Anais da Academia Brasileira de Ciências | Annals of the Brazilian Academy of Sciences
Printed ISSN 0001-3765 | 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 phototoxicative 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 a2 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 a2 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.
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>
<thead>
<tr>
<th>INTA Code (CIP Code)</th>
<th>Root colour (Skin/flesh)</th>
<th>INTA Code (CIP Code)</th>
<th>Root colour (Skin/flesh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 (CIP3338)</td>
<td>Purple/yellow cream</td>
<td>22 218 (DLP4727)</td>
<td>Cream/white</td>
</tr>
<tr>
<td>2 4</td>
<td>Purple/pale yellow</td>
<td>23 225</td>
<td>Cream/cream</td>
</tr>
<tr>
<td>3 7 (CIP31)</td>
<td>Cream/cream</td>
<td>24 227 (DLP4709)</td>
<td>White/white</td>
</tr>
<tr>
<td>4 15</td>
<td>Purple/dark cream</td>
<td>25 243 (DLP4879)</td>
<td>White/cream</td>
</tr>
<tr>
<td>5 20</td>
<td>Cream/pale yellow</td>
<td>26 246 (DLP4793)</td>
<td>Cream/dark cream</td>
</tr>
<tr>
<td>6 30 (CIP17)</td>
<td>Cream/cream</td>
<td>27 264 (DLP4803)</td>
<td>Orange/pale orange</td>
</tr>
<tr>
<td>7 31 (CIP16)</td>
<td>Cream/cream</td>
<td>28 281</td>
<td>Pink/cream</td>
</tr>
<tr>
<td>8 45</td>
<td>Purple/yellow with orange spots</td>
<td>29 324 (DLP4883)</td>
<td>Purple red/cream</td>
</tr>
<tr>
<td>9 59 (DLP3983)</td>
<td>White/white</td>
<td>30 358 (CIP9)</td>
<td>Purple/yellow with orange spots</td>
</tr>
<tr>
<td>10 83 (DLP4013)</td>
<td>Purple red/orange</td>
<td>31 6</td>
<td>Purple red/dark yellow</td>
</tr>
<tr>
<td>11 85 (DLP4017)</td>
<td>Red/yellow cream</td>
<td>32 18</td>
<td>Purple red/dark orange</td>
</tr>
<tr>
<td>12 93 (DLP4025)</td>
<td>Purple red/purple</td>
<td>33 35</td>
<td>Cooper/intermediate orange</td>
</tr>
<tr>
<td>13 100 (DLP4042)</td>
<td>Cream/cream</td>
<td>34 48</td>
<td>Cooper/cream</td>
</tr>
<tr>
<td>14 108 (DLP4053)</td>
<td>White/cream</td>
<td>35 306 (Beauregard)</td>
<td>Cooper/intermediate orange</td>
</tr>
<tr>
<td>15 127 (DLP4080)</td>
<td>White/cream</td>
<td>36 336</td>
<td>Red/pale yellow</td>
</tr>
<tr>
<td>16 129 (DLP4083)</td>
<td>Cream/white</td>
<td>37 338</td>
<td>Red/ dark yellow</td>
</tr>
<tr>
<td>17 134 (DLP4092)</td>
<td>White/cream</td>
<td>38 339</td>
<td>Red/ dark orange</td>
</tr>
<tr>
<td>18 148 (DLP4115)</td>
<td>Purple red/cream</td>
<td>39 340</td>
<td>Red/cream</td>
</tr>
<tr>
<td>19 166 (DLP4144)</td>
<td>Cream/yellow with orange spots</td>
<td>40 341</td>
<td>Dark Purple/white</td>
</tr>
<tr>
<td>20 171</td>
<td>Purple red/cream with purple spots</td>
<td>41 353 (USDA ID531154)</td>
<td>No data</td>
</tr>
<tr>
<td>21 178 (DLP4059)</td>
<td>Purple red/purple</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 primersto 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
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.
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⁻¹²⁹). 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).

**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...
(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. aristolochiifolia.

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. aristolochiifolia 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. aristolochiifolia, which presented better homology with I.
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 was the most polymorphic markers with 6 and 7 bands, respectively. Fifty bands (92.6%) were polymorphic among the 41 sweet potato accessions of *Ipomoea* species.
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*. aristolochiaeefolia). 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*. aristolochiaeefolia, two loci do not amplify in *I*. leucantha, three loci for *I*. amnicola and four loci for *I*. parasitica.
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. Leucanthera*, *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
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
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.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Number of alleles</th>
<th>Observed frequency</th>
<th>Allele type</th>
<th>Segregation ratio</th>
<th>$\chi^2$</th>
<th>Potential allelic dosage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ib-DFR2</td>
<td>1</td>
<td>51/61</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>0.8928</td>
<td>Aaaaaa x aaaaa</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>50/63</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>1.4956</td>
<td>Aaaaaa x aaaaa</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>83/30</td>
<td>Biparental</td>
<td>3:1</td>
<td>0.1450</td>
<td>Aaaaaa x Aaaaaa</td>
</tr>
<tr>
<td>Ib-F3H1</td>
<td>1</td>
<td>99/14</td>
<td>Polymorphic</td>
<td>4:1</td>
<td>4.0907</td>
<td>AAAAAA x AAAAAA</td>
</tr>
<tr>
<td>Ib-F3H3</td>
<td>1</td>
<td>52/54</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>0.0377</td>
<td>Aaaaaa x aaaaaa</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56/50</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>0.3396</td>
<td>Aaaaaa x aaaaaa</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>99/7</td>
<td>Biparental</td>
<td>14:1</td>
<td>0.0000</td>
<td>AAAAAA x AAAAAA</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>98/6</td>
<td>Biparental</td>
<td>14:1</td>
<td>0.13</td>
<td>AAAAAA x AAAAAA</td>
</tr>
<tr>
<td>Ib-CHI3</td>
<td>1</td>
<td>51/51</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>0.0000</td>
<td>Aaaaaa x aaaaaa</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>91/10</td>
<td>Biparental</td>
<td>9:1</td>
<td>0.004</td>
<td>AAAAAA x AAAAAA/AAAAAA x AAAAAA</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>51/51</td>
<td>Biparental</td>
<td>3:1</td>
<td></td>
<td>------ x aaaaaa (distorted)</td>
</tr>
<tr>
<td>Ib-CHI5</td>
<td>1</td>
<td>46/55</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>0.8019</td>
<td>Aaaaaa x aaaaaa</td>
</tr>
<tr>
<td>Ib-UF3GT</td>
<td>1</td>
<td>78/31</td>
<td>Biparental</td>
<td>3:1</td>
<td>0.688</td>
<td>Aaaaaa x Aaaaaa</td>
</tr>
<tr>
<td>Ib-ANS3</td>
<td>1</td>
<td>62/59</td>
<td>Polymorphic</td>
<td>1:1</td>
<td>1.522</td>
<td>Aaaaaa x aaaaaa</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>108/3</td>
<td>Biparental</td>
<td>14:1</td>
<td>2.80</td>
<td>AAAAAA x AAAAAA</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>84/27</td>
<td>Biparental</td>
<td>3:1</td>
<td>0.027</td>
<td>Aaaaaa x Aaaaaa</td>
</tr>
</tbody>
</table>

*Presence/absence of bands. p-value (NS: not significant; *: significant).
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 autoployploid 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 x 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-testcross 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-testcross 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, Krieger 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 (Rajapakte et al. 2004). Carotenoid biosynthesis genes have been used to analyze evidence of geographical subdivision and linkage disequilibrium in carrot (Cloutault 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**


SYBENG 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.


**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
How to cite

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

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.