-TÁRTARA SM, ZUNINO IM & MANIFESTO MM. 2022. Development of anthocyanin markers: gene mapping, genomic analysis and genetic diversity studies in *Ipomoea* species. An Acad Bras Cienc 94: e20210672. DOI 10.1590/0001-3765202220210672.

Manuscript received on April 28, 2021;
accepted for publication on October 4, 2021

CARLA M. ARIZIO[1,2]
https://orcid.org/0000-0003-3703-0329

SABRINA M. COSTA-TÁRTARA[3]
https://orcid.org/0000-0003-2859-2266

IGNACIO M. ZUNINO[4]
https://orcid.org/000-0002-3591-2427

MARÍA M. MANIFESTO[1]
https://orcid.org/0000-0001-7333-0826

[1]Instituto de Recursos Biológicos, Instituto Nacional de Tecnología Agropecuaria, Las Cabañas y Los Reseros s/n, Hurlingham (1686), Buenos Aires, Argentina

[2]Universidad Nacional de Luján, Departamento de Ciencias Básicas, Ruta 5 y 7 s/n, Luján (6700), Buenos Aires, Argentina

[3]Universidad Nacional de Luján, Departamento de Tecnología, Ruta 5 y 7 s/n, Luján (6700), Buenos Aires, Argentina

[4]Instituto Nacional de Tecnología Agropecuaria, Agencia de Extensión Rural Mercedes, Terminal de Omnibus, Mercedes (6600), Buenos Aires, Argentina

Correspondence to: **Carla Marcela Arizio**
E-mail: arizio.carla@inta.gob.ar, carlaarizio@yahoo.com.ar

Author contributions

CMA planned the research, designed PCR primers, analysis of raw data and drafted the manuscript. MMM helped to plan the experimental design and manuscript preparation. SMCT participated in manuscript preparation and revision and IMZ participated in manuscript revision. All authors read and approved the final manuscript.