Title

Functional polymorphism in lycopene beta-cyclase gene as a molecular marker to predict bixin production in Bixa orellana L. (achiote)

Journal

Molecular Breeding, 36(9)

ISSN

1380-3743

Authors

Trujillo-Hdz, José Abraham
Cárdenas-Conejo, Yair
Turriza, Pedro Escobar
et al.

Publication Date

2016-09-01

DOI

10.1007/s11032-016-0555-y

Peer reviewed

Mol Breeding (2016) 36:135
DOI 10.1007/s11032-016-0555-y

Functional polymorphism in lycopene beta-cyclase gene as a molecular marker to predict bixin production in *Bixa orellana* L. (achiote)

José Abraham Trujillo-Hdz · Yair Cárdenas-Conejo · Pedro Escobar Turriza · Margarita Aguilar-Espinosa · Victor Carballo-Uicab · Luz Elena Garza-Caligaris · Luca Comai · Renata Rivera-Madrid

Received: 26 February 2016 / Accepted: 2 September 2016 / Published online: 13 September 2016

Abstract Bixin is an apocarotenoid obtained from the seed aril of *Bixa orellana* L., a tropical plant known as achiote in Mexico. This compound is the second most commonly used natural colouring for food and pharmaceutical industries. *B. orellana* is an outcrossing species that displays high genetic variability. Recently, the colour traits of sexual organs were associated with the biosynthesis and accumulation of bixin in mature seeds. Herein, we describe a new approach for genotype–phenotype association by surveying lycopene beta-cyclase *(Boβ-LCY1)* gene variation in sixteen achiote accessions divided into three groups according to contrasting traits, such as flower colour, fruit colour and bixin production. Using a combination of single-strand conformational polymorphism techniques and the sequencing of polymorphic bands, we identified several single-nucleotide polymorphisms that divided the accessions into three haplotypes. Surprisingly, we observed that these three haplotypes were consistent with the same three groups previously characterized by phenotypic traits. We derived a putative sequence for the *Boβ-LCY1* gene and surveyed the variations in this sequence. The heterozygosity of *Boβ-LCY1* alleles resulted in a higher bixin content, likely associated with heterosis for this metabolite. These findings augment the

Electronic supplementary material The online version of this article (doi:10.1007/s11032-016-0555-y) contains supplementary material, which is available to authorized users.

J. A. Trujillo-Hdz · Y. Cárdenas-Conejo · P. E. Turriza · M. Aguilar-Espinosa · V. Carballo-Uicab · L. E. Garza-Caligaris · R. Rivera-Madrid (✉)
Centro de Investigación Científica de Yucatán A.C., Calle 43 No. 130, Col. Chuburná de Hidalgo, 97200 Mérida, Yucatán, Mexico
e-mail: abrahamth@hotmail.com

Y. Cárdenas-Conejo
e-mail: ycardenasco@conacyt.mx

P. E. Turriza
e-mail: pjeturriza@gmail.com

M. Aguilar-Espinosa
e-mail: mgf@cicy.mx

V. Carballo-Uicab
e-mail: victor.carballo@cicy.mx

L. E. Garza-Caligaris
e-mail: luzelenagarcal@hotmail.com

R. Rivera-Madrid
e-mail: renata@cicy.mx

L. Comai
Genome Center, University of California, Davis, CA 95616, USA
e-mail: lcomai@ucdavis.edu

Present Address:
Y. Cárdenas-Conejo
Laboratorio de Bioingeniería, Universidad de Colima, Km. 9 carretera Coquimatlán-Colima, 28400 Coquimatlán, Colima, Mexico

toolbox available for the selection and genetic improvement of *B. orellana* and provide a reliable phenotype–genotype association method for commercial varietal selection, contributing to the development of laboratory techniques to identify desirable traits of commercial plant species.

Keywords Annatto · *Boβ-LCY1* gene · Carotenoids · SNP · SSCP · Heterosis

Introduction

Achiote or *Bixa orellana* L. (locally known as 'achiote') is a perennial tree, native to the Neotropics (Arce 1999) and the only cultivated species in the *Bixaceae* family (Ambrósio Moreira et al. 2015). This species has one of the lowest estimates for the amount DNA among angiosperms, $4C = 0.78$ pg with $2n = 14$ chromosomes (Hanson et al. 2001; De Almeida et al. 2006). *Bixa orellana* L. was domesticated in the Neotropics, although it is not clear where or from which wild populations, and there is evidence that *Bixa urucurana* Willd. is a subspecies of *B. orellana* L. and likely the wild ancestor of cultivated annatto or achiote (Ambrósio Moreira, et al. 2015).

Bixin is a natural colouring apocarotenoid derived from the carotenoid biosynthesis pathway (Jako et al. 2002; Bouvier et al. 2003). This molecule accumulates in the seed aril of *B. orellana* L. and is primarily used as a colouring for dairy products, drinks and in the pharmaceutical industry (Giuliano et al. 2003). Additionally, achiote seeds are the basic ingredients of a condiment used to prepare traditional dishes in Mexico and other parts of the world (Godoy-Hernández 2000; Avendaño-Arrate et al. 2012). Nevertheless, the production and trade of this compound are affected by the heterogeneity of bixin levels accumulated in the mature seeds of different plant varieties, reflecting the high genetic variability of this species (Rivera-Madrid et al. 2006). Given the economic potential of bixin, the early identification of promising plants in breeding programmes would be advantageous. Previous studies aimed at genetic improvement have identified the phenotypic traits of achiote plants associated with differential pigment accumulation in mature seeds, such as the colour of the flower and fruit and pod dehiscence (Rivera-Madrid et al. 2006; Valdez-Ojeda et al. 2010; Rodríguez-Ávila et al. 2011). However, genetic selection based on sexual traits is difficult because the first flowering occurs in the third or fourth year after seed germination, followed by the annual flowering. Thus, a molecular marker that could be used in young individuals to detect more productive plants would be an important tool to improve field yields.

Carotenoids are metabolites widely distributed in plants, contributing to the flavour, colour and/or aroma of the flowers and fruits (Moise et al. 2014; Nisar et al. 2015). In both plants and algae, the carotenoid pathway involves the cyclization of lycopene. Both α- and β-cyclase are needed for the biosynthesis of α-carotene, which contains an ε- (produced by α-cyclase) and a β-ionone ring (produced by β-cyclase), while β-carotene is produced only by the enzyme lycopene-β-cyclase (β-LCY) (Cunningham et al. 1996; Cunningham and Gantt 1998; Yuan et al. 2015). Carotenoids can also be cleaved at any of their conjugated double bonds, resulting in a diverse set of apocarotenoids (Auldridge et al. 2006) that can be modified through dioxygenases, hydroxylases and epoxidases to produce physiologically important *β*-apocarotenoids, hormones such as abscisic acid (ABA) and strigolactones. In addition, apocarotenoid modification can produce pigments such as crocetin and aromatic compounds, such as β-cyclocitral, α- and β-damascenone, geranial, genaryl acetone, β-ionone, which are also commercially valuable products (Pfander and Schurtenberger 1982; Bouvier et al. 2003; Frusciante et al. 2014; Nisar et al. 2015; Yuan et al. 2015).

Bixa orellana produces bixin, an orange-red apocarotenoid that accumulates in high concentrations and accounts for 80 % of the total carotenoids present in seeds (Rivera-Madrid et al. 2006). Using a heterologous system, previous studies have shown that lycopene is transformed via a series of reactions, involving a CCD4 dioxygenase enzyme (BoLCD), an aldehyde dehydrogenase (BoBADH) and a methyltransferase (BoBMT), into bixin and bixin dimethyl ester (Bouvier et al. 2003).

The analysis of *B. orellana* gene transcripts encoding carotenoid biosynthesis enzymes revealed that *PDS*, *β-LCY* and *ε-LCY* genes are highly regulated at different stages of plant development and in plant accessions with different flower and fruit colours (Rodríguez-Ávila et al. 2011). Thus, in two accessions

of achiote differing in pigment content, *Boβ-LCY1* mRNA expression was correlated with carotene accumulation (Rodríguez-Ávila et al. 2011). Additionally, in some plants, such as capsicum, watermelon, papaya and maize, DNA sequence polymorphisms in cyclase genes are associated with the pigment phenotype (Bang et al. 2007; Ha et al. 2007; Harjes et al. 2008; Devitt et al. 2010). Based on these observations, the aim of the present study was to investigate the existence of polymorphisms in the *Boβ-LCY1* gene in different achiote accessions and test a potential correlation to bixin levels. Here, we identify molecular markers using single-strand conformation polymorphism (SSCP) and DNA sequencing, showing that polymorphism at *Boβ-LCY1* is associated with colour and bixin yield.

Materials and methods

Achiote plant accessions

Sixteen *B. orellana* individuals from the germplasm bank located in CRUPY (Centro Regional Universitario de la Peninsula de Yucatan, Mexico) were chosen for this study based on the associations observed previously between the plant colour traits and bixin accumulation in mature seeds (Rivera-Madrid et al. 2006; Valdez-Ojeda et al. 2010; Rodríguez-Ávila et al. 2011). These plants were originally collected from Yaxcaba, Cooperativa and Chicxulub Pueblo (the achiote plantation of the 'La Extra' Company), from the cultivation region of Yucatan State located at latitude 20°13′ to 20°29′ and longitude 88°55′ to 89°54′. Each of these *B. orellana* plants could result from selfing or from crossing (Valdez-Ojeda et al. 2010; Lombello and Pinto-Maglio 2014). They were genetically and morphologically characterized. Genetic analysis based on SNPs demonstrated their heterogeneity (Valdez-Ojeda et al. 2008). The plants were divided into three groups according to bixin content and the following contrasting morphological characteristics: group 1 showed white flowers, green and indehiscent fruit, and comprised five individuals, herein referred to as P12, P13, P19, N20 and P37; Seven plants (N1, N3, N4, N6, N9, BC and Co32) formed group 2 and showed pink flowers and green dehiscent fruit with red spines; and group 3 comprised accessions N5, N8p, C1 and C6, showing purple flowers and red or yellow dehiscent pods (Fig. 1). The young leaves were harvested from all accession plants and immediately stored in liquid nitrogen, transported to the laboratory and stored at −80 °C for subsequent analyses.

Determination of the bixin content

The bixin contents in mature seeds called stage 3 (S3) were spectrophotometrically quantified. This protocol was specifically developed for *B. orellana* seeds (Mckeown and Mark 1962) using a formula considering the extinction coefficient of chloroform. Pigment was extracted from the dried seeds of three biological replicates using chloroform (5 ml $CHCl_3$ per 100 mg) for 30 min, and each sample was measured in triplicate. The bixin content was measured at 404 and 500 nm using a spectrophotometer (DU6-650 Beckman) and calculated using the formula according to McKeown and Mark (1962).

An analysis of variance was carried out on the data obtained using Statgraphics and complete randomization. Duncan multiple comparison tests ($n = 3$; $p \leq 0.5$) were performed on the variables. Values labelled with the same letter are not significantly different.

Isolation and cloning of the lycopene beta-cyclase *Boβ-LCY1* gene

To isolate and verify the *Boβ-LCY1* cDNA sequence (KT359004) in annatto identified by Cárdenas-Conejo et al. (2015), we designed specific primers (5′ GAAC CACCATACGAGCTTTACC 3′ and 5′ CCCTTGGC CATAATCTCTATCC 3′). The predicted amplicon size based on oligonucleotides is 1770 bp and includes most of the open reading frame (ORF) of *Boβ-LCY1* and the 5′UTR. Total RNA was extracted from leaves using a PureLink™ Micro-to-Midi™ (Invitrogen,USA) following Rodríguez-Ávila et al. (2009). The cDNA was synthesized using the SuperScript III First-Strand Synthesis System for the RT-PCR kit (Invitrogen, San Diego, CA) according to the manufacturer's instructions. DNA was extracted from leaves using the DNeasy® Plant Mini Kit, according to the manufacturer's instructions (QIAGEN, Netherlands). We amplified the cDNA and genomic DNA of annatto leaves using the same primers.

The PCR reactions were performed in a total volume of 15 µl, containing 0.6 units of FastStart Taq DNA,

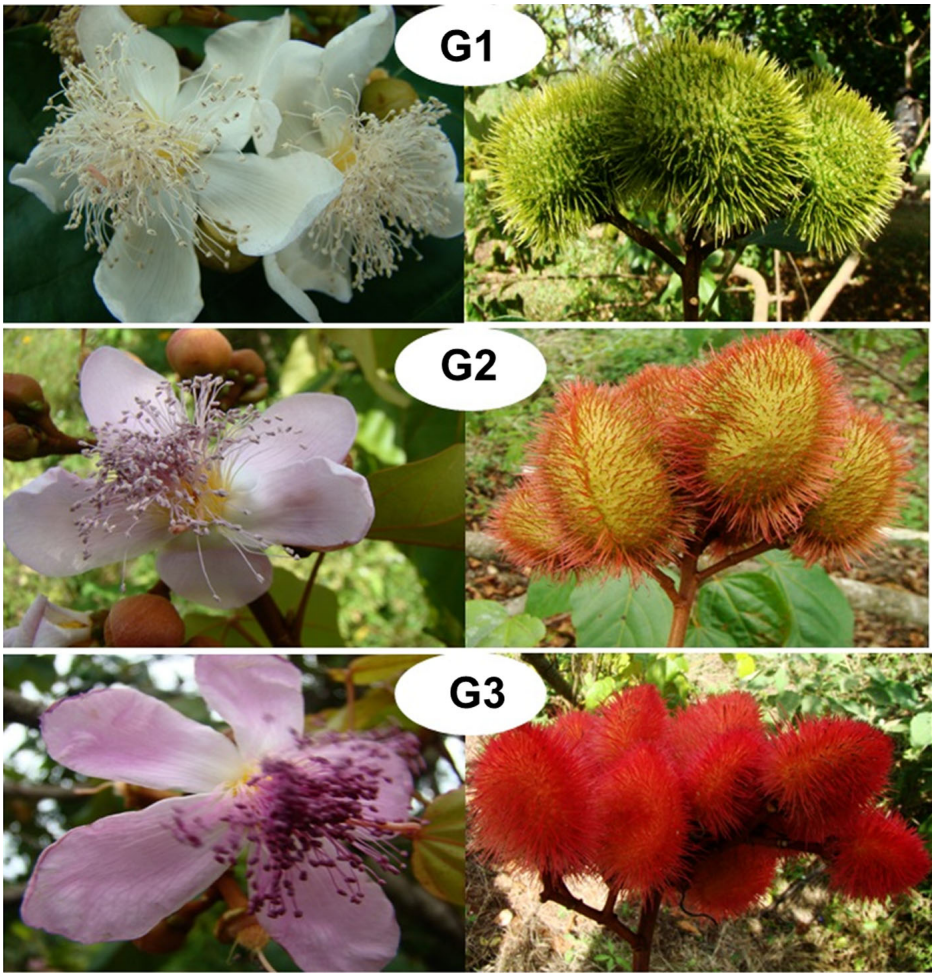


Fig. 1 Representative achiote phenotypes of the plants analysed in the present study. Group 1, **G1** green fruit with white flowers; Group 2, **G2**, green fruit with red spines and pink flowers; Group 3, **G3**, red fruit with purple flowers. (Color figure online)

0.6 mM 10× buffer with $MgCl_2$, 0.3× GC-RICH solution (F. Hoffmann-La Roche Ltd, Switzerland), 200 µM dNTPs (Invitrogen, Carlsbad, CA, USA), 1 µM of each primer (Sigma-Aldrich, USA) and 10 ng of cDNA or DNA, using the following programme: 1 cycle at 94 °C for 4 min, 40 cycles at 90 °C for 30 s, 62 °C for 30 s and 72 °C for 1.5 min and a final extension at 72 °C for 7 min. The purified PCR products were cloned into the pGEM-T Easy vector (Promega Corporation, Madison, USA) and sequenced at Clemson University Genomics Institute (CUGI, USA).

Phylogenetic and sequence analysis of *Boβ-LCY1*

A phylogenetic reconstruction was based on the alignment of 28 sequences of β-LCY protein from 16 plant species (Table S1). The analysis was performed using algorithms included in MEGA5 (Tamura et al. 2011). The phylogenetic tree was inferred using the maximum-likelihood method based on the Jones–Taylor–Thornton (JTT) substitution model (Jones et al. 1992) and gamma distributed (G). The substitution model was predicted using the best-fit substitution model (ML) function included in MEGA5. A phylogeny test was conducted using the bootstrap method (1000 replicates). All positions containing gaps and missing data were eliminated. The alignment was performed using a ClustalW algorithm with default parameters. The phylogenetic tree was rooted with β-LCY protein from *Chlamydomonas reinhardtii* a single-cell green alga.

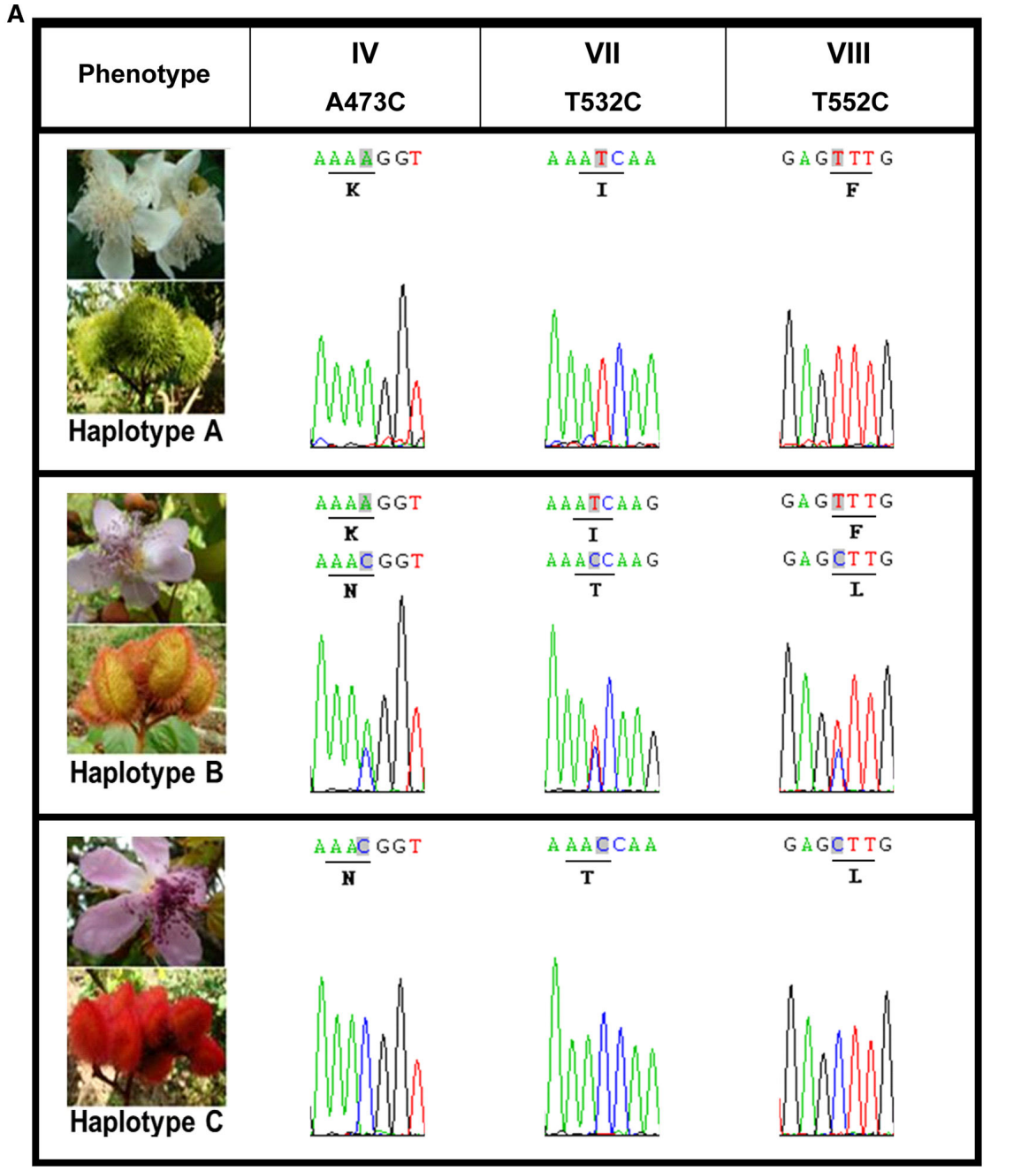
SNP detection using the SSCP method in *Bixa orellana* L.

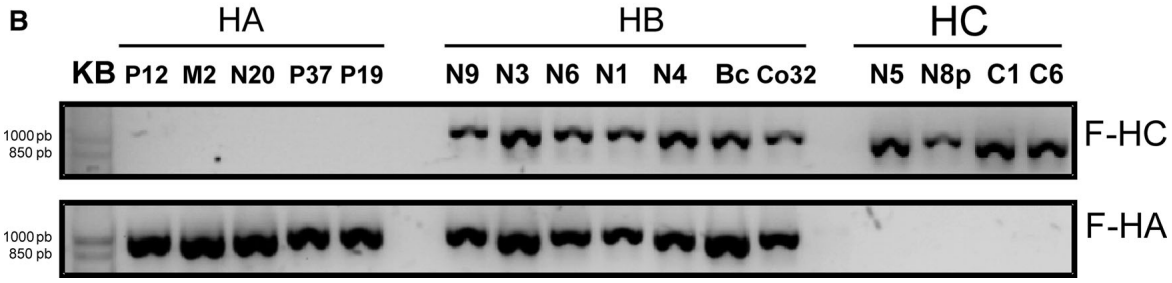
To detect SNPs along the *Boβ-LCY1* gene, the SSCP technique was used, and the resulting polymorphic bands were sequenced (Orita et al. 1989; Hayashi 1991). This is an effective method for the easy, rapid and economically feasible detection of DNA mutations or nucleotide changes. Changes in the sequences were detected based on differences in the migration of single strands of DNA in non-denaturing acrylamide gels and were verified through sequencing only the bands with differential profiles (Orita et al. 1989; Hayashi 1991).

Eight specific primer pairs (B2, B3, B4, B5, B6, B7, B8 and B9), which only amplify *Boβ-LCY1* under the established conditions were designed based on the sequence obtained from the transcriptome (Acc. No. KT359004) using Primer3 (Table S2). Fragments of approximately 200 bp along the length of the *Boβ-LCY1* gene were amplified. The genomic DNA of *Bixa orellana* L. for SNP detection was isolated from the selected accession plants using the DNeasy® Plant Mini Kit, according to the manufacturer's instructions (QIAGEN, Netherlands). Using 10 ng of genomic DNA from each sample as template, eight pairs of primers were used for amplification (Table S2). Each reaction comprised a final volume of 15 µl, with 0.6 units of FastStart Taq DNA, 0.6 mM 10x buffer with $MgCl_2$, 0.3× GC-RICH solution (F. Hoffmann-La Roche Ltd, Switzerland), 200 µM dNTPs (Invitrogen, Carlsbad, CA, USA), 1 µM of forward/reverse oligonucleotides (Sigma-Aldrich, USA) and the PCR programme included 1 cycle at 94 °C for 4 min, followed by 35 cycles at 90 °C for 30 s, 62 °C for 30 s and 72 °C for 50 s, with a final extension at 72 °C for 7 min. For the SSCP analysis (Orita et al. 1989), 3 µl of PCR product was used with 3 µl of LoadDye (0.05 % bromophenol blue, 0.05 % xylene cyanol, 24 % formamide and 20 mM EDTA). The samples were heated to 94 °C for 4 min for denaturing and immediately incubated on ice for 5 min. The samples were subsequently separated on acrylamide gels (Model S2, GibcoBRL), using 0.5X Acrylamide Solution for Mutation Detection (Sigma-Aldrich, USA). The electrophoresis conditions were 200 V, 15 mA and 5 W power for 20 h at 22 °C, using 1X TBE electrophoresis buffer. After electrophoresis, the gel was silver stained to visualize the bands according to the protocol of Bassam et al. (1991). The gels were subsequently scanned using a ScanJet 6300C (Hewlett–Packard, USA) to store the image for analysis. Polymorphic bands were cut from the gel using a scalpel and placed separately in a 1.5-ml Eppendorf tube, containing 50 µl of elution solution (0.5 M ammonium acetate, 10 mM magnesium acetate tetrahydrate, 1 mM EDTA (pH 8.0) and 0.1 % (w/v) SDS (Maxam and Gilbert 1977; Sambrook et al. 1989). The reactions were subsequently incubated for 12 h at 37 °C. One-third of the total volume of DNA solution from each band was used as a template for PCR reamplification using the primers mentioned above. The reaction resulting from this PCR was purified with the High Pure PCR Product Purification Kit (F. Hoffmann-La Roche Ltd, Switzerland), according to the manufacturer's instructions, and submitted for sequencing to Clemson University Genomics Institute (CUGI, USA).

Primer design for identification of specific alleles

To develop an easy and convenient assay for genotype–phenotype association, two suitable forward primers were designed for PCR amplification of the allele of interest in the *Boβ-LCY1* gene in achiote. The reverse primer B7 was used in both cases for amplification. These forward primers were named F-HA (Forward haplotype A) and F-HC (Forward haplotype C). The F-HA primer (5′-CCAATTCTAG-CACTCTCACG-3′) is located between nucleotide coordinates 496–515 of the Boβ-LCY1 cDNA sequence (KT359004). The last nucleotide in this primer (G) is positioned in the SNP G515T, this nucleotide (G) confer specificity to *Boβ-LCY1* gene from haplotype A plants. On the other hand, the F-HC primer (5′-CAAGAAGGAGAATCTTGAGC-3′) is located between nucleotide coordinates 533–552 of the *Boβ-LCY1* cDNA sequence (KT359004). In this case, the last nucleotide (C) is positioned in the SNP T552C, and this last nucleotide in F-HC primer (C) confers specificity to *Boβ-LCY1* gene from haplotype C plants. Therefore, this two-primer system can be used to detect the homozygotes plants from haplotype A (HA), haplotype C (HC) and the heterozygotes plants from haplotype B (HB). Each reaction comprised a final volume of 15 µl, containing 0.6 units FastStart Taq DNA, 0.6 mM 10× buffer with $MgCl_2$, 0.3× GC-RICH solution (F. Hoffmann-La


A
Phenotype
IV
A473C
VII
T532C
VIII
T552C
AAAAGGT
K
AAATCAA
I
GAGTTTG
F
Haplotype A
AAAAGGT
K
AAACGGT
N
AAATCAAG
I
AAACCAAG
T
GAGTTTG
F
GAGCTTG
L
Haplotype B
AAACGGT
N
AAACCAA
T
GAGCTTG
L
Haplotype C


B
HA
HB
HC
KB P12 M2 N20 P37 P19
N9 N3 N6 N1 N4 Bc Co32
N5 N8p C1 C6
1000 pb
850 pb
F-HC
1000 pb
850 pb
F-HA

◀**Fig. 2** Results of the SSCP banding profile using B3 primer in sixteen annatto accessions. **a** Structural diagram of the *Boβ-LCY1* gene of annatto. The *letters* inside the *boxes* correspond to the conserved motifs. The position and length of the primers are shown as *white rectangles*. **b** SSCP results for the banding profile for primer B3. The *lines* above the gels show the phenotypic groups correlated with each pattern. *HA* haplotype A, *HB* haplotype B, *HC* haplotype C

Roche Ltd, Switzerland), 200 μM dNTPs (Invitrogen, Carlsbad, CA, USA) and 1 μM of forward/reverse oligonucleotides (Sigma-Aldrich, USA). The PCR programme comprised 1 cycle at 95 °C for 5 min, followed by 28 cycles at 95 °C for 30 s, 62 °C for 30 s and 72 °C for 50 s and a final extension at 72 °C for 7 min.

The bands resulting from this amplification were sequenced to confirm the correct amplification (Fig. S4). In order to test the effectiveness of these primers, they were tested in eleven additional accessions from the tree groups (Fig. S5).

Expression analysis of *BoLCY1* and *BoLCY2* by Quantitative PCR

Total RNA was extracted from seeds of three different development stages S1 (0–7 days), S2 (7–14 days) and S3 (14–21 days) of achiote plants using the method of Rodríguez-Ávila et al. (2009). The cDNA was synthesized using the SuperScript III First-Strand Synthesis System for the RT-PCR kit (Invitrogen, San Diego, CA) according to the manufacturer's instructions.

Boβ-LCY1 and *Boβ-LCY2* genes expression was analysed using the iCycler IQ real-time PCR detection system (Bio-Rad, Hercules, CA, USA). The PCR mix comprised a final volume of 15 μl, containing 100 ng of cDNA as template, Platinum® SYBR® Green qPCR SuperMix-UDG (Invitrogen, Carlsbad, CA) and 1 μM of forward/reverse specific oligonucleotides form *Boβ-LCY1* and *Boβ-LCY2* (Table S1) (Sigma-Aldrich, USA). The amplification programme included incubation at 50 °C for 2 min and 95 °C for 1 min to activate the Platinum® Taq DNA polymerase, followed by 35 amplification cycles of 30 s at 95 °C, 30 s at 61.3 °C and 30 s at 72 °C, with a final extension of 10 min at 72 °C. The 18S ribosomal RNA was used as an internal control (FW 5′ CGGCTACCACATCCAA GGAA3′ and RW 5′ GCTGGAATTACCGCGGCT 3′; GenBank: AF206868). Each quantitative PCR experiment was conducted in duplicate, and the entire procedure was repeated three times to calculate the standard error for each sample. The real-time PCR results were analysed using the 2-ΔΔCt method (Livak and Schmittgen 2001), with appropriate validation experiments (Real-time PCR. Applications Guide Bio-Rad).

Results

Gene structure and identities with orthologous

The sequence obtained from cDNA amplification was 1770 bp long as predicted from the transcriptome (Cárdenas-Conejo et al. 2015), sharing 99.2 % identity with the cDNA of *Boβ-LCY1* (KT359004) (Fig. S1). The sequence from genomic DNA amplification was 1925 bp (KU640956). The genomic sequence revealed that the 5′UTR gene region contains an intron (155 bp).

The amino acid sequence predicted from the full-length cDNA of *Boβ-LCY1* isolated in this work showed an identity range between 77 and 88 % with β-LCY proteins of other dicotyledonous plants, such as *Jatropha curcas* (88 %), *Populus thrichocarpa* (87 %), *Prunus persica* (87 %), *Vitis vinifera* (86 %), *Citrus sinensis* (86 %), *Citrus maxima* (86 %), *Theobroma cacao* (85 %), *Carica papaya* (85 %) and *Ricinus communis* (85 %). Pairwise comparison between the Boβ-LCY1 protein and the previously isolated β-LCY protein for *B. orellana* (GenBank: AJ549288) showed 79 % identity, indicating that they are different proteins. The comparison of Boβ-LCY1 protein with its orthologue facilitated the identification of conserved regions between the cyclases (Fig. S2), consistent with the reports of Cunningham et al. (1996), Hugueney et al. (1995).

In most dicotyledonous plant species, two genomic copies of the lycopene beta-cyclase gene have been identified (Fig. S3). These two copies are expressed differentially and in different plant organs during development (Devitt et al. 2010; Ronen et al. 2000; Pecker et al. 1996). Consistent with observations in other species, two beta-lycopene cyclase genes (*Boβ-LCY1* and *Boβ-LCY2*) were identified in the *Bixa orellana* transcriptome (Cárdenas-Conejo et al. 2015). The phylogenetic analysis of the inferred amino acid

Table 1 SNPs identified after sequencing the polymorphic SSCP bands detected using primer B3

ACCESSION	Traits	Bixin mg/gDW	SNP IV A473C	SNP V G515T	SNPVI C527A	SNPVII T532C	SNPVIII T552C
Haplotype A							
P12	White flower, green fruit	8.84 ± 0.30a	A	G	C	T	T
P19		10.76 ± 0.03	A	G	C	T	T
P13		08.76 ± 0.79a	A	G	C	T	T
N20		09.23 ± 0.80a	A	G	C	T	T
P37		10.50 ± 0.65a	A	G	C	T	T
Haplotype B							
N9	Pink flower, green/red fruit	16.90 ± 0.90cd	A/C	G/T	C/A	T/C	T/C
N3		18.02 ± 0.76d	A/C	G/T	C/A	T/C	T/C
N6		14.40 ± 0.74b	A/C	G/T	C/A	T/C	T/C
N1		16.55 ± 0.69cd	A/C	G/T	C/A	T/C	T/C
N4		16.04 ± 0.52c	A/C	G/T	C/A	T/C	T/C
BC		20.56 ± 1.00f	A/C	G/T	C/A	T/C	T/C
Co32		19.28 ± 0.33	A/C	G/T	C/A	T/C	T/C
Haplotype C							
N5	Purple, red fruit	09.32 ± 0.47a	C	T	A	C	C
N8p	Purple, yellow fruit	8.52 ± 0.22a	C	T	A	C	C
C1	Purple, red fruit	6.25 ± 0.57	A/C	T	A	C	C
C6	Purple, red fruit	12.67 ± 2.45	A/C	T	A	C	C

sequence of the *Boβ-LCY1* gene revealed that this protein is evolutionarily related to beta-cyclases proteins that are expressed preferentially in photosynthetic tissue, whereas Boβ-LCY2 is related to beta-cyclases enzymes specific for the chromoplast (Figure S3).

Identification of polymorphism in the *Boβ-LCY1* gene using the PCR-SSCP method

Using the SSCP method, we explored allelic variation in the *Boβ-LCY1* gene among 16 accessions of *Bixa orellana* L. Eight primers pairs scanning the *Boβ-LCY1* gene were used to amplify ~200 bp regions from genomic DNA of the accessions (Fig. 2a, Table S1). The SNPs detected were located between nucleotide coordinates 153–354 (untranslated region, primers B2) and in the region from 334 to 574 (beginning of the ORF *Boβ-LCY1* gene sequence, primers B3). The amplicons corresponding to primers B3 displayed four different SSCP patterns among the 16 achiote accessions examined (Fig. 2b). Additionally, the changes detected in amplicon B3 corresponded to a plant-characteristic motif of the *Boβ-LCY1* gene (plant β conserved region) as summarized in Fig. S2 (Hugueney et al. 1995; Cunningham et al. 1996). The amplicons produced using the remaining primers (B4, B5, B6, B7, B8 and B9) did not detect any polymorphisms.

To verify the polymorphism in the band pattern and identify the potential SNPs, the polymorphic bands were isolated, reamplified through PCR and sequenced. The DNA products from two accessions in each group were sequenced. The sequence comparison of genomic DNA fragments from achiote plants revealed the presence of single-nucleotide polymorphisms (SNPs) in the B2 amplicon. We identified three changes (G249C, T278G and A293T), representing substitutions of a single nucleotide of the Boβ-*LCY1* gene. Similarly, the sequencing of the B3 PCR product revealed five polymorphic sites (A473C, G515T, C527A, T532C and T552C). The B3 changes were located in the ORF and confirmed in all accessions (Table 1). SNPs analysis defined three haplotypes, corresponding to the three groups of varieties previously defined according to flower and fruit phenotype, i.e. displaying different colour traits and bixin production. Plants belonging to group 1

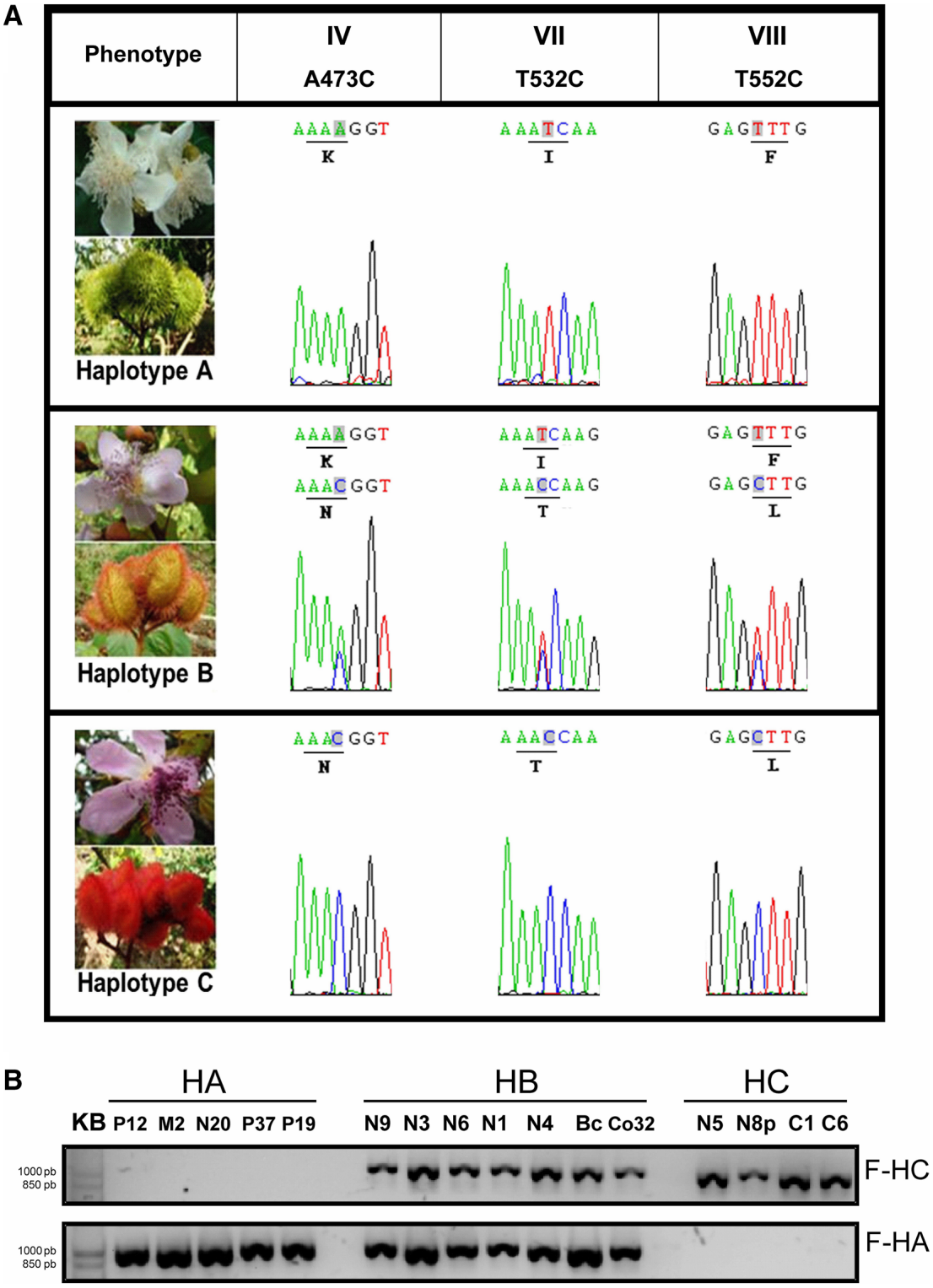


Fig. 3 **a** Amino acid substitutions resulting from the SNPs in the ORF of the gene encoding Boβ-LCY1 in contrasting phenotypes of annatto. The substituted amino acid resulting from the specific DNA change is shown on top of the chromatograms. **b** PCR identification of three haplotypes of annatto plants. *Top of the figure* F-HC primer should be absent in haplotype A (HA); present in haplotype pattern HB and HC. *Bottom of the figure* F-HA should be present in haplotype A (HA) and haplotype pattern HB and absent in haplotype C (HC)

showed no polymorphism compared with the reference sequence (cDNA of *Boβ-LCY1*), while the plants in group 3 showed polymorphisms in the five SNPs. The plants in group 2 displayed both alleles found in group 1 and group 3 (Table 1).

We investigated whether the discovered polymorphism caused any modification in the Boβ-LCY1 protein. The SNPs detected using the B2 primer (G249C, T278G and A293T) are located immediately prior to the start of the ORF of the sequence. The SNPs detected using the B3 primer are located at the beginning of the gene ORF (Fig. 2a). SNP A473C results in an asparagine to lysine substitution at amino acid position 48 (K48 N) (Fig. 3a and Fig. S2). Both T515G and C527A are silent. T532C leads to the substitution of an isoleucine for a threonine at position 68 (I68T) (Fig. 3a; Fig. S2). Moreover, T552C leads to the substitution of a phenylalanine for a leucine at position 75 (F75L) (Fig. 3a). Accessions with pink flowers, green fruit with red spines and increased bixin accumulation in the mature seeds (group 2) displayed these amino acid substitutions (Table 1; Fig. 3a).

Bixin contents in *Bixa orellana* accessions

The total bixin contents in mature seeds (S3 development stage) (Table 1) was closely associated with SNPs pattern, and the plants in groups 1 and 3 contained approximately half the bixin content compared with the plants in group 2. The plants in group 1 had white, green fruit and dehiscent pods and average bixin contents ranging from 8.76 to 10.76 mg/g DW. The plants in group 2 showed bright pink flowers, green fruit with red spines and dehiscent pods, with bixin contents ranging from 16.90 to 20.56 mg/g DW. The plants in the group 3 had 6.25–12.67 mg/g DW bixin content with purple flowers, bright red fruit and dehiscent pots (Table 1). These results demonstrate that DNA polymorphism is associated with reproductive traits and bixin contents, enabling seedling screening for these phenotypic traits.

Haplotype identification

To enable early and efficient detection of the *B. orellana* phenotype, specific primers were designed considering the SNPs in *Boβ-LCY1* (see "Materials and methods"). The PCR reaction facilitated the detection of the three putative genotypes, which were confirmed by sequencing, and determination of zygosity (Fig. 3b). Primer forward F-HA combined with primer reverse B7 amplified a product from plants belonging to haplotype A. Forward primer F-HC combined with reverse primer B7 amplified a product from plants belonging to haplotype C. Only pattern B corresponded to PCR products amplified using both combinations, confirming the presence of three genotypes, each unique to one group of plants (Fig. 3b). The plants from group 2 yielded a product from each set of specific primers, confirming heterozygosity.

The effectiveness of the use of these primers was strengthened after amplifying DNA from eleven additional accessions. We found that the banding pattern showed a perfect match to each of the phenotypic groups (Fig. S5).

Analysis of *Boβ-LCY1* and *Boβ-LCY2* gene expression in *Bixa orellana* seeds

Lycopene cyclases are known to have a major regulatory role in modulating the ratio of carotenoid biosynthesis. After lycopene is produced, carotenoid biosynthesis bifurcates to produce epsilon and beta-carotenoids by the enzymatic activity of two kinds of lycopene cyclases (review in Cazzonelli and Pogson 2010). Since lycopene is a bixin precursor, a third branch also leads to bixin production (Bouvier et al. 2003). Thus, the activity of the cyclase *Boβ-LCY1* may affect lycopene accumulation and therefore bixin production (review in Cazzonelli and Pogson 2010).

In order to evaluate *Boβ-LCY1* expression in the three haplotypes and correlate it with the bixin production, we performed RT-qPCR from RNA extracted from three development stages of the seeds (S1, S2 and S3), corresponding to the developmental window for bixin accumulation. Based on the results, we infer that higher levels of lycopene lead to higher bixin production. We found higher transcript levels of *Boβ-LCY1* gene in the S1 and S2 stages from white flower plants (Group 1). Pink and purple flower plants displayed lower *Boβ-LCY1* gene expression (Fig. 4a). We also analysed *Boβ-LCY2* expression finding an opposite expression pattern where the white flowers accumulated lower transcript levels (Fig. 4b).

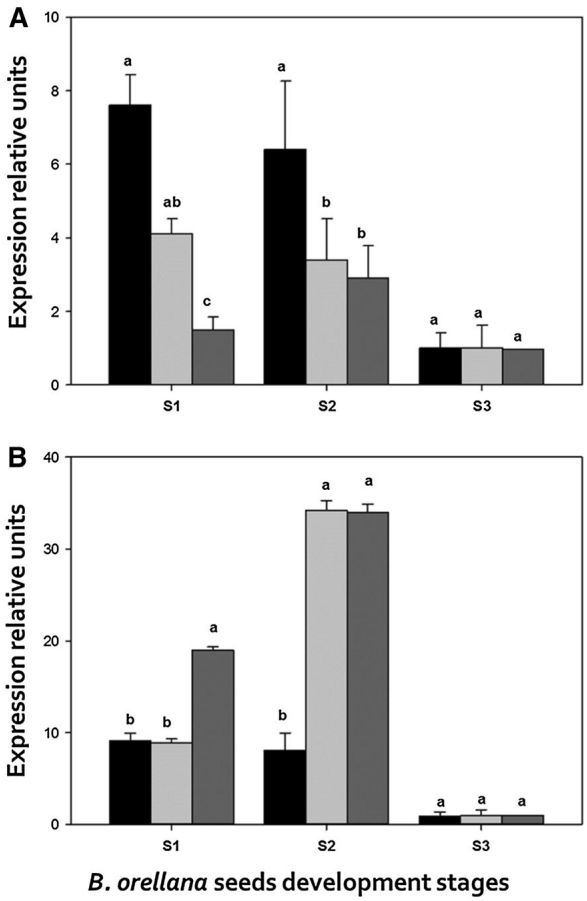


Fig. 4 Expression of *Boβ-LCY* genes in *Bixa orellana* seeds at three different development stages (S1, S2 and S3) from plants of each phenotypic group. **a** Expression of *Boβ-LCY1*. **b** Expression of *Boβ-LCY2*. The *black bars* indicate plant from group 1, the *grey bars* indicate plants from group 2, and the *dark grey* indicates plants from group 3. The transcripts were detected through RT-qPCR using gene-specific primers. An *18S* rRNA was included as a reference control. The data are presented as the mean ± standard deviation (SD). The data were subjected to analysis of variance (ANOVA), and the mean comparison (Tukey test at $p < 0.05$) was performed using OriginPro 8 software. The data in the figure represent the mean ± standard deviation. *Different letters* denote significant differences among treatments ($p \leq 0.05$)

Discussion

Achiote plants are a source of the orange pigment bixin. This species has high genetic and phenotypic variability that affects the production and trade of bixin because varying bixin levels are accumulated by different varieties. Thus, the identification of molecular markers is of applied importance. Previous studies using two achiote variants with contrasting phenotypes suggested an association between *Boβ-LCY1* gene expression and the bixin content in mature seeds and flower and fruit colour (Rivera-Madrid et al. 2006; Valdez-Ojeda et al. 2010; Rodríguez-Ávila et al. 2011). We therefore investigated allelic variation in the *Boβ-LCY1* gene using the SSCP method (Orita et al. 1989), a technique that detects DNA variation in small amplicons, to identify SNPs associated with phenotypic traits. This approach has been successfully used to detect molecular markers in papaya, maize and watermelon (Devitt et al. 2010; Harjes et al. 2008; Bang et al. 2007).

By scanning the gene sequence with multiple primer combinations, we identified polymorphic bands in two *Boβ-LCY1* genic regions. The sequencing of these products identified eight substitutions: G249C, T278G, A293T, A473C, G515T, C527A, T532C and T552C. The first three changes are transversions located 5′ of the coding region. Although these SNPs are not predicted to affect the β-LCY1 protein, the 5′ untranslated region (UTR) is important for the coupling of mRNA to translational machinery and could affect the translation rate. A sequence feature search (Huang et al. 2006; Grillo et al. 2010) identified two potential regulatory elements in the 5′ UTR. The first element is an upstream open reading frame (uORF), spanning nucleotide positions T178 to A260 and the second element is an internal ribosome entry site (IRES) beginning at nucleotide 89 and ending at the start codon. The SNPs localized in the predicted uORF (G249C) and IRES (T278G and A293T) are not predicted to disrupt the function of these putative regulatory elements (data not shown). The remaining substitutions detected using primer pair B3 were located in the ORF. The first three substitutions were transversions, while the latter two substitutions were transitions. The ORF substitutions T515G and C527A are silent, while A473C, T532C and T552C result in the amino acid substitutions K48N, I68T and F75L, respectively. K48N is in a phylogenetically variable region of the β-LCY1 protein. The remaining changes identified in amplicon B3, I68T and F75L are located in the β motif, a characteristic motif identified in the *β-LCY* protein of plants, which might mediate the association of β-LCY with membrane components (Fig. S2) (Hugueney et al. 1995; Cunningham et al. 1996; Bouvier et al. 1997). These changes are tolerated according to the Sorting Intolerant from Tolerant (SIFT) algorithm, which predicts the effect of coding variants on protein function

(Kumar et al. 2009). The predicted proteins were also analysed using the CBS prediction suite, TargetP 1.1 (Emanuelsson et al. 2007), which predicted the localization of both protein forms in the chloroplast. However, the processing site for the signal peptide was different for each form: the processing site for the form associated with the white flowered variety is predicted at amino acid 57, and the processing site for the form associated with the pink flowered variety is predicted at position 65, likely affecting the protein interaction with the chloroplast membrane. This change is also found in *Zea mays* (Fig S2), although we do not know whether both alleles are found in this plant as in *Bixa orellana*. Another interesting prediction was obtained using SUMOplot (Sumoplot, ABGENT). A SUMOylation site predicted at lysine 69 is consistent with the I68T polymorphism. Thus, the protein encoded by the *Boβ-LCY1* allele associated with white flowers has a presumably optimal IKKE recognition sequence for SUMOylation through SUMO-1, while the protein encoded by the *Boβ-LCY1* allele associated with purple flowers alters this site to potentially suboptimal TKKE. Such alterations might result in differential SUMOylation of Boβ-LCY1, thus altering enzyme activity or localization in haplotypes associated with purple flower phenotype (Miura et al. 2009; Park et al. 2011).

Boβ-LCY1 is a key enzyme involved in carotenoid synthesis and it can modulate carotenoid accumulation (Moreno et al. 2013). Thus, Boβ-LCY1 activity regulation could play an important role in either promoting the synthesis of cyclic metabolites, such as ABA (Rodríguez-Ávila et al. 2011), as likely occurs in the plants of group 1 (white flowers), or in promoting lycopene accumulation, thereby enhancing the bixin synthesis pathway (plants from group 2 and 3). However, the correlation between enzyme polymorphism and the differential bixin content is not direct because in this hypothesis more bixin should be expected in group 3. Instead, heterozygous group 2 has the highest concentration, suggesting that another type of regulation might be present since Boβ-LCY1 is a highly regulated enzyme or that the different alleles result in heterosis for bixin production (Birchler et al. 2010; Fiévet et al. 2010).

The polymorphism identified in *Boβ-LCY1* displays a remarkable correspondence to phenotype: varieties in group 1 are characterized by haplotype A and have low bixin contents, green fruit and white flowers; those in group 2 are characterized by haplotype pattern B, with greater bixin quantities, pink flowers and green fruit with red spines; varieties in group 3 display haplotype C and have intermediate bixin content, purple flowers, red fruit (Table 1). The SNP pattern of varieties in group 2, which combines the features of both haplotypes, is consistent with B, reflecting the heterozygosity of haplotypes A and C. The flower and fruit colour phenotype displayed by the B type is approximately semi-dominant, i.e. intermediate between A and C. Interestingly, this is not the case for bixin production, as the high amount is inconsistent with additivity. Non-additive phenotypes are associated with heterosis, a response displayed by hybrids between diverged parents and involving enhanced growth, stress tolerance, improved agronomic characteristics and also any transgressive trait (East 1936; Birchler et al. 2010; Ding et al. 2014). Metabolic heterosis has been documented in Arabidopsis (Fiévet et al. 2010), but this feature is relatively uncommon and primarily consistent with epistasis. In the case of *Boβ-LCY1*, heterozygosity at this locus, as displayed by the group B plants, may be all that is required to increase bixin concentration. Further genetic characterization and transcription and biochemical analysis of the genes encoding bixin biosynthesis will be necessary to fully examine this hypothesis and understand the associated mechanisms.

Phenotypically relevant variation of the *Boβ-LCY1* gene has also been observed in other species. Bang et al. (2007) characterized the *β-LCY* gene in two varieties of watermelon with contrasting fruit types and found that a SNPs resulting in a Phe to Val change at position 226 of the β-LCY protein was associated with the red to yellow colour change in the fruit. A similar case in papaya involved fruit carotene content (Devitt et al. 2010). The yellow-fruited variety accumulates lycopene and β-carotene, while the red-fruited variety accumulates lycopene in its mature fruit. One of the two *β-LCY* genes in papaya was expressed in chromoplasts of non-photosynthetic tissue. This *β-LCY* gene has a dinucleotide insertion and encodes an inactive, truncated enzyme resulting in the accumulation of lycopene in the red variety. In addition to having implications for carotene synthesis regulation, Devitt et al. (2010) noted that the corresponding molecular markers were important for the genetic improvement of papaya.

Polymorphism associated with phenotypic characteristics in plants has been demonstrated in other

members of the cyclase family of enzymes involved in carotenoid biosynthesis (Ronen et al. 2000). The CCS enzyme in *Capsicum annuum* adds κ terminals to antheraxanthin or violaxanthin molecules and β rings to *trans*-lycopene. Ha et al. (2007) identified a polymorphism in the corresponding gene resulting in non-synonymous amino acid changes and deletions affecting the ORF. These changes were associated with differences in the level and content of carotenoids in mature fruit. It has been suggested that inactivation of the *CCS* gene might result in lower quantities of carotenes in yellow-fruited *Capsicum* sp. varieties (Ha et al. 2007). Other changes in the genes of the cyclase family have been associated with altered carotene synthesis. Harjes et al. (2008) demonstrated in maize that mutations in the ε-*LCY* gene, which adds epsilon terminals to *trans*-lycopene, are associated with the differential accumulation of α-carotene versus β-carotene and β-cryptoxanthin. The identification of these polymorphic alleles of the *ε-LCY* gene facilitates the development of molecular markers suitable for selecting maize plants with increased β-carotene content in the seeds.

The expression level of the *Boβ-LCY1* gene during *Bixa orellana* seed development (Fig. 4a) was higher in the plants of flower group 1, haplotype A compared with those from plant N4, group 2 and plant N5 from haplotypes B and C. According to the bixin biosynthesis model (Cárdenas-Conejo et al. 2015), high *Boβ-LCY1* transcript levels (Bo-βLCY1 activity) in immature seeds correspond to less lycopene availability for bixin synthesis resulting in less bixin product. Therefore, a low *Boβ-LCY1* expression (less Boβ-LCY1 activity) promotes lycopene accumulation, which might subsequently be used for bixin synthesis. The behaviour of the *Boβ-LCY1* gene transcript and the low quantities of bixin in plants with white flowers in Group 1 (Fig. 3a; Table 1) support the bixin biosynthesis model (Cárdenas-Conejo et al. 2015), suggesting that in these plants lycopene is used by Boβ-LCY1 to produce other, non-bixin metabolites, which correlates with less bixin production. Boβ-LCY2 shows an opposite pattern (Fig. 4b) which does not correlate with this hypothesis. The *Boβ-LCY1* transcript levels and the quantities of bixin in haplotype B plants are also consistent with the bixin biosynthesis model (Cárdenas-Conejo et al. 2015); however, these features are not consistent with haplotype C plants. haplotype C plants displayed low *Boβ-LCY1* transcript levels and similar quantities of bixin compared with haplotype A plants. These results suggest an heterotic response because the bixin biosynthesis model predicts that haplotype C should display the highest quantities of bixin, but haplotype B, which has both alleles, is in fact the most productive group. We hypothesized that any modulation in expression would affect these responses if dosage sensitivity were involved. Furthermore, any dosage-sensitive gene in the regulatory hierarchy of the carotenoid pathway controlling this gene might also play a role in heterosis (Birchler et al. 2010). Additional experiments will be necessary to demonstrate this hypothesis. Still, no differences associated with the phenotype were observed using SSCP for other genes of the carotenoid route, including *Boβ-LCY2,* phytoene synthase, phytoene desaturase, epsilon cyclase or carotenoid cleavage dioxygenase.

Relevant to breeding, the achiote haplotypes differentiated by SNPs and associated with phenotypic traits, such as bixin production and flower and fruit colour, can easily be detected through PCR using specific primers (Fig. 3b).

Conclusions

The different alleles identified in the *Boβ-LCY1* gene of *B. orellana* associated with fruit and flower colour and bixin contents might also be involved in the regulation of carotenoid synthesis, as proposed for watermelon, papaya, chilli and maize (Bang et al. 2007; Ha et al. 2007; Harjes et al. 2008; Devitt et al. 2010). The identification of the genes involved in the bixin pathway is necessary to completely understand the regulation of this gene and the molecular basis for the heterosis phenotype associated with bixin production.

The discovery of these alleles in *Bixa orellana* provides molecular markers for selecting plants with increased bixin contents during early plant development, revealing an important tool for the genetic improvement of this woody plant, as a more accurate selection of productive individuals can now be made at an early stage of the plant. These findings have an important application in the production of bixin from *B. orellana* and provide incentives to study other key genes, such as *Boβ-LCY2* and *-cyclase*, to identify additional single-nucleotide polymorphisms (SNPs).

Acknowledgments This work was financially supported through grants from the Consejo Nacional de Ciencia y Tecnología (CONACYT) with Grant No. 46541, 98508, 220259. UC Mexus-CONACYT. J.A. Trujillo-Hernandez, P. Escobar Turriza, Y. Cárdenas-Conejo and LE. Garza-Caligaris were supported through CONACYT: MSc. Grants 254639 and 338189 and the postdoctoral Grant 290754, respectively. Victor Carballo-Uicab was supported by CONACYT PhD Grant 265369. The authors would also like to thank Dr. JA Huerta-Ocampo and Dr. C. Quiros for comments on the manuscript and anonymous reviewers for improving the manuscript. Work in the Comai laboratory was supported through a grant from the Mars Company.

References

Ambrósio Moreira P, Lins J, Dequigiovanni G, Veasey EA, Clement CR (2015) The domestication of Annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. Econ Bot 69:127–135. doi:10.1007/s12231-015-9304-0

Arce J (1999) El achiote *Bixa orellana* L. Cultivo promisorio para el trópico, 1a ed, Earth, Turrialba, Costa Rica

Auldridge ME, McCarty DR, Klee HJ (2006) Plant carotenoid cleavage oxygenases and their apocarotenoid products. Curr Opin Plant Biol 9:315–321. doi:10.1016/j.pbi.2006.03.005

Avendaño-Arrázate C, Pinzón-López L, Mendoza-López A, Campos-Rojas E, Correa-Navarro PJ, Godoy-Hernández G, Mijangos-Cortés J, Rivera-Madrid R (2012) Rescate y conservación del Achiote (*Bixa orellana* L.) en México. Agroproductividad 5:3–8

Bang H, Kim S, Leskovar D, King S (2007) Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β-cyclase (LCYB) gene. Mol Breed 20:63–72. doi:10.1007/s11032-006-9076-4

Bassam B, Caetano-Anolles G, Gresshoff P (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA (2010) Heterosis. Plant Cell 22:2105–2112. doi:10.1105/tpc.110.076133

Bouvier F, d'Harlingue A, Camara B (1997) Molecular analysis of carotenoid cyclase inhibition. Arch Biochem Biophys 346:53–64. doi:10.1006/abbi.1997.0278

Bouvier F, Dogbo O, Camara B (2003) Biosynthesis of the food and cosmetic plant pigment bixin (Annatto). Science 300:2089–2091

Cárdenas-Conejo Y, Carballo-Uicab V, Lieberman M, Aguilar-Espinosa M, Comai L, Rivera-Madrid R (2015) De novo transcriptome sequencing in *Bixa orellana* to identify genes involved in methylerythritol phosphate, carotenoid and bixin biosynthesis. BMC Genom 16:877. doi:10.1186/s12864-015-2065-4

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Rev Trends Plant Sci 15:266–274. doi:10.1016/j.tplants.2010.02.003

Cunningham FX Jr, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49:557–583. doi:10.1146/annurev.arplant.49.1.557

Cunningham FX Jr, Pogson B, Sun Z, McDonald KA, DellaPenna D, Gantt E (1996) Functional analysis of the beta and epsilon lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. Plant Cell 8:1613–1626

De Almeida P, Carvalho C, Clarindo WR (2006) Classical and molecular cytogenetic tools to resolve the Bixa karyotypes. Cytologia 71:391–398

Devitt LC, Fanning K, Dietzgen RG, Holton TA (2010) Isolation and functional characterization of a lycopene β-cyclase gene that controls fruit colour of papaya (*Carica papaya* L.). J Exp Bot 61:33–39. doi:10.1093/jxb/erp284

Ding H, Qin C, Luo X, Li L, Chen Z, Liu H, Gao J, Lin H, Shen Y, Zhao M, Lübberstedt T, Zhang Z, Pan G (2014) Heterosis in early maize ear inflorescence development: a genome-wide transcription analysis for two maize inbred lines and their hybrid. Int J Mol Sci 15:13892–13915. doi:10.3390/ijms150813892

East EM (1936) Heterosis. Genetics 21:375–397

Emanuelsson O, Brunak S, Gv Heijne, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nat Prot 2:953–971

Fiévet JB, Dillmann C, de Vienne D (2010) Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. Theor Appl Genet 120:463–473. doi:10.1007/s00122-009-1203-2

Frusciante S, Diretto G, Bruno M, Ferrante P, Pietrella M, Prado-Cabrero A, Rubio-Moraga A, Beyer P, Gomez-Gomez L, Al-Babili S, Giuliano G (2014) Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. Proc Natl Acad Sci USA 111:12246–12251. doi:10.1073/pnas.1404629111

Giuliano G, Rosati C, Bramley PM (2003) To dye or not to dye: biochemistry of annatto unveiled. Trends Biotechnol 21:513–516. doi:10.1016/j.tibtech.2003.10.001

Godoy-Hernández G (2000) El achiote una especie subexplotada. Cienc y Desarro 26:34–39

Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 38:D75–D80. doi:10.1093/nar/gkp902

Ha SH, Kim JB, Park JS, Lee SW, Cho KJ (2007) A comparison of the carotenoid accumulation in Capsicum varieties that show different ripening colours: deletion of the capsanthin-capsorubin synthase gene is not a prerequisite for the formation of a yellow pepper. J Exp Bot 58:3135–3144. doi:10.1093/jxb/erm132

Hanson L, McMahon KA, Johnson MAT, Bennett MD (2001) First Nuclear DNA C-values for 25 Angiosperm Families. Ann Bot 88:851–858. doi:10.1006/anbo.2000.1325

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Bermudez-Kandianis C, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333. doi:10.1126/Science.1150255

Hayashi K (1991) PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl 1:34–38

Huang H-Y, Chien C-H, Jen K-H, Huang H-D (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. Nucleic Acids Res 34:W429–W434. doi:10.1093/nar/gkl333

Hugueney P, Badillo A, Chen HC, Klein A, Hirschberg J, Camara B, Kuntz M (1995) Metabolism of cyclic carotenoids: a model for the alteration of this biosynthetic pathway in *Capsicum annuum* chromoplasts. Plant J 8:417–424

Jako C, Coutu C, Roewer I, Reed DW, Pelcher LE, Covello PS (2002) Probing carotenoid biosynthesis in developing seed coats of *Bixa orellana* (Bixaceae) through expressed sequence tag analysis. Plant Sci 163:141–145

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. doi:10.1093/bioinformatics/8.3.275

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081. doi:10.1038/nprot.2009.86

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 22^{DDCT} Method. Methods 25:402–408. doi:10.1006/meth.2001.1262

Lombello RA, Pinto-Maglio CAF (2014) Cytogenetics and reproductive biology of *Bixa orellana* L. (Bixaceae). Cytologia 79:379–386

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Mckeown G, Mark F (1962) The composition of oil soluble annatto food colour. J Assoc Agric Chem 45:761–766

Miura K, Lee J, Jin JB, Yoo CY, Miura T, Hasegawa PM (2009) Sumoylation of ABI5 by the Arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. Proc Natl Acad Sci USA 106:5418–5423. doi:10.1073/pnas.0811088106

Moise AR, Al-Babili S, Wurtzel ET (2014) Mechanistic aspects of carotenoid biosynthesis. Chem Rev 114:164–193

Moreno JC, Pizarro L, Fuentes P, Handford M, Cifuentes V, Stange C (2013) Levels of lycopene b-cyclase 1 modulate carotenoid gene expression and accumulation in *Daucus carota*. PLoS ONE 8:e58144–e58144. doi:10.1371/journal.pone.0058144

Nisar N, Li L, Lu S, Khin NC, Pogson BJ (2015) Carotenoid metabolism in plants. Mol Plant 8:68–82

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86:2766–2770. doi:10.1073/pnas.86.8.2766

Park HJ, Kim WY, Park HC, Lee SY, Bohnert HJ, Yun DJ (2011) SUMO and SUMOylation in plants. Mol Cells 32:305–316. doi:10.1007/s10059-011-0122-7

Pecker I, Gubbay R, Cunningham FX Jr, Hirschberg J (1996) Cloning and characterisation of the cDNA for lycopene ß-cyclase from tomato reveals a decrease in its expression during tomato ripening. Plant Mol Biol 30:806–819

Pfander H, Schurtenberger H (1982) Biosynthesis of C20-carotenoids in *Crocus sativus*. Phytochemistry 21:1039–1042

Rivera-Madrid R, Escobedo-GM RM, Balam-Galera E, Vera-Ku M, Harries H (2006) Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). Sci Hortic (Amsterdam) 109:165–172. doi:10.1016/j.scienta.2006.03.011

Rodríguez-Ávila NL, Narváez-Zapata JA, Aguilar-Espinosa ML, Rivera-Madrid R (2009) Full-length gene enrichment by using an optimized RNA isolation protocol in *Bixa orellana* recalcitrant tissues. Mol Biotechnol 42:84–90. doi:10.1007/s12033-008-9138-4

Rodríguez-Ávila NL, Narváez-Zapata JA, Aguilar-espinosa ML, Rivera-Madrid R (2011) Regulation of pigment-related genes during flower and fruit development of *Bixa orellana*. Plant Mol Biol Report 29:43–50. doi:10.1007/s11105-010-0207-z

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of β and old-gold color mutations in tomato. Proc Natl Acad Sci USA 97:11102–11107. doi:10.1073/pnas.190177497

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi:10.1093/molbev/msr121

Valdez-Ojeda R, Hernández-Stefanoni JL, Aguilar-Espinosa M, Ortiz R, Quiros CF, Rivera-Madrid R (2008)Assessing morphological and genetic variation in annatto (*Bixa orellana* L.) by sequence-related amplified polymorphism and cluster analysis. HortScience 43:2013–2017

Valdez-Ojeda R, Quiros CF, Aguilar-Espinosa MdL, Rivera-Madrid R (2010) Outcrossing rates in annatto determined by sequence-related amplified polymorphism. Agron J 102:1340–1345. doi:10.2134/agronj2009.0510

Yuan H, Zhang J, Nageswaran D, Li L (2015) Carotenoid metabolism and regulation in horticultural crops. Horticulture Research 2:15036. doi:10.1038/hortres.2015.36