



HAL
open science

Image analysis and polyphenol profiling unveil red-flesh apple phenotype complexity

Pierre Bouillon, Anne-Laure Fanciullino, Etienne Belin, Dimitri Bréard, Séverine Boisard, Béatrice Bonnet, Sylvain Hanteville, Frédéric Bernard, Jean-Marc Celton

► To cite this version:

Pierre Bouillon, Anne-Laure Fanciullino, Etienne Belin, Dimitri Bréard, Séverine Boisard, et al.. Image analysis and polyphenol profiling unveil red-flesh apple phenotype complexity. *Plant Methods*, 2024, 20 (1), pp.71. 10.1186/s13007-024-01196-1 . hal-04615546

HAL Id: hal-04615546

<https://hal.science/hal-04615546v1>

Submitted on 18 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access



Image analysis and polyphenol profiling unveil red-flesh apple phenotype complexity

Pierre Bouillon^{1,2}, Anne-Laure Fanciullino¹, Etienne Belin¹, Dimitri Bréard³, Séverine Boisard³, Béatrice Bonnet¹, Sylvain Hanteville¹, Frédéric Bernard² and Jean-Marc Celton^{1*}

Abstract

Background The genetic basis of colour development in red-flesh apples (*Malus domestica* Borkh) has been widely characterised; however, current models do not explain the observed variations in red pigmentation intensity and distribution. Available methods to evaluate the red-flesh trait rely on the estimation of an average overall colour using a discrete class notation index. However, colour variations among red-flesh cultivars are continuous while development of red colour is non-homogeneous and genotype-dependent. A robust estimation of red-flesh colour intensity and distribution is essential to fully capture the diversity among genotypes and provide a basis to enable identification of loci influencing the red-flesh trait.

Results In this study, we developed a multivariable approach to evaluate the red-flesh trait in apple. This method was implemented to study the phenotypic diversity in a segregating hybrid F1 family (91 genotypes). We developed a Python pipeline based on image and colour analysis to quantitatively dissect the red-flesh pigmentation from RGB (Red Green Blue) images and compared the efficiency of RGB and CIEL*a*b* colour spaces in discriminating genotypes previously classified with a visual notation. Chemical destructive methods, including targeted-metabolite analysis using ultra-high performance liquid chromatography with ultraviolet detection (UPLC-UV), were performed to quantify major phenolic compounds in fruits' flesh, as well as pH and water contents. Multivariate analyses were performed to study covariations of biochemical factors in relation to colour expression in CIEL*a*b* colour space. Our results indicate that anthocyanin, flavonol and flavanol concentrations, as well as pH, are closely related to flesh pigmentation in apple.

Conclusion Extraction of colour descriptors combined to chemical analyses helped in discriminating genotypes in relation to their flesh colour. These results suggest that the red-flesh trait in apple is a complex trait associated with several biochemical factors.

Keywords *Malus domestica*, Phenolic compounds, Image analysis, Red-flesh

Introduction

Studying nature's palette had always fascinated scientists. Mendel's seminal laws of inheritance resulted from his studies on white and pink pea flowers. This colouration is attributed to anthocyanin, a class of phenolic compounds, that are major determinants of plant organ colours (e.g. leaves, flowers, fruits skin and flesh) with a wide range of hue variation, from orange-red to violet-blue [1]. They mostly serve to attract pollinators or seed dispersers [2] and protect against various stresses [3, 4].

*Correspondence:

Jean-Marc Celton
jean-marc.celton@inrae.fr

¹ Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France

² IFO, 49140 Seiches sur le Loir, France

³ SONAS, SFR QUASAV Univ Angers, SONAS, SFR QUASAV, Univ Angers, F-49000 Angers, France



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

In recent years, our knowledge of the mechanisms leading to the synthesis and stability of anthocyanin pigments has been enhanced by progress in genetic and biochemistry [5, 6]. The transcriptional control of the anthocyanin biosynthetic pathway has been characterized [7], highlighting a common regulatory network in Eudicots [8]. More recently, molecular engineering has enabled production of anthocyanins in non-anthocyanin plants [9]. Studies have also identified the involvement of epigenetic mechanisms in anthocyanin biosynthesis and degradation [10, 11]. Some fruits naturally exhibit red pigmentation [12] and are therefore preferential models to study red colour biochemical and phenotypical expression. The major phenolic compounds in apple are hydroxycinnamic acids, flavanols, flavonols, dihydrochalcones and anthocyanins [13].

Anthocyanins are main determinants of red colour in apple [12]. Some cultivars display an ectopic accumulation of anthocyanins in the fleshy part of the fruits leading to the 'red-flesh' trait. This appealing phenotype [14] originates from a wild species of *Malus sieversii* [15] and is studied for its innovative aspects and potential health benefits [16, 17]. The genetic basis of flesh colour development has been characterised [18–20] in apple but current models do not explain the observed variations in intensity and pigment distribution. Studies have revealed a large diversity of phenolic compounds among red-flesh apple cultivars [21, 22], as well as their seasonal variability [23]. Environmental factors linked to anthocyanin accumulation, stability and colour expression have also been listed [16]. Among them, light intensity, water deficit and low temperature may promote anthocyanin synthesis in red-flesh cultivars. Other factors associated with expression of red colour may also be involved [24] such as pH [16, 25], interaction with other phenolic compounds leading to copigmentation events [26], or temperature during fruit storage [27].

Many plant studies rely on colour measurement [10, 28–30] and image-based phenotyping [31–34] to dissect the genetic determinism of colour development. Image-based colour measurement uses mostly colour conversion from RGB (Red Green Blue) images to different colour spaces to quantify colour variations among individuals [35]. Indeed, the RGB colour space is an additive colour model and is not suitable for colour comparison given that each colour is represented by a mix of various proportions of three distinct stimuli colours of light [36] thus, by definition, each colour channel can not be interpreted individually. Moreover, the RGB model is not a perceptually uniform space, the differences among colours in RGB space do not correspond to colour differences as perceived by humans

[37]. RGB and CIEL*a*b* values have been directly used to detect quantitative trait loci (QTL) associated with colour variations in grape [38] or salvia [39]. Li et al. [40] used Gaussian estimation of pixel distributions to study colour patterning in the foliar ornamental coleus. Deep learning techniques have also been used to overcome limitations of pixel colour information (confounding effects and continuity of colours) in the case of complex plant vegetation segmentation [41]. Colour measurement can also deliver useful information on fruit intrinsic physiological changes (i.e. organoleptic, nutritional, visual or non-visual defects) [42] during fruit development, ripening and post-harvest conservation [36, 37]. For example, a convolutional neural network has been trained to evaluate fruit maturity based on starch index in apple and pear [43].

In apple, computer vision systems have been developed to sort fruits according to skin colour [44–46] or to assess the maturity stage of fruits based on skin colour measurements [47, 48]. Other systems have been designed to detect visual defects on fruits [49, 50]. An index based on colour measurements in CIEL*a*b* has been established to characterize enzymatic browning of apple slices [51, 52]. The red-flesh trait in apple is difficult to assess visually due to non-homogeneous and non-continuous colour distribution leading to pigmentation pattern and variations in red hue. Various procedures to evaluate red-flesh colour based on imaging techniques have already been developed in apple [53, 54]. These methods include the evaluation of anthocyanin contents derived from images of red-flesh apple fruit section [55], the estimation of average overall red colour [54] and the non-destructive analysis of fruits to predict red-flesh colour using interactance spectroscopy [56, 57]. However, current methods based on spectroscopy are less predictive than destructive methods [57], while most RGB image-based methods do not consider evaluation of non-homogeneous surface [38]. Moreover absolute quantification of anthocyanins is laboratory-based and involves costly and time-consuming steps.

In this study, we developed a new method to dissect flesh colour from fruit section images acquired with a RGB scanner and the conversion of RGB images into CIEL*a*b* colour space. This method was applied to identify biochemical factors involved in colour variations in a F1 apple progeny segregating for the red-flesh trait and displaying a large variability in red-flesh phenotypes. Colorimetric variables from the CIEL*a*b* colour space were used to build a Partial Least Squares regression model allowing the identification of the main biochemical factors controlling pigmentation variations in red-flesh apple.

Methods

Experimental approach

Our objective was to develop a robust method to differentiate apple fruits based on colour parameters using image analysis. Colour descriptors from RGB and CIEL*a*b* colour space were used to discriminate genotypes from a F1 apple progeny segregating for the red-flesh trait. Major phenolic compounds potentially involved in red pigmentation were quantified, and correlations among these compounds and the colour descriptors were estimated. Finally, models based on colours descriptors (a*, b* and hue) variations were established to identify biochemical factors involved in red-flesh pigmentation.

Plant material

Fruit harvest was conducted from late August to mid-October in 2022 and 2023 in IFO orchard (L'Anguicherie, 49,140 Seiches-sur-le-Loir, France/ GCS: 47°37'52.5"N 0°19'38.4"W). Our study was carried out on 91 genotypes from a F1 hybrid family segregating for the red-flesh trait (plantation year: 2017). Each genotype is represented by one tree grafted on M9 Pajam®2 Cepiland C.O.V. For each genotype harvested at maturity (brix values varying from 13 to 22; starch index between 6 and 8), four representative fruits were dedicated to image analysis for estimation of the red-flesh intensity and distribution,

while four other fruits were sampled for the quantitation of phenolic compounds. Fruits that were positioned in the middle of each tree, with similar exposure to light, same developmental stage, and similar diameter were harvested preferentially to limit intra- and inter-tree bias. Image acquisition was performed for fruits harvested in 2022 and 2023 while phenolic compounds quantification was carried out only on fruits harvested in 2022. Colour descriptors and biochemical factors were estimated for four apples per genotype.

Image analysis

Evaluation of fruit flesh colour

Prior to image acquisition, fruits were visually classified into six intensity classes according to a red colour scale (from white flesh 0 to dark red-flesh 5, Figure S1). The mean of the four fruits provided the colour index of each genotype and enabled image analysis conformity evaluation.

Image acquisition

In an affordable phenotyping approach [58, 59], images of fruits were acquired using a RGB flatbed scanner Canon LIDE 400 (Fig. 1A). A transversal section of each fruit was scanned immediately after cutting to avoid enzymatic browning. A shade box was positioned upon the

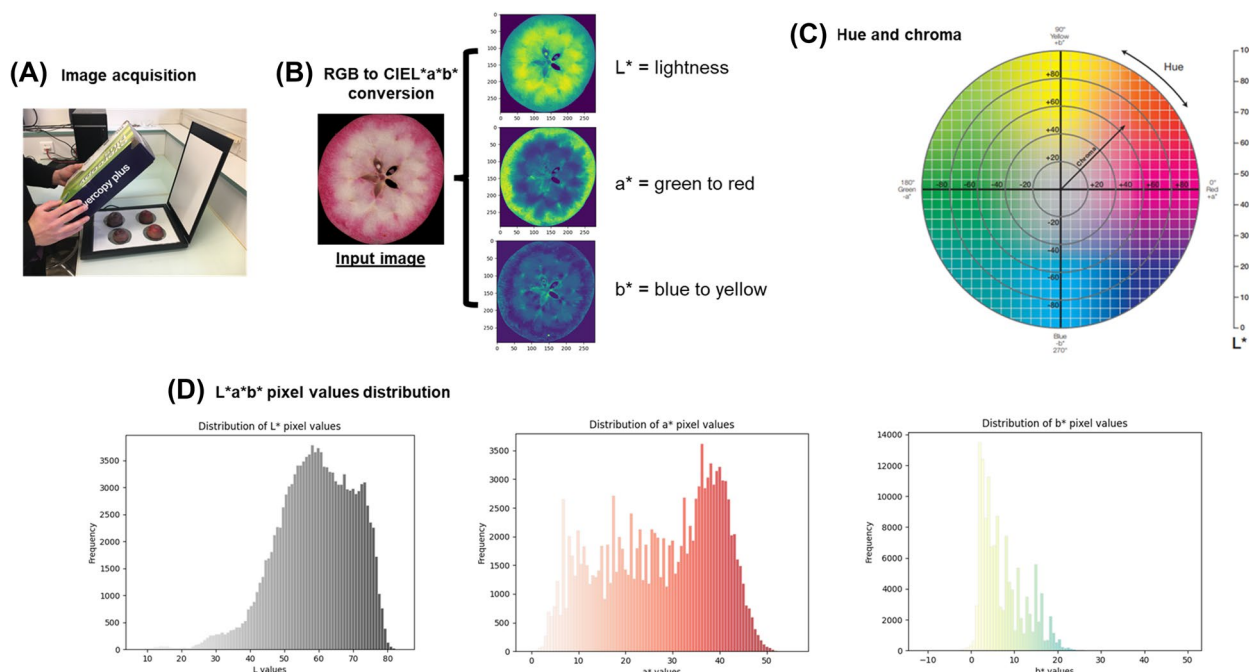


Fig. 1 Overview of the analysis pipeline **A**: image acquisition is performed with a RGB flatbed scanner. **A** shade box ensures reproducible lighting conditions. **B**: RGB images of apple sections are converted to CIEL*a*b* colour space. This conversion is nonlinear. Therefore, the RGB colour space was transformed into CIEXYZ before CIEL*a*b* conversion. **C**: chroma and hue are estimated from a* and b* colour descriptors. **D**: statistical descriptors are calculated (mean and standard deviation) on CIEL*a*b* pixel distribution to approximate colour for a non-homogeneous surface

system to reduce external light pollution. Image acquisition was achieved with the IJ scan utility software and images were stored in png format.

Image processing

An in-house object detection pipeline was used to obtain individual apple section. Firstly, an Otsu's thresholding was performed to separate apple objects from the background. Then, a Connected Component Analysis isolated each apple object and drew a bounding box around the object contours, giving four apple section images from a single file. Final images were converted in.png format and stored for further analysis. Apple section images resolution are approximately 400*400 pixels.

Image analysis

The image analysis pipeline was written in Python. Firstly, RGB images of apple sections were converted to CIEXYZ, then to CIEL*a*b* with default parameters Illuminant = "D65" (Fig. 1B) using Scikit-image [60]. The image acquisition system provided repeatable lightning conditions and ensured relative comparison between genotypes. CIEL*a*b* is a device-independent and uniform colour space derived from CIE XYZ space. In the CIEL*a*b* colour space, L* is associated with the lightness of the colour (L* = 0 means black and L* = 100 refers to white). a* is related to the colour tonalities from red (+) to green (-), and b* is associated with yellow (+) to blue (-) colour variations [61]. Basic statistical descriptors of pixel values distribution were calculated (mean and standard deviation) with Numpy library [62] (Fig. 1C). Hue angle was calculated with the formula $h = \arctan\left(\frac{a^*}{b^*}\right)$ and chroma $C^* = \sqrt{a^2 + b^2}$ (Fig. 1D).

Determination of phenolic content

Metabolites extraction

Fruits without peel were cut and two flesh pieces from the sun-exposed side of the fruit were selected (about 5 g fresh weight per fruit). Samples were immediately frozen in liquid nitrogen, then stored at -80°C and freeze-dried. This dried material was crushed into a fine and homogeneous powder and stored until analysis. About 50 mg of this powder was precisely weighed in a 2 mL microtube. Polyphenols extraction was adapted from [63]. Samples were extracted during 30 min in an ultrasonic bath with 1.5 mL of methanol (MeOH) containing 5% acetic acid (v/v) and spiked with 2 µg of internal standard (IS) in order to control the extraction recovery. After centrifugation (14,800 rpm/10 min), supernatants were filtered with RC filters (0.2 µm/13 mm; Macherey-Nagel) and transferred to LC vials.

Phenolic compound detection and quantification

First, pools of ten hybrids from the F1 progeny were analyzed using an ACQUITY quantitative ultra-high performance liquid chromatography (UPLC) H-Class Series system coupled to a Xevo G2-XS mass spectrometer (Waters) in order to identify the metabolites by comparing the retention time and *m/z* data from literature and/or with data obtained from authentic standards. The UPLC-Q-TOF was equipped with a quaternary solvent manager, a sample manager and a column compartment as chromatographic part and coupled to a quadrupole time-of-flight (Q-TOF) mass spectrometer. The data was processed with the Progenesis QI software (Waters, Elstree, UK). Ten phenolic compounds were identified with level 1 of confidence according to [64] and appeared to be the major phenolic compounds in our samples. These ten phenolic compounds were quantified in all our genotypes. Water content was calculated based on fresh weight (FW) and dry weight (DW) with the following equation: $WC = 1 - \frac{DW}{FW}$. pH was measured using a pH-meter Seven Compact S210 (Metler Toledo) after fruit powder re-suspension to match corresponding individual fruit water content.

The calibration curves were set to the concentration range expected for each compound and their linearity was assessed by injecting 6 levels of calibration standards in three replicates. Residuals (difference between nominal concentration and calculated concentration by the linear model) and their distribution (normally distributed around the mean) were monitored (Minitab 19 software). Precision was evaluated by repeated analysis of standards within different analytical batches. Limit of detection (LOD) and limit of quantification (LOQ) were defined as 3- and 10- fold the signal-to-noise ratio, respectively. When considering samples with area under the curve (A.U.C) inferior to LOD, they were expressed as LOD/2 to limit bias for statistical analysis.

Each sample was spiked with the same level of IS at the start of the extraction procedure in order to evaluate the recovery of extraction. The precision of the extraction and quantification process was evaluated from the analysis of 5 replicates of the 2 parents of the segregating F1 progeny.

UPLC-UV analysis

Methanol (HiPerSolv CHROMANORM for LC-MS) was purchased from VWR. Acetic acid, formic acid and dimethylsulfoxide (analytical reagent grade) were purchased from Fisher Scientific. Ultrapure water was obtained from a MilliQ advantage A10 purification system (Millipore). Quercetin 3-galactoside, cyanidin 3-galactoside chloride, procyanidin B1, procyanidin B2,

procyanidin C1, (+)-catechin, (–)-epicatechin, chlorogenic acid and phlorizin standards were purchased from Extrasynthese. 4-*p*-coumaroylquinic acid was purchased from Ambinter. Daidzein (internal standard) was purchased from Molekula. All standards were furnished with a certificate of analysis. Stock solutions of procyanidin B1, procyanidin B2, procyanidin C1, phlorizin, quercetin 3-galactoside and daidzein were prepared in DMSO at a concentration of 5 mg.mL⁻¹. Stock solutions of (+)-catechin and (–)-epicatechin were prepared in MeOH at a concentration of 5 mg.mL⁻¹. Stock solutions of chlorogenic acid and 4-*p*-coumaroylquinic acid were prepared in H₂O at a concentration of 5 mg.mL⁻¹. Stock solution of cyanidin 3-galactoside chloride was prepared in MeOH containing 5% v/v acetic acid at a concentration of 5 mg.mL⁻¹. All the dilutions were then carried out in MeOH except for cyanidin 3-galactoside chloride in MeOH containing 5% v/v acetic acid. Each calibration curve was prepared using six different standards concentrations. Three replicates were used for each calibration level to determine LOD and LOQ (Table 1). Quantitative ultra-high performance liquid chromatography with ultraviolet detection (UPLC-UV) analysis was performed on a ThermoFisher Scientific Vanquish Flex UPLC system equipped with a quaternary solvent manager, a sample manager and a column compartment as chromatographic part and coupled to a variable wavelength detector as detection part. A sample volume of 10 µL was injected onto a ZORBAX RRHD StableBond Aq column (2.1 x 150 mm; 80 Å; 1.7 µm). The samples were kept at 10°C and the column was maintained at 30°C with a flow rate of 0.3 mL.min⁻¹. The mobile phase consisted of 0.1%

formic acid in both H₂O (A) and MeOH (B) used in gradient mode as follows: from 5 to 15% B in 5 min, then 15 to 35 % B from 5 to 30 min, then 35 to 50% B from 30 to 35 min, then 50 to 100% B from 35 to 37 min, hold at 100% B from 37 to 39 min, and afterward the column was re-equilibrated at initial conditions during 8 min. The UV detection was measured at different wavelengths optimized for individual compounds and listed in Table 1. Data were processed using Xcalibur software.

Statistical analysis

Broad-sense heritability (h^2) of each colour descriptors was estimated by intra-class correlation analysis [65] with the following formula: $h^2 = \sigma_B^2 / (\sigma_B^2 + \sigma_\epsilon^2)$ where σ_B^2 and σ_ϵ^2 were the individual genetic and residual variances respectively (Table 2). Genetic and individual variances were estimated from the linear model: $y_{jk} = \mu + B_j + \epsilon_{jk}$, with μ the population mean of the trait (treated as a fixed effect), B_j the “true effect” of the j th individual, and ϵ_{jk} the special environmental error. B_j is assumed to be a random variable sampled from a normal distribution with a mean zero and variance σ_B^2 .

Principal Component Analyses (PCA) were carried out to investigate portability from RGB to CIEL*a*b* colour space, adequation with visual notation and ability to discriminate genotypes. Data were normalized by subtracting means and then dividing every measure by the standard deviation. Squared cosine were calculated to confirm interpretability of explanatory variables on first components. The squared cosine shows the importance of a component for a given observation and corresponds to the square of the cosine of the angle from the

Table 1 Method parameters

Compound	Retention time (min) ^a	UV detection λ max (nm)	LOD (ng) ^b	LOQ (ng) ^c	Linearity range (ng) ^d	R ²
Procyanidin B1	12.75 ± 0.12	279	0.24	0.47	0.9 – 47.3	1.000
(+)- catechin	13.45 ± 0.34	279	0.50	1	2.5 – 99.8	0.998
Procyanidin B2	16.26 ± 0.07	279	0.24	0.48	9.6 – 478.5	1.000
(-)- epicatechin	16.74 ± 0.25	279	0.25	0.49	4.9 – 247.3	0.996
Chlorogenic acid	17.44 ± 0.39	325	0.49	0.99	98.9 – 1483.5	0.997
Cyanidin 3-galactoside	18.62 ± 0.11	512	2.40	4.7	4.7 – 236.3	0.999
4- <i>p</i> -coumaroylquinic acid	20.79 ± 0.34	312	0.13	0.25	5 – 250	1.000
Procyanidin C1	21.99 ± 0.33	279	0.90	2.3	4.6 – 232	1.000
Phlorizin	33.41 ± 0.11	284	0.09	0.23	0.9 – 46.2	0.998
Quercetin 3-galactoside	36.21 ± 0.17	350	0.24	0.47	0.5 – 9.5	0.976
Daidzein (IS)	38.73 ± 0.19	250	0.10	0.25	2.5 – 49	0.997

^a Values are mean ± SD (n=3)

^b The signal-to-noise was set to 3:1

^c The signal-to-noise was set to 10:1

^d Suitable for samples

right triangle made with the origin, the observation, and its projection on the component.

To further explore the relationship between the various biochemical factors and our descriptors, we built Partial Least Square (PLS) regression models (Table S4). An advantage of PLS-regression is its ability to deal with multiple colinearity compared to classical regression model [66]. The objective of PLS regression was to quantitatively dissect covariation of CIEL*a*b* colour parameters as response variable: a* / b* / hue with biochemical predictors and to identify relevant biochemical factors involved in colour expression. First, relations between colour descriptors and each polyphenol content were represented in order to identify skewed distribution that require transformation (Figure S2). We performed log-transformation for anthocyanin contents. Variable of Importance (VIP) values were then calculated following [67] with the *getVIPVn* function to extract important biochemical factors in colour expression. A variable with a VIP < 0.8 is a variable that appears to have a negligible impact on the model. The number of PLS components were determined as follow: a new component *h* is added to the model if the percentage of Y dispersion explained by component *h* is more than 1 percent. Model validation was achieved with internal validation based on calculation of Root Mean Square Error (RMSE) on test set (20% of dataset). R4.0.3 software was used to perform statistical analysis [68]. *PCA* function implements in *factomineR* package [69] and *ropls* package [70] were used for respectively PCA and PLS approaches. A graphical display of a correlation matrix using *ggplot2* of *ggcorrplot* package [71] was chosen to represent the correlation matrix.

Results

Study of colour descriptors from RGB and CIEL*a*b* spaces

Colour descriptors from RGB and CIEL*a*b* colour spaces were calculated from four fruits per genotype (Table S2). In RGB colour space, red-coloured fruits were associated with low R, G and B values (Fig. 2). The green channel exhibited more variations for mean and standard deviation values. Considering CIEL*a*b* colour space, L* and a* parameters exhibited more variations than b*. Highly-coloured fruits were associated with high a* values and low L* values. Moreover, standard deviation of a* values exhibited more variations than L* and b* channels. Multivariate analyses of colour descriptors from RGB and CIEL*a*b* colour space were performed. Our aim was to confirm the possibility of converting RGB into CIEL*a*b* data and to evaluate the ability of the two colour spaces to distinguish genotypes for the red-flesh pigmentation. A PCA was performed on RGB colour descriptors. First and second axis accounted for 61.8% and 25.4% (Fig. 3)

of the total variation, respectively. Axis 1 separates genotypes according to flesh colour (colour index) while axis 2 mostly separates the homogeneous from the heterogeneous flesh colour genotypes. However, less-coloured genotypes (colour index 0,1 and 2) were closely grouped. R, G and B values were strongly correlated to each other and negatively correlated to colour index.

R, G and B colour descriptors showed h^2 values between 2022 and 2023 image data (Table 2, raw image data: Table S2) with h^2 values of 0.755, 0.783 and 0.763, respectively. For CIEL*a*b* colour descriptors, L*, a* and b* values were 0.786, 0.787 and 0.744, respectively. Visual notation exhibited lower broad-sense heritability than RGB and CIEL*a*b* colour descriptors with h^2 values of 0.67.

Considering CIEL*a*b* colour space, red-flesh fruits were associated with high a* and chroma values, and low L*, b* and hue values (Fig. 2). Our results are consistent with CIEL*a*b* colour space definition [36], a* and b* values characterised respectively green to red and blue to yellow variations. Standard deviation values of a* and L* channels exhibited more variations, higher values should be associated with non-homogeneously pigmented phenotypes. Chroma indicated saturation of colour and was correlated to a* and colour index (Pearson correlation of 0.98 and 0.96, respectively - Fig. 5). Principal component analysis based on colour descriptors from the CIEL*a*b* colour space permitted to distinguish genotypes according to their flesh colour index. First and second axis accounted for 63.7% and 20.4% (Fig. 4) of the total variation, respectively. Hue values, ranging from red (0) to yellow (60), exhibited a negative correlation with axis 1. Concomitantly, within the RGB colour space, axis 1 effectively distinguished genotypes based on red-flesh colour (colour index), while axis 2 discriminated between genotypes characterized by homogeneous and heterogeneous pigmentation.

Phenolic compound composition and biochemical factors

Several biochemical parameters (phenolic compound contents, dry matter content and pH values) were measured to evaluate their influence on colour expression (Fig. 5A). Polyphenol concentrations are expressed in $\mu\text{g}\cdot\text{g}^{-1}$ of fresh weight (FW) and summarised in total content per class (Table S1, Fig. 5B, C). Relative standard deviation varied from 0.78 to 4.89 with an overall mean of 2.74 when considering the two parents for the ten phenolic compounds (Table S3). Accordingly, two replicates for each genotype (two distinct metabolites extractions) were analyzed and the metabolite content was expressed as the mean of these two values. Hydroxycinnamic acids, among all measured compounds, reached the highest concentrations:

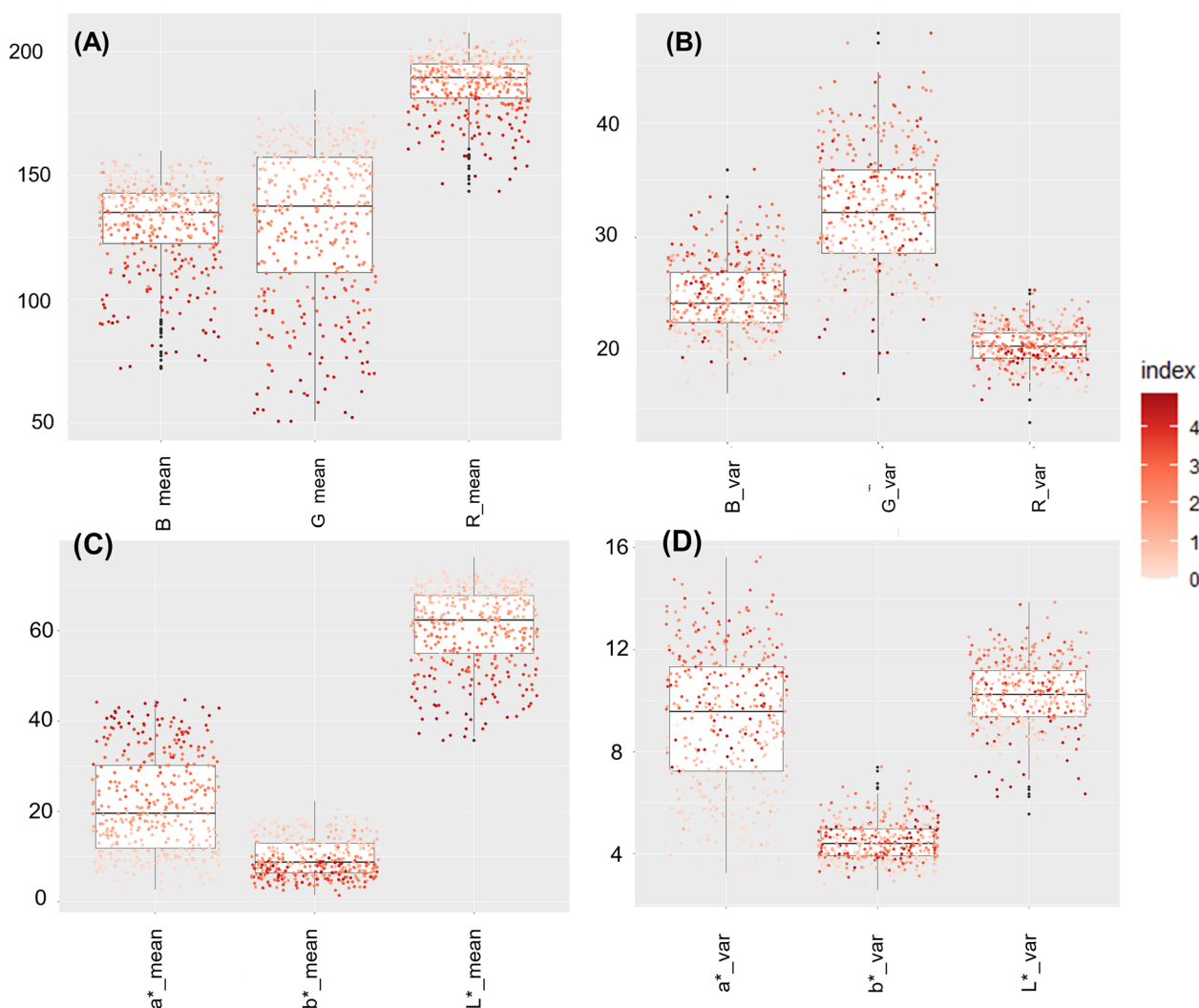


Fig. 2 Boxplot showing colour descriptors distribution for RGB and CIEL*a*b* colour space. Colour descriptors were averaged for four apples per genotype. Dots are coloured according to the colour index. The median (denoted by a horizontal bar in the box), the 25th percentile (denoted by the bottom edge of the box), the 75th percentile (denoted by the top edge of the box) and the dots indicate single observations. **A:** Mean of RGB colour descriptors. **B:** Standard deviation of RGB colour descriptors. **C:** Mean of CIEL*a*b* colour descriptors. **D:** Standard deviation of CIEL*a*b* colour descriptors

from 7.92 to 141.07 and from 49.93 to 804.25 $\mu\text{g}\cdot\text{g}^{-1}$ of FW for 4-*p*-coumaroylquinic acid and chlorogenic acid, respectively. Anthocyanins were represented by cyanidin 3-galactoside and, as expected, their content varied greatly among our samples: from 2.37 to 309.47 $\mu\text{g}\cdot\text{g}^{-1}$ of FW. It was followed by flavanols: between 0.45 and 64.59, 2.02–194.16, 0.21–49.47, 2.96–282.59, 2.32–123.79 $\mu\text{g}\cdot\text{g}^{-1}$ of FW for (+)-catechin, (–)-epicatechin, procyanidin B1, procyanidin B2, and procyanidin C1, respectively. Phlorizin (class of dihydrochalcone) contents ranged from 0.63 to 32.38 $\mu\text{g}\cdot\text{g}^{-1}$ of FW. Finally, within the flavonol group,

quercetin 3-galactoside was the most concentrated: contents ranged from 1.26 to 11.65 $\mu\text{g}\cdot\text{g}^{-1}$ of FW. Water contents and pH varied among individuals with a mean water content of 81.5% (range from 76.5% to 91.0%) and a mean measured pH value of 3.18 (range from 2.91 to 3.83). Red-flesh colour showed a clear relationship to anthocyanin contents (Pearson correlation = 0.75) and confirmed the possibility to estimate anthocyanin contents from red colour [55]. However, logarithmic evolution of anthocyanin contents suggests the existence of a saturation point. Beyond this, the red colour intensity does not change and could

Table 2 Broad-sense heritability of colour descriptors measured for fruits harvested in 2022 and 2023

Descriptor	Broad-sense heritability (h ²)
R mean	0.755
G mean	0.783
B mean	0.763
R var	0.372
G var	0.604
B var	0.612
L* mean	0.786
a* mean	0.787
b* mean	0.744
L* var	0.59
a* var	0.649
b* var	0.392
Chroma	0.758
Hue	0.79
Index (visual notation)	0.67

lead to underestimation of anthocyanin contents in a linear model.

Correlation between biochemical factors and colour data

Pearson correlations were calculated to study association between biochemical factors and colour descriptors. As observed in RGB-PCA (Fig. 3), R, G and B channels were highly correlated (correlation = 0.96). RGB descriptors were also negatively correlated to colour index (from -0.91 to -0.97) and anthocyanin contents (from -0.81 to -0.86).

L*a*b* and RGB colour descriptors were highly correlated with absolute values oscillating between 0.91 and 1 for L* and a* parameters. b* exhibited lower values with correlations of 0.66, 0.75 and 0.53 for R, G and B parameters respectively. a* parameter was strongly correlated to colour index (0.97) and anthocyanin contents (0.75).

Modelling of colour descriptors with biochemical factors

Hue values (Fig. 6A) distribution was coherent with colour variations from white-off to red [72]. Hue was expressed through a single component model (R² = 0.73 and RMSE = 6.92 - Fig. 6B) which combine predictor variables found in a* and b*: Anthocyanin, flavanol, flavanol contents and pH have VIP values of 2.02, 1.08, 0.88 and 1.97, respectively (Fig. 6C). One component was determined for a* values with R² = 0.84 (Fig. 7A) and RMSE = 3.26. Anthocyanin, flavanol and flavanol contents had the greatest impact on the model construction with VIP values equal to 2, 1.1 and 1.04, respectively (Fig. 7B). One component was also sufficient

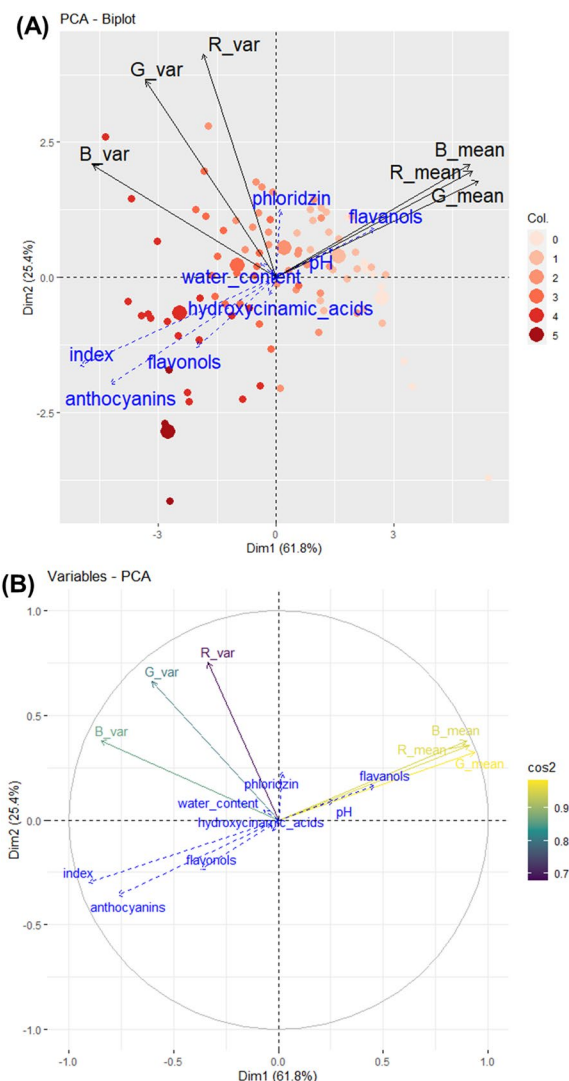


Fig. 3 Multivariate analysis based on RGB colour space PCA was performed for colour descriptors of the 91 genotypes. R, G and B mean and standard deviation values were considered as active variables in component calculation. Biochemical factors and colour index values were added as supplementary variables. Data were normalized before performing PCA (feature scaling). **A:** The biplot shows the PCA scores of the explanatory (in black) and supplementary (in blue) variables as vectors. Individuals are coloured according to colour index. Six bigger dots represent barycentres of colour groups. **B:** Correlation circles of colour variables. Arrow colour indicates the cos2 of each explanatory variables on dimensions 1–2

for b* values to capture most of the inertia with R² = 0.56 (Fig. 8A) and RMSE = 2.27. Anthocyanin, flavanol contents and pH were significant in the model construction with VIP values equal to 1.92, 1.18 and 1.07, respectively (Fig. 8B).

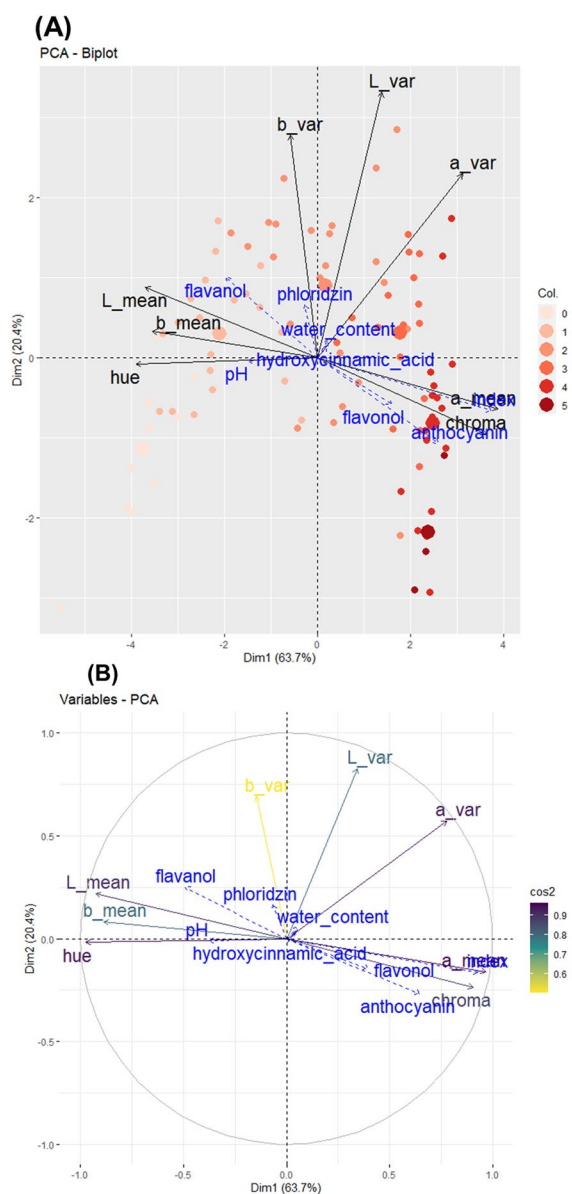


Fig. 4 Multivariate analysis of CIEL*a*b* colour space Principal component analysis (PCA) was performed for colour descriptors of the fruits of the 91 genotypes. L*, a* and b* mean and standard deviation were considered as active variables in component calculation. Biochemical factors and index were added as supplementary variables. Data were normalized before performing PCA (mean scaling). Representation of each genotype on the biplot is coloured according to colour index. **A:** The biplot shows the PCA scores of the explanatory (in black) and supplementary (in blue) variables as vectors. Individuals are coloured according to colour index. Six bigger dots represent barycentres of colour groups. **B:** Correlation circles of colour variables. Arrow colour indicates the cos2 of each explanatory variables on dimensions 1–2

Discussion

Distribution of the contents of phenolic compounds

Total phenolic compounds and individual concentrations

(Table S1) were consistent with those observed in some wild apple species [73] or cider apples [74, 75]. We have to consider that the genetic background of red-fleshed apple varieties is close to that of the wild apple species [15].

As expected, anthocyanin contents varied greatly among our samples. Some genotypes displayed higher anthocyanin concentrations, with anthocyanin contents superior to 100 µg.g⁻¹ of FW for 19 samples, than those found in literature [21, 22, 63]. Indeed for these particular genotypes, anthocyanin content was comparable to that of other anthocyanin-rich species such as strawberry (*Fragaria × ananassa*) or blueberry (*Vaccinium sp.*) [76]. These results suggest that red-flesh apples are an interesting natural food source of anthocyanins. However, these contents could have been overestimated (in comparison to whole fruit) by our sampling method: we always selected flesh tissues from the sun-exposed side of the fruit to allow comparison between genotypes. The sun-exposed side of apple fruits is generally more concentrated in phenolic compounds [22].

Dissection of the red-flesh trait in the CIEL*a*b* colour space

In this study, we confirmed the transferability of RGB to CIEL*a*b* colour space and the suitability of these two colour spaces with visual notation. Both colour spaces were well-adapted in discriminating genotypes. Moreover, colour descriptors were more accurate than visual notations because of their higher broad-sense heritability values. Indeed, colour index broad-sense heritability was 0.67 while values were comprised between 0.744 and 0.787 for colour descriptors. a* exhibited slightly higher broad-sense heritability than the other colour descriptors.

R, G and B parameters were individually correlated to colour index. Strong associations of RGB variables in PCA space confirm the difficulty of studying and interpreting each of them. This behaviour is relative to additive colour theory. By definition, low R, G or B values taken separately are associated with dark colours. In our study, dark colours corresponded to red-flesh colours. In other anthocyanin-rich species [1], phenotyping colours can be more challenging using RGB colour space. Only considering dark colours could lead to overlap divergence between dark-red and deeper purplish phenotypes. Consequently, considering R, G and B individually will lead to bias in distinguishing these phenotypes, given that, colour in RGB space is defined by mixing R, G and B values together [77].

Interestingly, colour heterogeneity was also repeatable between 2022 and 2023 with h² values comprised between 0.372 and 0.649, indicating a potential genetic

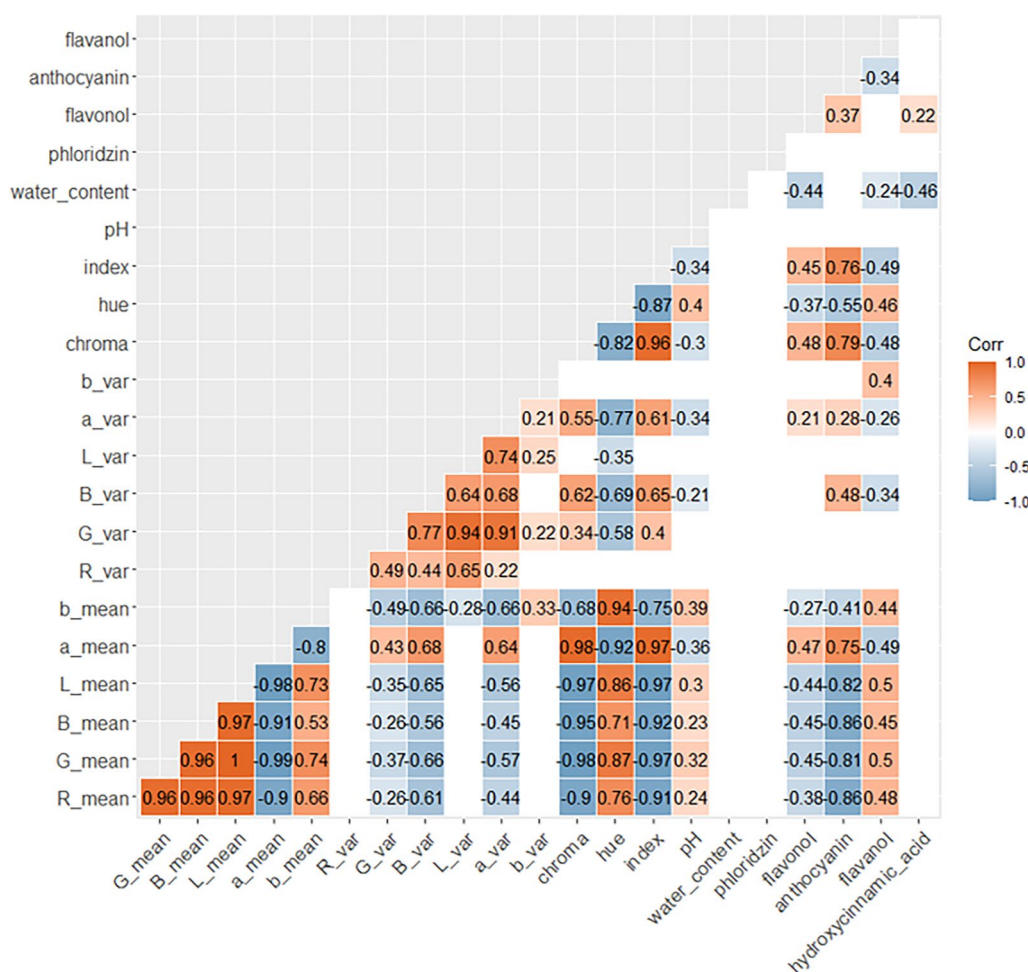


Fig. 5 Pearson correlation matrix of biochemical factors and colour descriptors. Only significant values were considered ($P < 0.05$). Positive correlations are highlighted in orange, while negative correlations are highlighted in blue

control. However, we did not identify any statistical association between colour heterogeneity (L^* , a^* and b^* variance) and particular phenolic contents (data not shown). This lack of association could be attributed to the limited-area in the sampling method which does not distinguish red and white parts in heterogeneous pigmented fruits, or, alternatively, to the regulation of pigmentation heterogeneity by an activator/repressor system during fruit development [78]. Underhill et al. [38], argued that a major challenge in phenotyping grape berry colour is the classification of fruits into discrete classes despite a continuous variation within and among genotypes. Our image analysis pipeline enables relative colour comparison by considering each apple image as pixel matrices and then estimates statistical descriptors. The use of quantitative colour measurements rather than qualitative categorisation could help in detecting minor QTL related to flesh colour variations [39].

Identification of biochemical factors involved in red pigmentation

PLS regression models enabled identification of important biochemical factors involved in colour expression. Altogether, we found that anthocyanins, flavonols, flavanols and pH were involved in hue variations and, therefore, in red-flesh colour expression with model coefficients of -0.51 , -0.22 , 0.27 and 0.24 , respectively. a^* colour variation was associated with anthocyanins, flavonols and flavanols (Fig. 7). The importance of non-anthocyanin compounds in red colour expression supported copigmentation events where flavonols could be copigments of anthocyanins, leading to a hyperchromic shift providing a deep red colouration [26]. In contrast to anthocyanins and flavonols, flavanols were negatively associated with higher a^* values (coefficients of 0.53 , 0.28 and -0.29 , respectively). Distribution of flavanol and anthocyanin contents suggested an imbalance between

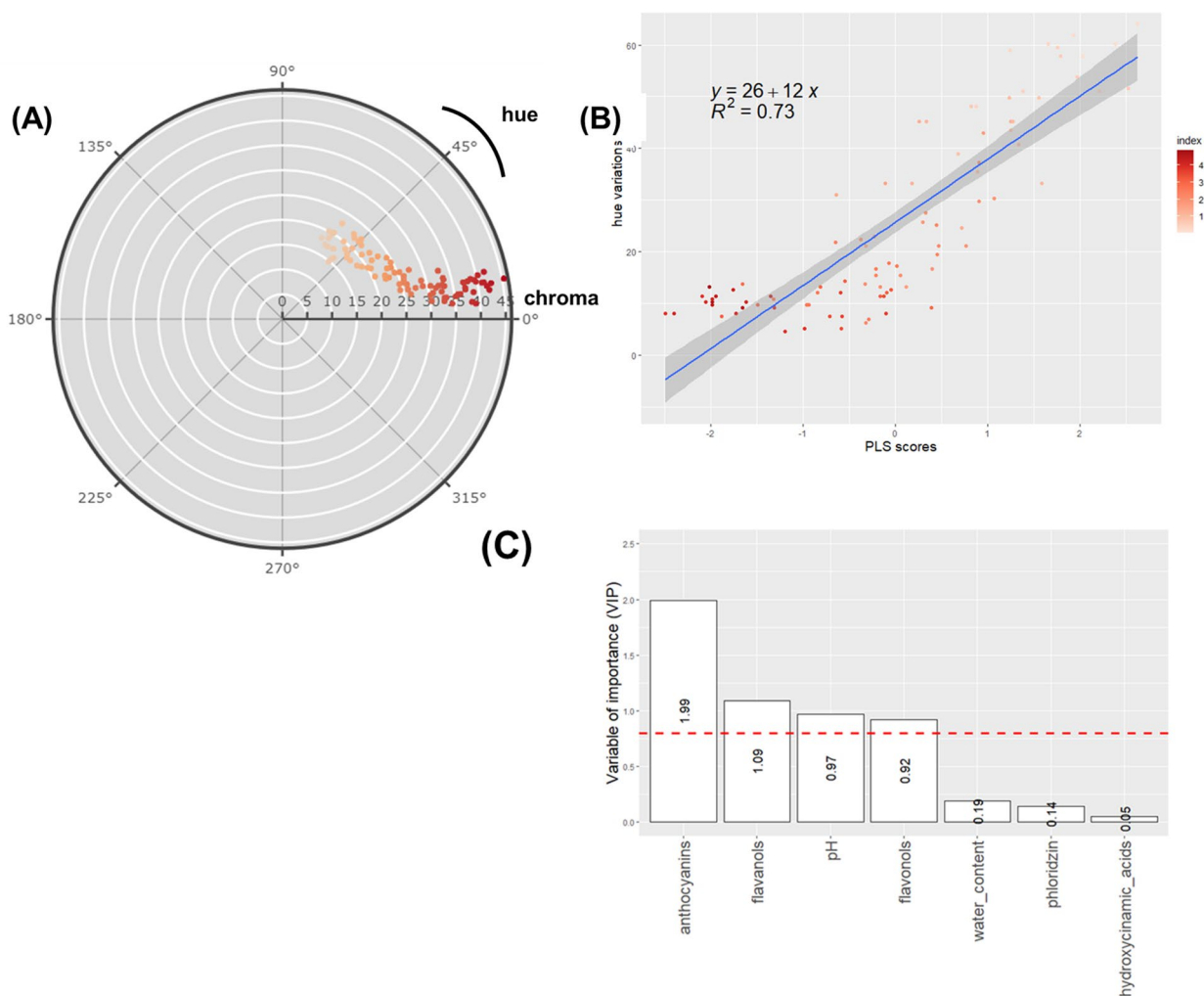


Fig. 6 Modelling of hue colour by biochemical factors Supervised multivariate analyses were performed using Partial Least Square (PLS) regression. Hue was considered as a response variable (Y) and biochemical factors as predictors (X). **A:** Circular plot showing distribution of hue and chroma. Dots are coloured according to hue. **B:** Scatter plot of observed hue and predicted values of the PLS model. Dots are coloured according to their hue from green (lower values) to red (higher values). **C:** Variable of importance (VIP) scores. A variable with a VIP value above 0.8 is a variable that appears to have a considerable impact on the model

anthocyanin and flavanol accumulation. Indeed, genotypes that exhibit redder fruits tend to have lower (8-fold) flavanol contents than non-red fruits, which would be consistent with a competition between two end-products of the flavonoid biosynthetic pathway [79]. Other studies on fruit berries, reported that increased anthocyanin accumulation results in a lower non-anthocyanin phenolic production because of the competition for the same substrates [80, 81]. Moreover, anthocyanin degradation occurs more rapidly in the presence of flavanols and could decrease the chemical stability of flavanol/anthocyanin solutions [82]. Knowing the prevalence of Flesh Browning Disorder (FBD) in red-flesh cultivars [83], decorelation of flavanol and anthocyanin production

could help in breeding red-flesh cultivars less prone to develop FBD.

Interestingly, b^* variation was linked to anthocyanins, flavanols and pH with model coefficients of -0.42 , 0.26 and 0.24 , respectively, highlighting the importance of vacuolar pH in colour expression [84, 85]. The lowest pH values were associated with the lowest b^* values and therefore to a blue colour. At similar anthocyanin levels, a more acidic vacuolar environment could lead to a chromatic shift from red to purple, resulting in a deeper colouration. During fruit development, anthocyanin contents in the flesh decrease [86]. This phenomenon is well known as the dilution process and could lead to a decrease in colour intensity. However, water content does

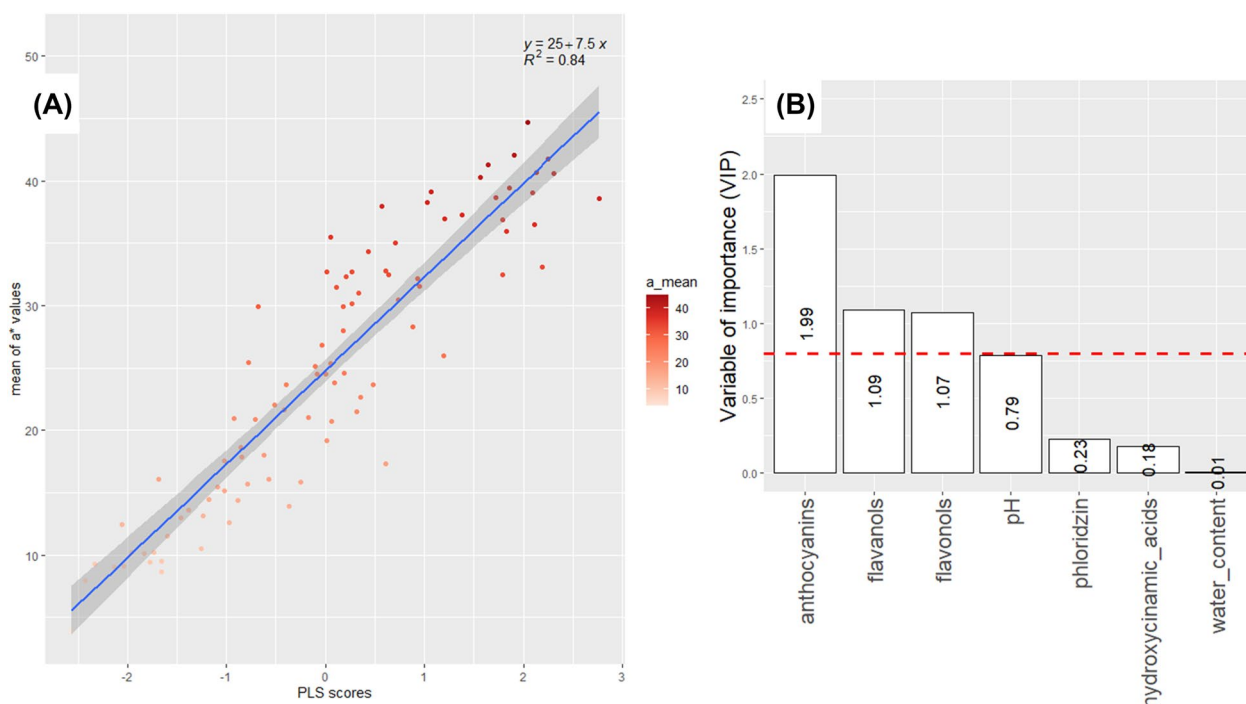


Fig. 7 Modelling of a* by biochemical factors Supervised multivariate analyses were performed using Partial Least Square (PLS) regression. a* values were considered as response variable (Y) and biochemical factors as predictors (X). **A:** Scatter plot of observed a* and predicted values of the PLS model. Dots are coloured according to a* from green (lower values) to red (higher values). **B:** Variable of importance (VIP) scores. A variable with a VIP value above 0.8 is a variable that appears to have a considerable impact in the model

not seem to be involved in red-flesh colour differences among mature fruits when compared at the same developmental stage.

Applications of colour analysis in red-flesh breeding

One of the main challenge in plant breeding is the development of robust phenotyping methods to accurately measure large number of plants in an increasingly constrained environment (e.g. time, cost and data management) [87]. Fruits are plant organs that are prone to damages and biochemical degradation, making fast and portable acquisition systems a necessity to reduce time from harvest to feature acquisition. Our acquisition system provides fast image acquisition (10 s/image) under repeatable lightning conditions (LED). Moreover, this system requires low storage capacity (200 Ko/image, 569 Mo for 3402 images) facilitating a deployment close to field trials. It provides a robust and accurate method for evaluating flesh colour and may be deployed for other species.

The rapid advancements in spectroscopy [88], coupled with the identification of key biochemical factors influencing red-flesh colour in apples, offer the potential to develop models capable of predicting phenolic compound contents. This would circumvent the need

for labor-intensive, expensive, and time-consuming laboratory-based steps. However, characterisation of non-homogeneous surface remains a challenge in spectroscopy due to the limited acquisition area. The integration of hyperspectral imaging [89] and RGB-colour analysis holds promise for assessing the phenolic profile at a pixel scale. Extending this methodology to other F1 families of red-flesh apples will help to better understand the genetic determinism of phenolic profiles and colour expression in other red-flesh pedigree.

Conclusion

In this study, we designed a Python pipeline to extract colorimetric descriptors from RGB images acquired with a flatbed scanner. This pipeline can be easily implemented in more versatile image analysis tools like “PhenoBox” [90]. We confirmed RGB and CIEL*a*b* colour spaces efficiency in discriminating genotypes in comparison with visual notation. Acquisition of biochemical data associated with colour analysis enabled the identification of relevant biochemical factors involved in flesh colour in apple. These analyses revealed that anthocyanin, flavanol and flavanol concentrations, as well as pH are closely related to hue variations, highlighting the multifactorial determinism of the red-flesh trait. Altogether, these

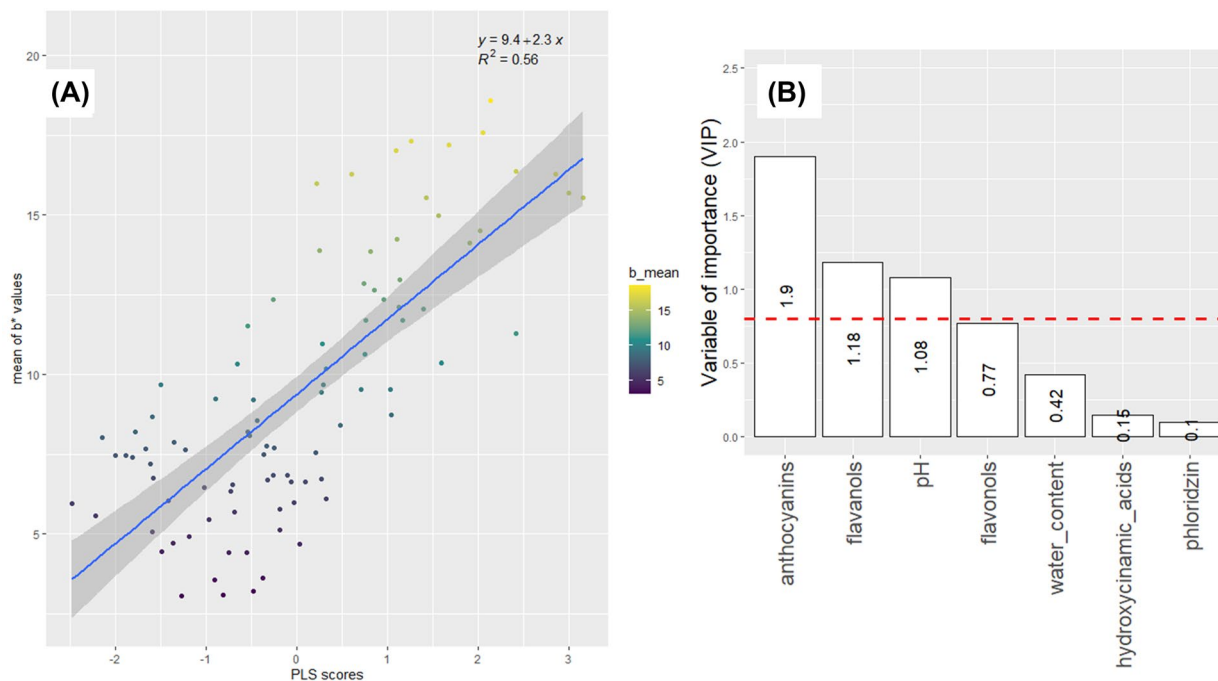


Fig. 8 Modelling of b^* by biochemical factors Supervised multivariate analyses were performed using Partial Least Square (PLS) regression. b^* values were considered as response variables (Y) and biochemical factors as predictors (X). **A:** Scatter plot of observed b^* and predicted values of the PLS model. Dots are coloured according to b^* from blue (lower values) to yellow (higher values). **B:** Variable of importance (VIP) scores. A variable with a VIP value above 0.8 is a variable that appears to have a considerable impact on the model

results will help in deciphering the genetic determinism of red-flesh trait in apple.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01196-1>.

Supplementary file 1.

Acknowledgements

The authors acknowledge PTM PHYTO (SFR QUASAV) for the instruments used for determination of phenolic content.

Author contributions

P. Bouillon wrote the code, analyzed the results and written the manuscript. D. Bréard developed the metabolites extraction and quantification methods. B. Bonnet, A-L. Fanciullino and P. Bouillon prepared samples for UHPLC analysis. S. Hanteville and P. Bouillon carried out fruit harvest, sampling and image analysis. A-L. Fanciullino, E. Belin, S. Boisard, F. Bernard and J-M. Celton provided feedback on the analysis and manuscript and contributed to reviewing and refining the manuscript. J-M Celton and F. Bernard supervised the project.

Funding

This work was financed by ANRT, IFO and IRHS (CIFRE convention No 2021/0182)

Availability of data and materials

The datasets generated and/or analysed during the current study are available in supplementary files (Table S1 and S2). Script designed for image analysis is

public and can be found at: https://github.com/pibouillon/colour_val/blob/main/colour_val.py

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no Competing interest.

Received: 13 July 2023 Accepted: 28 April 2024

Published online: 16 May 2024

References

1. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments anthocyanins, betalains and carotenoids. *Plant J*. 2008;54(4):733–49. <https://doi.org/10.1111/j.1365-3113X.2008.03447.x>.
2. Fairnie ALM, Yeo MTS, Gatti S, Chan E, Travaglia V, Walker JF, Moyroud E. Eco-evo-devo of petal pigmentation patterning. *Essays Biochem*. 2022;66(6):753–68. <https://doi.org/10.1042/ebc20220051>.
3. Steyn WJ, Wand SJE, Holcroft DM, Jacobs G. Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytol*. 2002;155(3):349–61. <https://doi.org/10.1046/j.1469-8137.2002.00482.x>.

4. Liu H, Liu Z, Wu Y, Zheng L, Zhang G. Regulatory mechanisms of anthocyanin biosynthesis in apple and pear. *Int J Mol Sci.* 2021;22(16):8441. <https://doi.org/10.3390/ijms22168441>.
5. Cone KC, Burr FA, Burr B. Molecular analysis of the maize anthocyanin regulatory locus *c1*. *Proc Natl Acad Sci.* 1986;83(24):9631–5. <https://doi.org/10.1073/pnas.83.24.9631>.
6. Holton TA, Cornish EC. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell.* 1995. <https://doi.org/10.1105/tpc.7.7.1071>.
7. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends Plant Sci.* 2015;20(3):176–85. <https://doi.org/10.1016/j.tplants.2014.12.001>.
8. Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell.* 2014;26(3):962–80. <https://doi.org/10.1105/tpc.113.122069>.
9. Zhang Y, Butelli E, Martin C. Engineering anthocyanin biosynthesis in plants. *Current Opin Plant Biol.* 2014;19:81–90. <https://doi.org/10.1016/j.pbi.2014.05.011>.
10. Sun L, Li S, Jiang J, Tang X, Fan X, Zhang Y, Liu J, Liu C. New quantitative trait locus (QTLs) and candidate genes associated with the grape berry color trait identified based on a high-density genetic map. *BMC Plant Biol.* 2020. <https://doi.org/10.1186/s12870-020-02517-x>.
11. Zhao Y, Sun J, Cheronso S, An J-P, Allan AC, Han Y. Colorful hues: insight into the mechanisms of anthocyanin pigmentation in fruit. *Plant Physiol.* 2023;192(3):1718–32. <https://doi.org/10.1093/plphys/kiad160>.
12. Allan AC, Hellens RP, Laing WA. MYB transcription factors that colour our fruit. *Trends Plant Sci.* 2008;13(3):99–102. <https://doi.org/10.1016/j.tplants.2007.11.012>.
13. Treutter D. Biosynthesis of phenolic compounds and its regulation in apple. *Plant Growth Regulation.* 2001;34(1):71–89. <https://doi.org/10.1023/a:1013378702940>.
14. Espley RV, Bovy A, Bava C, Jaeger SR, Tomes S, Norling C, Crawford J, Rowan D, McGhie TK, Brendolise C, Putterill J, Schouten HJ, Hellens RP, Allan AC. Analysis of genetically modified red-fleshed apples reveals effects on growth and consumer attributes. *Plant Biotechnol J.* 2013;11(4):408–19. <https://doi.org/10.1111/pbi.12017>.
15. Wang N, Jiang S, Zhang Z, Fang H, Xu H, Wang Y, Chen X. *Malus sieversii*: the origin, flavonoid synthesis mechanism, and breeding of red-skinned and red-fleshed apples. *Horticul Res.* 2018. <https://doi.org/10.1038/s41438-018-0084-4>.
16. Chen Z, Yu L, Liu W, Zhang J, Wang N, Chen X. Research progress of fruit color development in apple (*malus domestica* borkh). *Plant Physiol Biochem.* 2021;162:267–79. <https://doi.org/10.1016/j.plaphy.2021.02.033>.
17. Cappellini F, Marinelli A, Toccaceli M, Tonelli C, Petroni K. Anthocyanins: from mechanisms of regulation in plants to health benefits in foods. *Front Plant Sci.* 2021;12: 748049. <https://doi.org/10.3389/fpls.2021.748049>.
18. Chagné D, Carlisle CM, Blond C, Volz RK, Whitworth CJ, Oraguzie NC, Crowhurst RN, Allan AC, Espley RV, Hellens RP, Gardiner SE. Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. *BMC Genomics.* 2007;8:212. <https://doi.org/10.1186/1471-2164-8-212>.
19. Chagné D, Lin-Wang K, Espley RV, Volz RK, How NM, Rouse S, Brendolise C, Carlisle CM, Kumar S, De Silva N, Micheletti D, McGhie T, Crowhurst RN, Storey RD, Velasco R, Hellens RP, Gardiner SE, Allan AC. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol.* 2013;161(1):225–39. <https://doi.org/10.1104/pp.112.206771>.
20. Wang N, Genetics Chen X, genomics of fruit color development in apple. *Genetics and genomics of fruit color development in apple.* 2021. https://doi.org/10.1007/978-3-030-74682-7_13.
21. Bars-Cortina D, Macià A, Iglesias I, Romero MP, Motilva MJ. Phytochemical profiles of new red-fleshed apple varieties compared with traditional and new white-fleshed varieties. *J Agric Food Chem.* 2017;65(8):1684–96. <https://doi.org/10.1021/acs.jafc.6b02931>.
22. Ceci AT, Bassi M, Guerra W, Oberhuber M, Robatscher P, Mattivi F, Franceschi P. Metabolomic characterization of commercial, old, and red-fleshed apple varieties. *Metabolites.* 2021;11(6):378. <https://doi.org/10.3390/metabo11060378>.
23. Bars-Cortina D, Macià A, Iglesias I, Garanto X, Badiella L, Motilva M-J. Seasonal variability of the phytochemical composition of new red-fleshed apple varieties compared with traditional and new white-fleshed varieties. *J Agric Food Chem.* 2018;66(38):10011–25. <https://doi.org/10.1021/acs.jafc.8b03950>.
24. Sunil L, Shetty NP. Biosynthesis and regulation of anthocyanin pathway genes. *Appl Microbiol Biotechnol.* 2022;106(5–6):1783–98. <https://doi.org/10.1007/s00253-022-11835-z>.
25. Soto-Vaca A, Gutierrez A, Losso JN, Xu Z, Finley JW. Evolution of phenolic compounds from color and flavor problems to health benefits. *J Agric Food Chem.* 2012;60(27):6658–77. <https://doi.org/10.1021/jf300861c>.
26. Boulton R. The copigmentation of anthocyanins and its role in the color of red wine: a critical review. *Am J Enol Viticul.* 2001;52(2):67–87. <https://doi.org/10.5344/ajev.2001.52.2.67>.
27. Liang Z, Liang H, Guo Y, Yang D. Cyanidin 3-o-galactoside: a natural compound with multiple health benefits. *Int J Mol Sci.* 2021;22(5):2261. <https://doi.org/10.3390/ijms22052261>.
28. McCallum S, Woodhead M, Hackett CA, Kassim A, Paterson A, Graham J. Genetic and environmental effects influencing fruit colour and QTL analysis in raspberry. *Theor Appl Genet.* 2010;121(4):611–27. <https://doi.org/10.1007/s00122-010-1334-5>.
29. Calle A, Serradilla MJ, Wunsch A. QTL mapping of phenolic compounds and fruit colour in sweet cherry using a 6+9K SNP array genetic map. *Scientia Horticulturae.* 2021;280: 109900. <https://doi.org/10.1016/j.scienta.2021.109900>.
30. Shibaya T, Kuroda C, Tsuruoka H, Minami C, Obara A, Nakayama S, Kishida Y, Fujii T, Isobe S. Identification of QTLs for root color and carotenoid contents in Japanese orange carrot f2 populations. *Sci Rep.* 2022. <https://doi.org/10.1038/s41598-022-11544-7>.
31. Keyser ED, Lootens P, Bockstaele EV, Riek JD. Image analysis for QTL mapping of flower colour and leaf characteristics in pot azalea (*Rhododendron simsii* hybrids). *Euphytica.* 2012;189(3):445–60. <https://doi.org/10.1007/s10681-012-0809-7>.
32. Zhang S, Hu X, Miao H, Chu Y, Cui F, Yang W, Xu S, Guo J, Fu C, Song X, Hou M, Qiu J, Chen J. Imaged-based phenotyping accelerated QTL mapping and qtl x environment interaction analysis of testa colour in peanut (*Arachis hypogaea*). *Plant Breeding.* 2021;140(5):884–95. <https://doi.org/10.1111/pbr.12905>.
33. Zingaretti LM, Monfort A, Pérez-Enciso M. Automatic fruit morphology phenome and genetic analysis: an application in the octoploid strawberry. *Plant Phenomics.* 2021;2021:2021. <https://doi.org/10.4133/2021/9812910>.
34. Denoyes B, Prohaska A, Petit J, Rothan C. Deciphering the genetic architecture of fruit color in strawberry. *J Experim Botany.* 2023. <https://doi.org/10.1093/jxb/erad245>.
35. Li M, Frank MH, Migicovsky Z. Colourquant: a high-throughput technique to extract and quantify color phenotypes from plant images. *Plant Phenotyping Methods Protocols.* 2022. https://doi.org/10.1007/978-1-0716-2537-8_9.
36. Pathare PB, Opara UL, Al-Said FA-J. Colour measurement and analysis in fresh and processed foods: a review. *Food Bioprocess Technol.* 2012;6(1):36–60. <https://doi.org/10.1007/s11947-012-0867-9>.
37. Wu D, Sun D-W. Colour measurements by computer vision for food quality control - a review. *Trends Food Sci Technol.* 2013;29(1):5–20. <https://doi.org/10.1016/j.tifs.2012.08.004>.
38. Underhill AN, Hirsch CD, Clark MD. Evaluating and mapping grape color using image-based phenotyping. *Plant Phenomics.* 2020. <https://doi.org/10.34133/2020/8086309>.
39. Yu Y, Yang Z, Jiang Y, Wang L, Wu Y, Liao J, Yang R, Zhang L. Inheritance and QTL mapping for flower color in *Isalvia miltiorrhiza*/i bunge. *J Heredity.* 2022;113(3):248–56. <https://doi.org/10.1093/jhered/esac012>.
40. Li M, Coneva V, Robbins KR, Clark D, Chitwood D, Frank M. Quantitative dissection of color patterning in the foliar ornamental coleus. *Plant Physiol.* 2021;187(3):1310–24. <https://doi.org/10.1093/plphys/kiab393>.
41. Serouart M, Madec S, David E, Velumani K, Lozano RL, Weiss M, SegVeg Baret F. Segmenting RGB images into green and senescent vegetation by combining deep and shallow methods. *Plant Phenomics.* 2022. <https://doi.org/10.34133/2022/9803570>.
42. Kasajima I. Measuring plant colors. *Plant Biotechnol.* 2019;36(2):63–75. <https://doi.org/10.5511/plantbiotechnology.19.0322a>.
43. Bonora A, Bertolotti G, Bresilla K, Grappadelli LC, Manfrini L. A convolutional neural network approach to detecting fruit physiological

- disorders and maturity in 'abbé fétel' pears. *Biosyst Eng.* 2021;212:264–72. <https://doi.org/10.1016/j.biosystemseng.2021.10.009>.
44. PalSinghChauhan A, Partap Singh A. Intelligent estimator for assessing apple fruit quality. *IJCA.* 2012;60(5):35–41. <https://doi.org/10.5120/9691-4130>.
 45. Suresha S, Soumya. Apples grading based on SVM classifier. *Int J Computer Appl.* 2012;975:8878.
 46. Xiaobo Z, Jiewen Z, Yanxiao L. Apple color grading based on organization feature parameters. *Pattern Recogn Lett.* 2007;28(15):2046–53. <https://doi.org/10.1016/j.patrec.2007.06.001>.
 47. Cárdenas-Pérez S, Chanona-Pérez J, Méndez-Méndez JV, Calderón-Domínguez G, López-Santiago R, Perea-Flores MJ, Arzate-Vázquez I. Evaluation of the ripening stages of apple (Golden Delicious) by means of computer vision system. *Biosyst Eng.* 2017;159:46–58. <https://doi.org/10.1016/j.biosystemseng.2017.04.009>.
 48. Garrido-Novell C, Pérez-Marín D, Amigo JM, Fernández-Novales J, Guerrero JE, Garrido-Varo A. Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras. *J Food Eng.* 2012;113(2):281–8. <https://doi.org/10.1016/j.jfoodeng.2012.05.038>.
 49. Leemans V, Destain M-F. A real-time grading method of apples based on features extracted from defects. *J Food Eng.* 2004;61(1):83–9. [https://doi.org/10.1016/S0260-8774\(03\)00189-4](https://doi.org/10.1016/S0260-8774(03)00189-4).
 50. Xiao-bo Z, Jie-wen Z, Yanxiao L, Holmes M. In-line detection of apple defects using three color cameras system. *Computers Electron Agric.* 2010;70(1):129–34. <https://doi.org/10.1016/j.compag.2009.09.014>.
 51. Shimizu T, Okada K, Moriya S, Komori S, Abe K. A high-throughput color measurement system for evaluating flesh browning in apples. *J Amer Soc Hort Sci.* 2021;146(4):241–51. <https://doi.org/10.21273/JASHS05027-20>.
 52. Subhashree SN, Sunoj S, Xue J, Bora GC. Quantification of browning in apples using colour and textural features by image analysis. *Food Quality Safety.* 2017;1(3):221–6. <https://doi.org/10.1093/fqsafe/fyx021>.
 53. Wellner A, Grimm E, Knoche M. Effect of orchard management factors on flesh color of two red-fleshed apple clones. *Horticulturae.* 2019;5(3):54. <https://doi.org/10.3390/horticulturae5030054>.
 54. Kumar S, Deng CH, Molloy C, Kirk C, Plunkett B, Lin-Wang K, Allan A, Espley R, Kumar S, Deng CH, Molloy C, Kirk C, Plunkett B, Lin-Wang K, Allan A, Espley R. Extreme-phenotype GWAS unravels a complex nexus between apple (*Malus domestica*) red-flesh colour and internal flesh browning. *Fruit Res.* 2022;2(1):1–14. <https://doi.org/10.4130/FruRes-2022-0012>.
 55. Grimm E, Kuhnke F, Gajdt A, Ostermann J, Knoche M. Accurate quantification of anthocyanin in red flesh apples using digital photography and image analysis. *Horticulturae.* 2022;8(2):145. <https://doi.org/10.3390/horticulturae8020145>.
 56. Ye X, Ishioka S, Zhang S. Estimation of the degree of red coloration in flesh of a red-fleshed apple cultivar 'Kurenai no Yume' with a UV-vis-NIR intercan device. *Postharv Biol Technol.* 2017;124:128–36. <https://doi.org/10.1016/j.postharvbio.2016.09.005>.
 57. Ye X, Doi T, Arakawa O, Zhang S. A novel spatially resolved intercan spectroscopy system to estimate degree of red coloration in red-fleshed apple. *Sci Rep.* 2021;11(1):21982. <https://doi.org/10.1038/s41598-021-01468-z>.
 58. Reynolds D, Baret F, Welcker C, Bostrom A, Ball J, Cellini F, Lorence A, Chawade A, Khaffi M, Noshita K, Mueller-Linow M, Zhou J, Tardieu F. What is cost-efficient phenotyping? optimizing costs for different scenarios. *Plant Sci.* 2019;282:14–22. <https://doi.org/10.1016/j.plantsci.2018.06.015>.
 59. Rossi R, Leolini C, Costafreda-Aumedes S, Leolini L, Bindi M, Zaldei A, Moriondo M. Performances evaluation of a low-cost platform for high-resolution plant phenotyping. *Sensors.* 2020. <https://doi.org/10.3390/s20113150>.
 60. Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. scikit-image: image processing in python. *Peer J.* 2014;2:453.
 61. Chacon WDC, dos Santos Alves MJ, Monteiro AR, González SYG, Valencia GA. Image analysis applied to control postharvest maturity of papayas (*icaria* papaya/1 l). *J Food Process Preserv.* 2022. <https://doi.org/10.1111/jfpp.16999>.
 62. ...Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. *Nature.* 2020;585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2>.
 63. Malec M, Le Quééré J-M, Sotin H, Kolodziejczyk K, Bauduin R, Guyot S. Polyphenol profiling of a red-fleshed apple cultivar and evaluation of the color extractability and stability in the juice. *J Agric Food Chem.* 2014;62(29):6944–54. <https://doi.org/10.1021/jf500336v>.
 64. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol.* 2014;48(4):2097–8. <https://doi.org/10.1021/es5002105>.
 65. Xu S. Quantitative genetics. 2022. <https://doi.org/10.1007/978-3-030-83940-6>.
 66. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab Syst.* 2001;58(2):109–30. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
 67. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. *Chemometr Intell Lab Syst.* 2005;78(1–2):103–12. <https://doi.org/10.1016/j.chemolab.2004.12.011>.
 68. R Core Team: R: A language and environment for statistical computing (2021). R Foundation for Statistical Computing.
 69. Lê S, Josse J, Husson F. FactoMineR: a package for multivariate analysis. *J Stat Softw.* 2008;25(1):1–18. <https://doi.org/10.18637/jss.v025.i01>.
 70. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res.* 2015;14(8):3322–35. <https://doi.org/10.1021/acs.jproteome.5b00354>.
 71. Wickham H. ggplot2: Elegant graphics for data analysis 2016
 72. McLellan LR, Lind LR, Kime RW. Hue angle determinations and statistical analysis for multiquadrant hunter L, a, b data. *J Food Quality.* 1995;18(3):235–40. <https://doi.org/10.1111/j.1745-4557.1995.tb00377.x>.
 73. Farneti B, Masuero D, Costa F, Magnago P, Malnoy M, Costa G, Vrhovsek U, Mattivi F. Is there room for improving the nutraceutical composition of apple? *J Agric Food Chem.* 2015;63(10):2750–9. <https://doi.org/10.1021/acs.jafc.5b00291>.
 74. Verdu CF, Childebrand N, Marnet N, Lebal G, Dupuis F, Laurens F, Guillet D, Guyot S. Polyphenol variability in the fruits and juices of a cider apple progeny. *J Sci Food Agric.* 2013;94(7):1305–14. <https://doi.org/10.1002/jsfa.6411>.
 75. Guyot S, Marnet N, Sanoner P, Drilleau J-F. Variability of the polyphenolic composition of cider apple (*malus domestica*) fruits and juices. *J Agric Food Chem.* 2003;51(21):6240–7. <https://doi.org/10.1021/jf0301798>.
 76. Ockermann P, Headley L, Lizio R, Hansmann J. A review of the properties of anthocyanins and their influence on factors affecting cardiometabolic and cognitive health. *Nutrients.* 2021;13(8):2831. <https://doi.org/10.3390/nu13082831>.
 77. Ibraheem N, Hasan M, Khan RZ, Mishra P. Understanding color models: a review. *ARPN J Sci Technol.* 2012;2(3):265.
 78. Wang W, Moss SMA, Zeng L, Espley RV, Wang T, Lin-Wang K, Fu B, Schwinn KE, Allan AC, Yin X. The red flesh of kiwifruit is differentially controlled by specific activation-repression systems. *New Phytol.* 2022;235(2):630–45. <https://doi.org/10.1111/nph.18122>.
 79. Han Y, Vimolmangkang S, Soria-Guerra RE, Korban SS. Introduction of apple ANR genes into tobacco inhibits expression of both CHI and DFR genes in flowers, leading to loss of anthocyanin. *J Experim Botany.* 2012;63(7):2437–47. <https://doi.org/10.1093/jxb/err415>.
 80. Salvatierra A, Pimentel P, Moya-León MA, Herrera R. Increased accumulation of anthocyanins in fragaria chiloensis fruits by transient suppression of FcMYB1 gene. *Phytochemistry.* 2013;90:25–36. <https://doi.org/10.1016/j.phytochem.2013.02.016>.
 81. Xie S, Qiao X, Chen H, Nan H, Zhang Z. Coordinated regulation of grape berry flesh color by transcriptional activators and repressors. *J Agric Food Chem.* 2019;67(42):11815–24. <https://doi.org/10.1021/acs.jafc.9b05234>.
 82. Gordillo B, Rodríguez-Pulido FJ, González-Miret ML, Quijada-Morín N, Rivas-Gonzalo JC, García-Estévez I, Heredia FJ, Escribano-Bailón MT. Application of differential colorimetry to evaluate anthocyanin–flavonol–flavanol ternary copigmentation interactions in model solutions. *J Agric Food Chem.* 2015;63(35):7645–53. <https://doi.org/10.1021/acs.jafc.5b00181>.
 83. Espley RV, Leif D, Plunkett B, McGhie T, Henry-Kirk R, Hall M, Johnston JW, Punter MP, Boldingh H, Nardoza S, Volz RK, O'Donnell S, Allan AC. Red

- to brown: an elevated anthocyanic response in apple drives ethylene to advance maturity and fruit flesh browning. *Front Plant Sci.* 2019. <https://doi.org/10.3389/fpls.2019.01248>.
84. Khoo HE, Azlan A, Tang ST, Lim SM. Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr Res.* 2017;61(1):1361779. <https://doi.org/10.1080/16546628.2017.1361779>.
 85. Chen C-C, Lin C, Chen M-H, Chiang P-Y. Stability and quality of anthocyanin in purple sweet potato extracts. *Foods.* 2019;8(9):393. <https://doi.org/10.3390/foods8090393>.
 86. Zhang X, Xu J, Xu Z, Sun X, Zhu J, Zhang Y. Analysis of antioxidant activity and flavonoids metabolites in peel and flesh of red-fleshed apple varieties. *Molecules.* 2020;25(8):1968. <https://doi.org/10.3390/molecules25081968>.
 87. Fiorani F, Schurr U. Future scenarios for plant phenotyping. *Ann Rev Plant Biol.* 2013;64(1):267–91. <https://doi.org/10.1146/annurev-arplant-050312-120137>.
 88. Manzoor MF, Hussain A, Naumovski N, Ranjha MMAN, Ahmad N, Karrar E, Xu B, Ibrahim SA. A narrative review of recent advances in rapid assessment of anthocyanins in agricultural and food products. *Front Nutr.* 2022. <https://doi.org/10.3389/fnut.2022.901342>.
 89. Chen S, Zhang F, Ning J, Liu X, Zhang Z, Yang S. Predicting the anthocyanin content of wine grapes by nir hyperspectral imaging. *Food Chem.* 2015;172:788–93. <https://doi.org/10.1016/j.foodchem.2014.09.119>.
 90. Czedik-Eysenberg A, Seitner S, Güldener U, Koemeda S, Jez J, Colombini M, Djamei A. The 'phenobox', a flexible, automated, open-source plant phenotyping solution. *New Phytol.* 2018;219(2):808–23. <https://doi.org/10.1111/nph.15129>.
 91. Hyson DA. A comprehensive review of apples and apple components and their relationship to human health. *Advan Nutr.* 2011;2(5):408–20. <https://doi.org/10.3945/an.111.000513>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.