

SOFTWARE

Open Access



BerryPortraits: Phenotyping Of Ripening Traits in cranberry (*Vaccinium macrocarpon* Ait.) with YOLOv8

Jenyne Loarca^{1,2*} , Tyr Wiesner-Hanks^{3*} , Hector Lopez-Moreno^{1,2} , Andrew F. Maule^{1,2} , Michael Liou⁴ , Maria Alejandra Torres-Meraz^{1,2}, Luis Diaz-Garcia^{1,2,5} , Jennifer Johnson-Cicalese⁶, Jeffrey Neyhart^{6,7} , James Polashock^{6,7}, Gina M. Sideli^{6,8} , Christopher F. Strock³ , Craig T. Beil^{3*} , Moira J. Sheehan^{3*†} , Massimo Iorizzo^{9,10} , Amaya Atucha^{1†} and Juan Zalapa^{1,2*†}

Abstract

BerryPortraits (Phenotyping of Ripening Traits) is open source Python-based image-analysis software that rapidly detects and segments berries and extracts morphometric data on fruit quality traits such as berry color, size, shape, and uniformity. Utilizing the YOLOv8 framework and community-developed, actively-maintained Python libraries such as OpenCV, BerryPortraits software was trained on 512 postharvest images (taken under controlled lighting conditions) of phenotypically diverse cranberry populations (*Vaccinium macrocarpon* Ait.) from the two largest public cranberry breeding programs in the U.S. The implementation of CIELAB, an intuitive and perceptually uniform color space, enables differentiation between berry color and berry brightness, which are confounded in classic RGB color channel measurements. Furthermore, computer vision enables precise and quantifiable color phenotyping, thus facilitating inclusion of researchers and data analysts with color vision deficiency. BerryPortraits is a phenotyping tool for researchers in plant breeding, plant genetics, horticulture, food science, plant physiology, plant pathology, and related fields. BerryPortraits has strong potential applications for other specialty crops such as blueberry, lingonberry, caneberry, grape, and more. As an open source phenotyping tool based on widely-used python libraries, BerryPortraits allows anyone to use, fork, modify, optimize, and embed this software into other tools or pipelines.

Keywords Computer vision, Digital phenotyping, Image-based phenotyping, Image segmentation, Plant breeding, Pomology, Fruit quality

[†]Moira J. Sheehan, Amaya Atucha, and Juan Zalapa authors share senior authorship.

*Correspondence:

Jenyne Loarca

jloarca@wisc.edu

Tyr Wiesner-Hanks

tw372@cornell.edu

Craig T. Beil

ctb94@cornell.edu

Moira J. Sheehan

moirasheehan@cornell.edu

Juan Zalapa

juan.zalapa@usda.gov; jezalapa@wisc.edu

Full list of author information is available at the end of the article

Background

The evolution of human response to visual cues is an outcome of a commensurate evolutionary relationship between humans, seeking sustenance, and fruit, whose seeds are subsequently dispersed [1]. Whether due to natural selection or human-mediated selection, the plethora of fruit forms and hues are an avid area of research by horticulturists, plant geneticists, plant breeders, and food scientists [2]. However, measuring fruit characteristics in a high throughput manner has long been recognized as a major bottleneck in crop research [3]. For example, traditional measures of fruit color rely on categorical color



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024, corrected publication 2025. **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/>.

bins, established by eye, in specialty crops such as plum [4], apricot [5], and eggplant [6]. Furthermore, many of the fruit quality traits that contribute to the hedonic human experience of fruit, such as color, are difficult to measure precisely. Genetic variation in photoreceptors among humans results in color perceptual differences; even humans without color deficient vision vary considerably in perceptual judgments of color due to variations in rod/cone receptors [7]. Thus, color estimations that rely on human perception reduce data accuracy, especially when differentiating between similar hues. In berry crops such as cranberry, grape, lingonberry, or blueberry, fruit color can be quantified with spectrophotometric measurements of the berry's total anthocyanin (TAcY) pigments that humans perceive as red, purple, or blue, respectively. TAcY quantity is a continuous trait and permits finer color granularity than categorical phenotyping, but in cranberry and other berry crops, such assays are costly, laborious, destructive to fruit samples, and require specialized equipment and training [8, 9]. While fruit size and shape characteristics in berry crops such as cranberry, are generally less subjective attributes to quantify, these traits are also tedious, time-consuming, laborious, and costly to measure manually [10]. Traditionally, fruit length and width are measured by hand with calipers, and size metrics, such as surface area and volume, are calculated from these measurements [11]. This process requires substantial time and labor, especially when a large number of fruits are required to accurately characterize one population [12, 13].

The development of automated, high throughput image-based methods for crop phenotyping was motivated by the substantial cost, time, tedious, and difficulty of hand-measuring, estimating, or assaying these important quantitative traits [14, 15].

Computer vision has been implemented for phenotyping visual aspects of quality, ripeness, and harvestability with varying levels of throughput and accuracy in horticultural crops (Supp. table 1). Field-based imaging, or imaging fruit on the plants under natural lighting conditions, has been developed under various crop production environment (open field, greenhouse, orchard, vineyard) in a variety of specialty crops, such as, pumpkin [16], tomato [17, 18], pepper [19], and including perennial fruit crops such as citrus [20], quince and raspberry [21], blueberry [22–24], grape [25–28], myriad other horticultural specialty crops [29]. Comparison among these methods has recently been reviewed [30].

High throughput phenotyping methods are critical tools to advance research with accuracy and efficiency, and suitability of the platform varies depending on the crop, trait, and imaging environment [30]. Imaging fruit under natural lighting conditions can reduce color image

quality due to variable daylight conditions. Fruit size distortion can occur due to the variable distances between the camera lens and fruits on the plant. Similarly, angular distortion can alter perception of fruit shape for non-spherical fruit, which occurs with variable angles of fruit on the plant relative to the lens of the camera. Finally, in the field, plant architecture and plant foliage may obscure imaging of the fruit, resulting in information loss due to occlusion [31]—at the time of publication, field imaging methods are still under development to impute or predict characteristics of obscured fruit in strawberry [31].

Postharvest fruit imaging in the laboratory mitigates all of the above issues by creating an environment with standardized lighting conditions, enabling repeatable and highly precise measurement of fruit quality traits. These types of standard conditions have been used in digital phenotyping pipelines for lettuce [33], carrot root and shoot [34, 35], tomato, winter squash [36], and in a variety of other specialty crops [37], including perennial fruit crops such as apple [38–40], cranberry [41], pear [40], blackberry [42], raspberry [43], strawberry [44–48], blueberry [49–52], and grape [27, 53]. Imaging on a stage, with all fruit at a fixed distance from the camera lens, eliminates fruit size distortion, while standard berry orientation eliminates fruit shape angular distortion.

Though there are numerous tools for other crops, there are very limited options for specialty crops such as cranberries. GiNA was the first high throughput phenotyping software program developed for postharvest imaging of cranberries under controlled lighting conditions [41]. This software was validated by measuring cranberries with a hand-held caliper and demonstrated very high correlations ($0.92 \leq r \leq 0.97$) between hand measurement and GiNA's image-based measurement. GiNA has been proven effective for accurately phenotyping fruit color and has been leveraged as a high throughput image-based phenotyping tool in cranberry [8, 10], blueberry [54, 55], *Silphium integrifolium* [56], and kura clover (*Trifolium ambiguum*) [57].

Despite a track record of successful utilization, GiNA has high barriers to entry for users. GiNA runs on MATLAB (highly specialized and costly proprietary software), is unmaintained, and underperforms when cranberry shape deviates from a perfect ellipse, such as with pyriform, spindle, or ovate berries [58]. GiNA was validated on a small sample of fruit from one uniformly dark red commercial cultivar, which means that the software's performance has not been tested on lighter berry colors that are abundant in diverse germplasm (e.g., green, yellow, white, and pink berries). These drawbacks required the creation of open source, user-friendly berry phenotyping software and subsequent validation on a larger and more diverse cranberry population.

Here we present BerryPortraits: Phenotyping Of Ripening Traits [‘with Rapid Automated Imaging Tools and Software,’ for those disinclined towards brevity]) (<https://github.com/Breeding-Insight/BerryPortraits/>). This project is a collaboration with Breeding Insight, a USDA-ARS funded effort that works with identified speciality crops, providing support and development of phenotyping and genotyping tools. As the software name suggests, *BerryPortraits* measures the visual aspects of fruit quality and ripeness—berry color, size, shape, and uniformity—that are targets for selection in cranberry breeding programs. As a scalable, affordable, and high throughput phenotyping pipeline, BerryPortraits can extract berry data from images for color, size, and shape traits measured by GiNA. Trait accuracy can be measured by comparing BerryPortraits performance to its validated predecessor, GiNA. We also introduce BerryPortraits’s additional functionality of the CIELAB color space and improved calculations for measuring size characteristics such as volume and surface area. To improve accessibility and lower the barrier to entry for new users, BerryPortraits is open-source and was developed in Python, which is widely used for data science and has strong support with an active community of users troubleshooting problems in web-based forums. This pipeline also has the capacity to integrate with BreedBase [59] and the Android app Field Book [60], enabling precise traceability of phenotypic data and reducing probability of data loss or collection errors. With improved accessibility, affordability, throughput, and accuracy, BerryPortraits has innumerable possible creative applications for researchers in horticulture, plant genetics, plant breeding, and food science.

Materials & methods

Plant materials

The cranberry population under study, CNJ02-1, is the most well-studied diploid cranberry population available for genetic research [8, 9, 61–65]. This population was developed at Rutgers University, Chatsworth, NJ, by Nicholi Vorsa and consists of 171 cranberry genotypes representing a wide range of phenotypic variation in external fruit quality characteristics. Propagules from each genotype were clonally propagated and established at a commercial cranberry farm in Necedah, Wisconsin. Each genotype was grown in a distinct 2.25 m² plot [61, 62] in beds that were professionally and uniformly managed. In September 2019, 30 mature cranberry fruit were randomly harvested from each plot and placed in mesh bags. Samples were placed in a styrofoam cooler and transported to the USDA-ARS Vegetable Crops Research Unit at the University of Wisconsin-Madison. Berry samples were maintained at 10 °C until ready for imaging.

Berry image acquisition

Berries were photographed in an enclosed light box for controlled and standard lighting conditions (Supp. Figure 1). Images were taken with a DSLR camera (Canon EOS REBEL T6; focal length = 28 mm; F-stop = 11; exposure length = 1/20 s) with lens positioned directly over the fruit, approximately 70 cm above the stage. A sample of 30 cranberries from each plot were arranged on the stage in rows and columns of 5 × 6 berries (Supp. Figure 2) (this feature can be enabled or disabled in BerryPortraits). The stage background was white and included six 2.54-cm diameter black scale markers on either side of the fruit. One image was taken per 30-fruit sample for 171 unique genotypes (25–30 cranberries per image; image dimensions = 5184 × 3456 pixels in JPG format). Each image was evaluated in GiNA and in BerryPortraits for comparison of berry size, berry shape, and berry color values. Photo sample files were manually labeled with genotype ID.

For the purposes of training the instance segmentation model that underlies BerryPortraits, two other image sets were used in order to make the model more robust against variations in lighting, background, layout, and any non-target objects present in the image. The first set consisted of 281 images captured in the same manner described above, with three core differences: the image background was blue instead of white to increase contrast between the background and unripe berries; each image contained an X-Rite ColorChecker Classic card and a small human-readable card; and only six size markers were used instead of twelve. The second set consisted of 60 images (25–50 berries per image) captured on white or speckled gray-black divot trays with a wide range of berry sizes, image layouts, and presence and position of size and/or color markers (personal communication—Jeff Neyhart). For representative examples of each image type, see Supp. Figure 3.

Running BerryPortraits

The core output of BerryPortraits is a.csv file with metrics for each detected berry in each input image, similar to the output of GiNA. If berries are arranged into rows and columns, BerryPortraits can auto-detect the number of rows, assign berries a row and column, and organize the output by row/column. If berries are not organized, the output values can be tied to a given berry in the image using the pixel coordinates of the berry centroid.

For visualization and quality control, BerryPortraits can save annotated versions of each input image showing the berry mask boundaries, size marker mask boundaries, and oriented bounding boxes (OBBs) around each berry. Examining these annotated images by hand is a fast

way to check that the model is detecting and segmenting objects correctly, and we recommend users do so as a matter of course. Several output values are checked during runtime as a basic QC. Any berries with solidity < 0.95 (see below for explanation) are flagged and reported, as these are typically berries with incorrect segmentation. Any berries with an area > 3 standard deviations from the mean area of the image are similarly flagged.

BerryPortraits object identification and segmentation

The first step of the BerryPortraits analysis pipeline is the detection and segmentation of berries and size markers. The default method uses YOLOv8 [66], a widely-used state of the art (SOTA) model (at the time of publication, cited more than 12,700 times) for detection and instance segmentation of four object classes: cranberries, black circular size markers (2.54-cm diameter in the image sets here), sample labels (black text on white cards), and X-Rite ColorChecker Classic color calibration cards. The YOLOv8 architecture was selected because it performed well on validation images and it offered pre-trained instance segmentation models, unlike pure detection models such as YOLOv9 [67], YOLOv10 [68], or YOLO-NAS [69], which must be combined with downstream segmentation methods. This single-step instance segmentation greatly simplified training the model, deploying the script, and retraining the model on additional data. Meta's Segment Anything Model (SAM) [70], a general-purpose instance segmentation model, was tested during development and performed well, but the large sizes of SAM model files (358 MB to 2.56 GB) made it a less attractive choice than YOLOv8 in terms of distributing this tool.

The dataset used for training, validation, and testing consisted of 512 images: 171 white-background images and 60 variable-layout images (described above under *Image Acquisition*), and 281 blue-background images (25 berries per image). Images were annotated with polygon borders around objects in the four target classes and partitioned into a 70/20/10 split between training/validation/test. A pretrained YOLOv8s-seg base model was then trained for 100 epochs on an NVidia A100 with single-image batch size at an image resolution of 3200 using hyperparameters set automatically by the Ultralytics library based on the image and annotation metadata.

We also included a traditional computer vision method for detecting and segmenting cranberries and size markers using OpenCV. This approach uses several static thresholding steps on the hue, saturation, and value channels and several. This method is more sensitive to non-target objects and variation in background or object layout, but is extremely fast and reliable for images laid

out on a clean background in a consistent manner with no non-target objects.

Because GiNA does not organize berries into rows and columns, we used the pixel coordinates of berry centroids to match berries 1-to-1 in order to compare the size and color metrics reported below between GiNA and BerryPortraits. For a given image, the Euclidean distance between each pair of berry centroids from GiNA and BerryPortraits were calculated and berries were considered the same if they were the reciprocally closest berries to one another.

BerryPortraits size and shape metrics

Size metrics of detected berries are calculated from the segmentation mask around the berry and are reported in pixels by default or are dimensionless. If the user has included circular size markers of a known diameter, this is used to adjust size metrics to cm. *Length* and *width* of each berry are defined by the length and width of the oriented bounding box (OBB) surrounding the berry, the direction of which is taken from the primary axis of the ellipse of best fit, thus berries do not have to be oriented before measuring (Table 1). This method attempts to match the length and width measurement that would be taken with a hand-held caliper.

Perimeter and *area* are defined as the arc length and area of the segmentation mask. *Solidity* is defined as the ratio of the area of the segmentation mask to the area of the convex hull around the mask. *Roundness* is defined using the ISO 110 definition: the ratio of the radius of the largest circle inscribed by the segmentation mask to that of the smallest circle inscribing the mask. In contrast to GiNA, BerryPortraits estimates surface area and volume by modeling the berry as a series of stacked cylinders. Volume is approximated by modeling the berry as a series of stacked cylinders with diameter equal to the segmentation mask width and a height of 1 pixel. Similarly, surface area is approximated by modeling the berry as a series of stacked frustums (truncated cones) with upper and lower diameters taken from the segmentation mask at pixel height h and height $h + 1$ (Supp. Figure 4). Testing of these traits on ellipsoids drawn at different pixel resolutions found the volume estimation to be within 0.05% of the true value and the surface area approximation to be within 5% of the standard approximation (as the surface area of an ellipsoid has no closed-form solution).

We identified formulaic discrepancies in GiNA that inexplicably or erroneously multiplied surface area and volume values by a factor of 4 and 50, respectively. For our analyses, we performed a transformation on the GiNA output to enable comparison. The formulae used for each platform are reported (Table 1).

Table 1 Shape and size formulas and descriptions used by GiNA vs. BerryPortraits

Trait	GiNA description/formula	BerryPortraits description/formula
Length	Length of the major axis that has the same normalized second central moments as the region. It is always larger than the width, therefore this parameter will be con-founded in objects that are wider than they are long	Length of the oriented bounding box (OBB) around the berry segmentation mask, ori-ented along the primary/major axis of the ellipse of best fit
Width	Length of the minor axis that has the same normalized second central moments as the region. It is always shorter than the length	Width of the OBB around the berry segmentation mask, orthogonal to the primary/major axis
Area	Number of squared units in the recognized object. The same pixels counter in this parameters are used to determine color	Area of the berry segmentation mask
Surface area	Estimation of the surface area assuming that the object is spherical or egg shaped $2\pi a^2 + \mu a * 2(\frac{b^2}{\sqrt{b^2 - a^2}} \frac{a}{b})$ where a and b are 0.5*width and 0.5*length, respectively Note: GiNA software inexplicably or erroneously multiplied this formula by a factor of 50	Approximated as the sum of lateral areas of a series of stacked truncated cones, each with an axis 1 pixel high on the major axis of the OBB and base/top widths equal to the width of the berry mask, orthogonal to the major axis, at given pixel heights. Base and top area are added only for the first and last truncated cone: $\pi[(\frac{w_1}{2})^2 + (\frac{w_h}{2})^2 + \sum_{i=1}^{h-1}(\frac{w_i}{2} + \frac{w_{i+1}}{2})\sqrt{(\frac{w_i}{2} - \frac{w_{i+1}}{2})^2 + 1}]$ where w_1 = width at the mask base, w_h = width at the top of the mask, w_i = mask width at a given pixel height i , and h = total height in pixels of the mask
Perimeter	Distance around the boundary region of the object. It is computed by calculating the distance between each adjoining pair of pixels around the border of the region	Arc length of the segmentation mask
Volume	Estimation of the object volume assuming that the object is spherical or egg shaped $\frac{4\pi}{3} a^2 (a + b)$, where a and b are 0.5*width and 0.5*length, respectively Note: GiNA software inexplicably or erroneously multiplied this formula by a factor of 4	Approximated as the sum of volumes of a series of stacked cylinders, each with an axis 1 pixel high on the major axis of the OBB and a width equal to that of the berry mask, orthogonal to the major axis, at a given pixel height: $\sum_{i=1}^{h-1} (\frac{w_i}{2})^2$, where h = total height in pixels of the mask and w_i = width in pixels of the mask at pixel height i
Shape/Roundness	Describe how round is the object. A value of 1 represents a circle-like object Length: width	ISO 110 Roundness definition: $\frac{r_{inscribed}}{r_{enclosing}}$ Where $r_{inscribed}$ = radius of the largest circle inscribed by the berry segmentation mask and $r_{enclosing}$ = radius of the smallest circle circumscribing the berry mask
Solidity	-	$\frac{A_{mask}}{A_{hull}}$ Where A_{mask} = area of the berry mask and A_{hull} = area of the convex hull circumscribing the mask

GiNA description column reprinted with permission from Table S1, Diaz-Garcia et al. [41]; Description of the parameters generated by GiNA. Mathematical formulae and description of parameters returned by the software

BerryPortraits color metrics

As illumination and camera settings can differ between image sets, BerryPortraits uses the ColorCorrectionML library (https://github.com/collinswakholi/ML_Color_Correction_tool) to perform post hoc color correction using linear correction, least-squares regression, or partial least-squares regression. At the time of writing, ColorCorrectionML can correct images captured with a standard D50 illuminant and that contain an X-Rite ColorChecker Classic. Color correction is disabled by default.

Berries with considerable epicuticular waxes tend to strongly reflect the background surface along the edge of the berry, distorting color metrics. To mitigate this, BerryPortrait erodes a 10-pixel border off of the edge of each berry before calculating the mean and standard deviation of the $L^*a^*b^*$ and RGB color channels.

We were interested in comparing the outputs for color traits measured by both GiNA and BerryPortraits. Both platforms evaluate berries using all three RGB color channels, which were then used to calculate grayscale (Table 2).

We also developed additional color metrics that were not readily available in GiNA. Defined by the International Commission on Illumination (*CIE: Commission Internationale de l'Eclairage*), the CIELAB color space is a perceptually uniform color space that expresses color as three values based on human color perception: L^* for luminance (perceptual lightness), a^* for red/green color coordinates, and b^* for blue/yellow color coordinates—four unique colors of human vision from the visible spectrum [71]. Because these features are not available in GiNA, no comparison between platforms was possible for these traits.

Statistical analyses

We were interested in all pairwise comparisons of geometric and color traits measured by both GiNA and BerryPortraits. All analyses were performed in R (version 4.3.1) [72] with assistance from the *tidyverse* suite of packages [73]. Two-sample paired t-tests were performed for all traits between platforms. We quantify trait preservation (Tables 3 & 4) with mean trait difference between platforms, normalized by the smaller range of trait values within a single platform—we refer to this effect size as *range-normalized mean difference* (rmd) defined as the following with $\bar{d}_t$ as the mean difference for trait (t), $x_{(n)}, x_{(1)}, y_{(n)}, y_{(1)}$ as the max and min of BerryPortraits (x) and GiNA (y) respectively:

$$\text{rmd}_t = \frac{\bar{d}_t}{\min(x_{(n)} - x_{(1)}, y_{(n)} - y_{(1)})}$$

Since the magnitude of raw mean difference will be distorted by the scale of the trait, the chosen measure of effect size (rmd) aims to place all traits on a comparable scale so as to identify traits that are more variable. We normalize by the range within a single platform to emphasize that measurements from a single platform will likely vary more than a change between the platforms.

Pearson correlation coefficients were calculated to measure strength of correlation between both platforms for each trait. Correlation among traits in each platform was evaluated with a scatterplot matrix. Curvilinear regression lines were fit by local regression (LOESS) for each pair of traits for each platform. Means for each trait are reported with 95% confidence intervals for each platform. We also visually compared sample distributions between platforms for each trait with density plots and histograms. Overlapping density plots (similar means and variances) indicate substantially similar performance

Table 2 Color formulae used by GiNA and BerryPortraits

Trait	GiNA [41]	BerryPortraits
RGB-color	Arithmetic mean and median of red, green, and blue color channel values, considering all the pixels in the detected object	Same as GiNA
Grayscale	A transformation of the RGB to a gray-color space: Red_median * 0.299 + Blue_median * 0.114 + Green_median * 0.587	Same as GiNA
Color variation	Variance of the RGB color and grayscale color across all the pixels in the object	Same as GiNA
L^*	–	Perceptual lightness Scale: 0 (black)—100 (white) From CIELAB
a^*	–	Green (–)/Red (+) color axis From CIELAB
b^*	–	Blue (–)/Yellow (+) color axis From CIELAB

Description of parameters returned by the software

Table 3 For each trait and each platform: comparison of population means, population minimum values, population maximum values, lower confidence intervals (LCI), upper confidence intervals (UCI), mean differences (md), range-normalized mean differences (rmd), and Pearson correlation coefficients (r) among shape and size parameters comparing BerryPortraits and GiNA

Trait	Platform	Mean	Min	Max	LCI	UCI	md	rmd	r
Length (cm)	BerryPortraits	2.03	1.27	2.90	2.03	2.04	0.08	0.05	> 0.98
	GiNA	1.95	1.24	2.81	1.94	1.95			
Width (cm)	BerryPortraits	1.66	1.10	2.19	1.66	1.66	0.06	0.06	> 0.97
	GiNA	1.60	1.04	2.09	1.60	1.60			
Perimeter (cm)	BerryPortraits	6.19	3.98	8.50	6.17	6.20	0.63	0.14	> 0.97
	GiNA	5.56	3.54	7.40	5.55	5.58			
Area (cm ²)	BerryPortraits	2.59	1.07	4.37	2.58	2.60	0.13	0.04	> 0.99
	GiNA	2.46	1.01	4.13	2.44	2.47			
Surface area (cm ²)	BerryPortraits	9.94	4.15	16.78	9.89	9.98	0.25	0.02	> 0.93
	GiNA	9.69	3.94	17.13	9.64	9.74			
Volume (cm ³)	BerryPortraits	2.91	0.78	6.24	2.89	2.93	1.5	0.29	> 0.99
	GiNA	1.33	0.35	2.85	1.32	1.34			

GiNA from Diaz-Garcia et al. [41]

BerryPortraits from present study, Loarca et al.

Table 4 For each trait and each platform across all images (N = 171): Comparison of population means, population minimum values, population maximum values, lower confidence intervals (LCI), upper confidence intervals (UCI), mean differences (md), range-normalized mean differences (rmd), and Pearson correlation coefficients (r) among color parameters comparing BerryPortraits and GiNA

Trait	Platform	Mean	Min	Max	LCI	UCI	md	rmd	r
Red channel (RGB)	BerryPortraits	95.65	26	188	94.66	96.63	1.95	0.01	> 0.99
	GiNA	97.59	29	189	96.64	98.55			
Green channel (RGB)	BerryPortraits	36.72	17	193	36.01	37.43	2.80	0.02	> 0.99
	GiNA	39.52	19	194	38.81	40.22			
Blue channel (RGB)	BerryPortraits	31.46	16	132	31.04	31.88	2.48	0.02	> 0.99
	GiNA	33.94	18	132	33.53	34.35			
Grayscale (RGB)	BerryPortraits	53.73	24	183	53.03	54.45	2.51	0.02	> 0.99
	GiNA	56.25	26	184	55.55	56.94			
L* (CIELAB)	BerryPortraits	60.34	21	195	59.56	61.13	–	–	–
	GiNA	–	–	–	–	–			
a* (CIELAB)	BerryPortraits	24.63	– 18	44	24.35	24.92	–	–	–
	GiNA	–	–	–	–	–			
b* (CIELAB)	BerryPortraits	16.37	0	51	16.15	16.60	–	–	–
	GiNA	–	–	–	–	–			

GiNA software from Diaz-Garcia et al. [41]

BerryPortraits software from present study, Loarca et al.

between platforms. Non-overlapping density plots, with larger mean differences or more unequal variances, indicate dissimilar trait performance between platforms.

Results

YOLOv8 model performance

The YOLOv8 model was able to detect and segment cranberries with high accuracy across a variety of

backgrounds. On a pixel-wise segmentation mask classification basis, the precision (true positives/true positives + false positives) was 0.997 and the recall (true positives/true positives + false negatives) was 0.999, for a pixel-wise F1 value of 0.998. The mask mAP50 (area under the precision-recall curve averaged over all classes when an intersection-over-union of 0.5 is considered a true overlap) was 0.995. The mask mAP50-95 (mean area

under the precision-recall curve across all classes, averaged over intersection-over-union values ranging from 0.5 to 0.95 in 0.05 increments) was 0.927.

The only task where the model consistently struggled was in correctly segmenting very light berries from a white background. In all instances that we could find, an incorrect berry segmentation mask was accompanied by a berry solidity < 0.95; given this, we added a flag at runtime notifying the user of any berry with solidity under this value.

Platform comparison for shape and size traits

The output of each platform's image analysis includes size and shape values for each berry in each image. Trait means, 95% confidence intervals, and range-normalized mean differences are provided for shape and size parameters (Table 1).

Range-normalized mean differences (rmd) for shape and size traits were generally very low ($0.02 \leq \text{rmd} \leq 0.06$), except for perimeter (rmd = 0.14) and volume (rmd = 0.29). Correlation coefficients between platforms for shape and size traits were extremely high ($0.97 \leq r \leq 0.99$) to very high ($r = 0.94$ for surface area) (Fig. 1a). Due to the large sample size ($N = 171$ photos with 30 fruit per photo), all t-tests comparing trait performance between platforms yielded statistically significant results ($p \leq 0.05$; not shown), though the estimated differences for many of the traits are too small to be of meaningful concern. Overlapping density plots (similar means and variances) indicate substantially similar performance between platforms for all shape and size traits except perimeter area and volume (Fig. 1b, diagonals). Regression equations for each trait pair were also highly similar between platforms (Fig. 1b, lower panels). Correlations among traits, factored by platform, typically differed by 0.0–0.02 (Fig. 1b, upper panels). The largest differences in correlations between platforms were for surface area $\times$ length ($0.84 \leq r \leq 0.97$), surface area $\times$ width ($0.73 \leq r \leq 0.90$), and volume ($0.93 \leq r \leq 1$). Coefficients of the regression equations for each trait pair were also highly similar between platforms, except for all relationships related to volume (Fig. 1b, lower panels). Comparing solidity between platforms, we found that BerryPortraits (3 berries) had fewer berries below the 0.95 threshold than GiNA (9 berries out of 5130), indicating that, while both platforms demonstrate excellent ability to segment fruit, BerryPortraits has higher overall performance for solidity (Supp. Figure 5). There is no overlap between the berries that segmented poorly in either platform. Within the 0.95 quality threshold, segmenting ability in GiNA appears to decrease with pink or light red fruit, while in the same region for Berry Portraits, solidity appears to decrease with yellow fruit.

Platform comparison for color traits

The output of each platform's image analysis includes color trait values for each berry in each image. Color trait means, 95% confidence intervals, and range-normalized mean differences are provided (Table 4).

RGB and grayscale demonstrated very low range-normalized mean differences ($0.01 \leq \text{rmd} \leq 0.02$). Correlation coefficients between platforms for RGB were nearly identical ($r > 0.99$) (Fig. 2a—top row) and identical for grayscale calculated from RGB ($r = 1$) (Fig. 2a, bottom row). Luminance (L^* , perceptual lightness) in BerryPortraits correlates almost identically ($r > 0.99$) with grayscale calculated from RGB values in both GiNA and BerryPortraits. Overlapping density plots (similar means and variances) indicate substantially similar performance between platforms for all color traits (Fig. 2b, diagonals). In RGB and grayscale, correlation coefficients among traits were highly similar for both platforms and density plots overlapped substantially for all traits (Fig. 2b, upper panels). Correlation coefficients differed by 0.00–0.01. Coefficients of the regression equations for each trait pair were also highly similar between platforms, except for red channel $\times$ blue channel—while the slopes were nearly identical, the intercepts varied between platforms (Fig. 2b, lower panels). We visualized output of the CIELAB color space by converting values to hex code. Each data point represents the average cranberry color from each berry in each photo (Supp. Figure 6) and enables rapid evaluation of the range of color variation of berries in our population.

Discussion

BerryPortraits is an open source high throughput image-analysis pipeline for cranberry morphometric traits, making this software broadly applicable to horticultural research by food scientists, plant physiologists, plant breeders, and plant pathologists, and with strong potential for application in other fruit crops. We tested this software on cranberry populations from the two largest public cranberry breeding programs in the U.S., which makes this the largest cranberry image dataset ever to be validated. BerryPortraits has all the capacity to accurately measure digital images for all morphometric traits (length, width, area, volume, surface area, RGB, grayscale) measured by the previous unsupported, unimproved, and outdated software (GiNA), as well as new traits that were not available before, such as an intuitive color space (CIELAB and HSV), shape estimation improvements (surface area and volume), and uniformity metrics. BerryPortraits output enables rapid and accurate phenotyping for traits that require a large number of berries per genotype to accurately characterize, thus reducing the time and labor in phenotyping. Though the

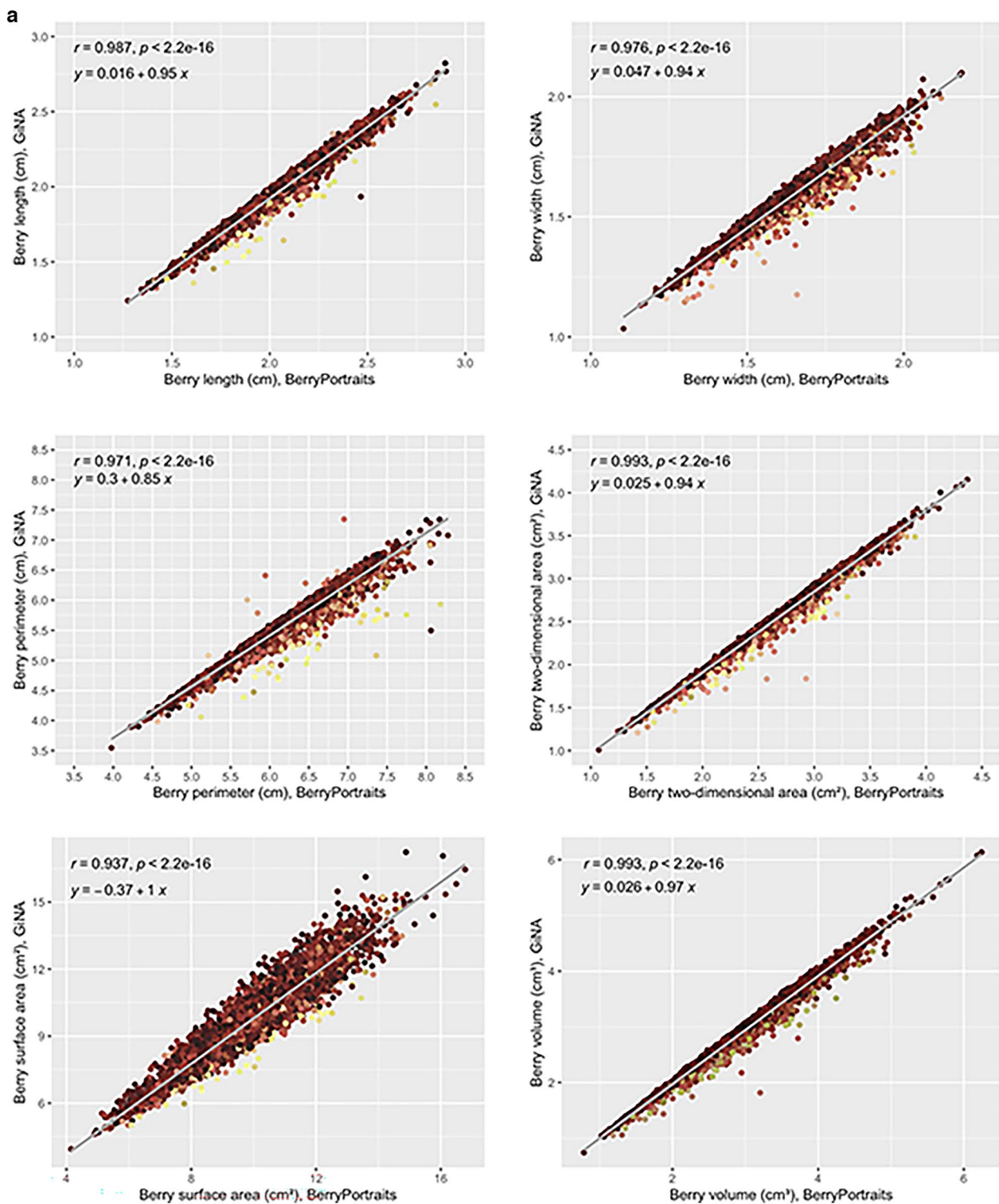


Fig. 1 **a** Pearson's correlation between size and shape parameters in BerryPortraits and GiNA, $n = 171$ samples. **b** Pearson correlation coefficients and single linear regression formulae among shape and size traits in BerryPortraits and GiNA. Shape and size relationships are generally conserved between platforms

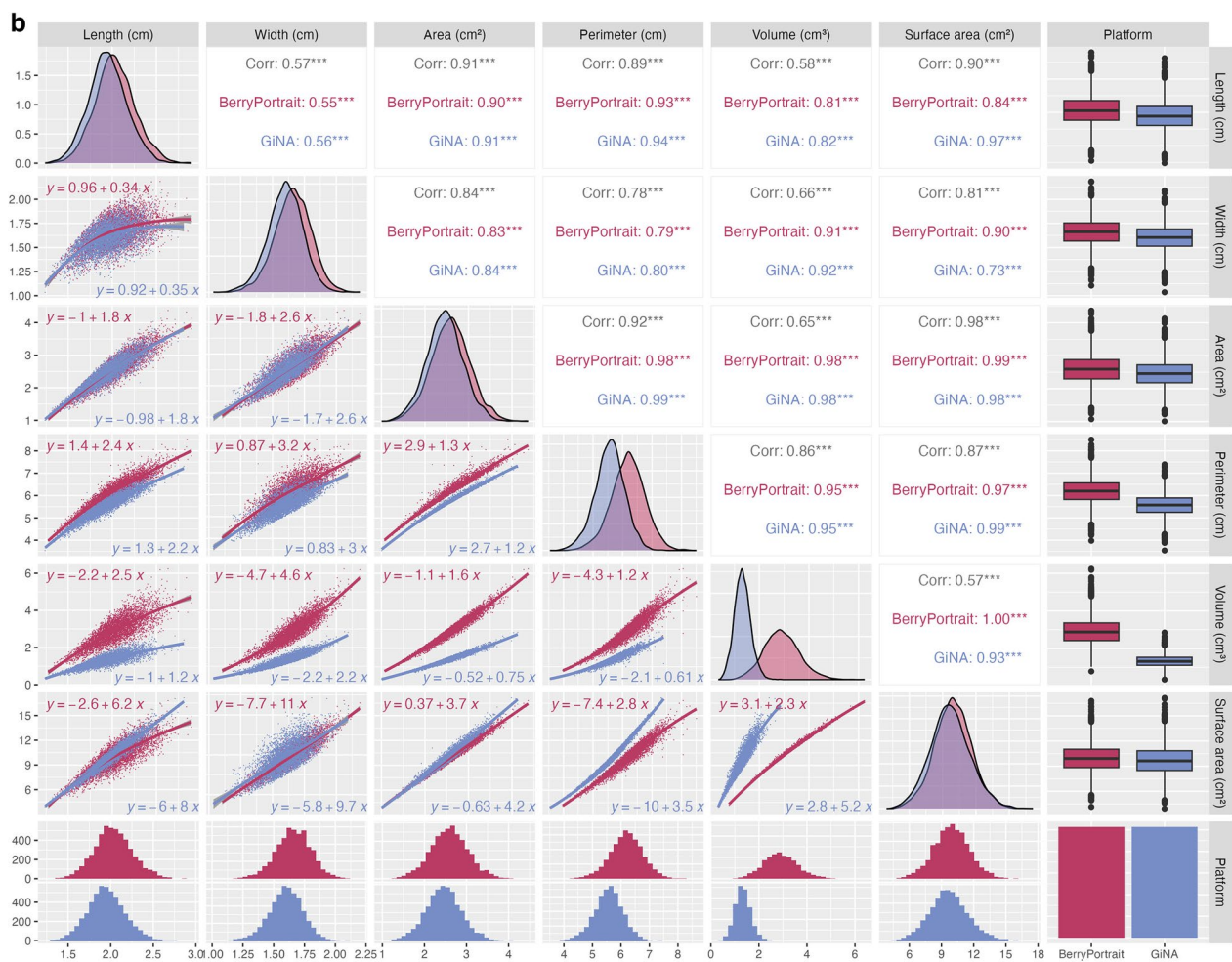


Fig. 1 continued

tool was originally conceived and designed for cranberry breeding populations, it has many potential applications for a wide range of plant science researchers to efficiently and accurately collect data on fruit quality traits.

Platform performance

In this study, we are using correlation between BerryPortraits with the validated software, GiNA, as a measure of ground-truthing. In the original paper, GiNA was validated using correlation between red color in berry images and berry anthocyanin content ($r=0.93$) [41]. BerryPortraits red channel correlated very highly with GiNA ($r_R > 0.99$). GiNA length and width measurements were validated by comparing hand caliper measurement of individual berries to length and width extracted from images ($0.92 < r < 0.97$), indicating very high accuracy. BerryPortraits length and width correlated very highly with GiNA's measurements ($r_L > 0.98$; $r_W > 0.97$; Table 3).

Platform methodology comparison

Both platforms offer classical computer vision and artificial neural network (ANN) approaches to perform semantic segmentation, a necessary preprocessing step needed to isolate foreground berry instances from extraneous background image components. However, the two platforms differ in the sophistication and capability of their ANNs to conduct segmentation.

BerryPortraits' underlying neural network segmentation framework is based on YOLOv8, a widely used SOTA model. YOLOv8 is pre-trained on a set of components useful to the berry inference process, such as berries, color cards, a variety of scale marker types, and text labels. This pre-trained and validated YOLOv8 is subsequently used to detect pertinent objects, returning an inference confidence value, the predicted class label, and a bounding box. BerryPortraits subsequently calculates the center of the object's bounding box and inputs the associated points and object class to a

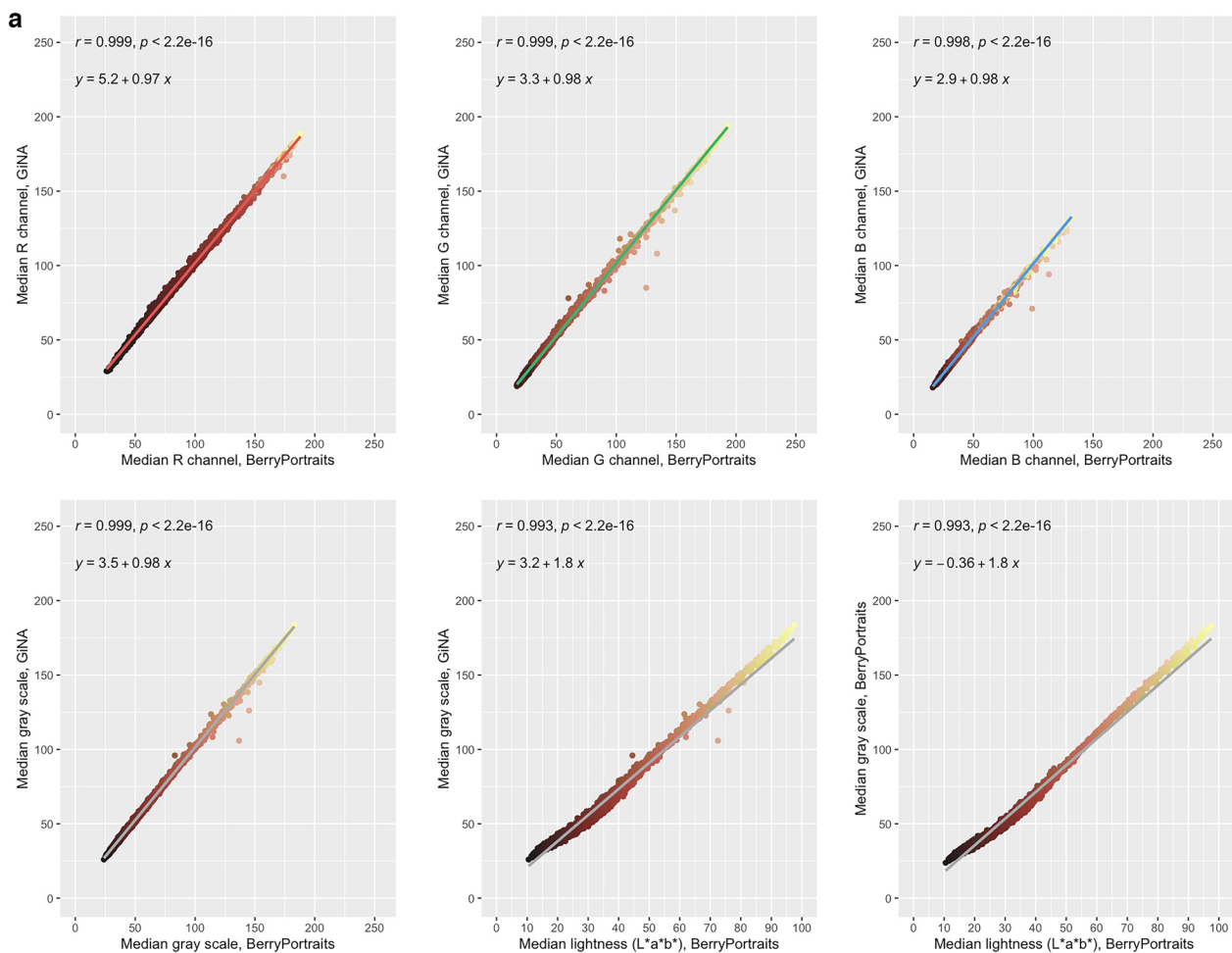


Fig. 2 a Pearson's correlation between RGB, grayscale, and lightness ($L^*a^*b^*$) values in BerryPortraits and GiNA. **b** Pearson correlation coefficients and single linear regression formulae among RGB and grayscale in BerryPortraits and GiNA. Color trait relationships are conserved between platforms

pre-trained model to extract a refined segmentation mask of relevant objects. This segmentation method balances classification performance with flexibility to accurately extract multi-class elements from an image containing berries. In contrast, GiNA uses a multi-layer perceptron (MLP) with a single hidden layer to do simple binary classification of pixels into foreground berries or background elements. This MLP is trained using two RGB images: a foreground image containing a manual sample of representative colors of berry pixels and a background image containing characteristic background colors. This binary classifier is synonymous with classical computer vision segmentation approaches using thresholded colorspace channels, in that it lacks the fidelity to robustly incorporate and classify other useful objects in an imaging scene, such as a sample label, color card, and scale markers—a capability that is built in to BerryPortraits.

Geometric features

BerryPortraits can quickly inspect outlines and oriented bounding boxes across dozens of images to visually confirm that the berry shapes as detected by BerryPortraits are correct. Length and width can easily be validated by hand, and even if the original fruits are no longer available, data can be compared to the typical values for berry size to confirm that it falls within the expected ranges.

Nearly one-to-one correlations between size and shape traits measured by BerryPortraits and GiNA (Fig. 1a) indicates a very strong relationship between two measurements. In most cases, corresponding traits that had high correlations between platforms also had means and variances that were substantially in agreement between platforms (diagonal of Fig. 1a—overlapping density plots). Therefore, BerryPortraits provides an equivalent estimate of length, width, area, and surface area, and can be used with confidence by researchers.

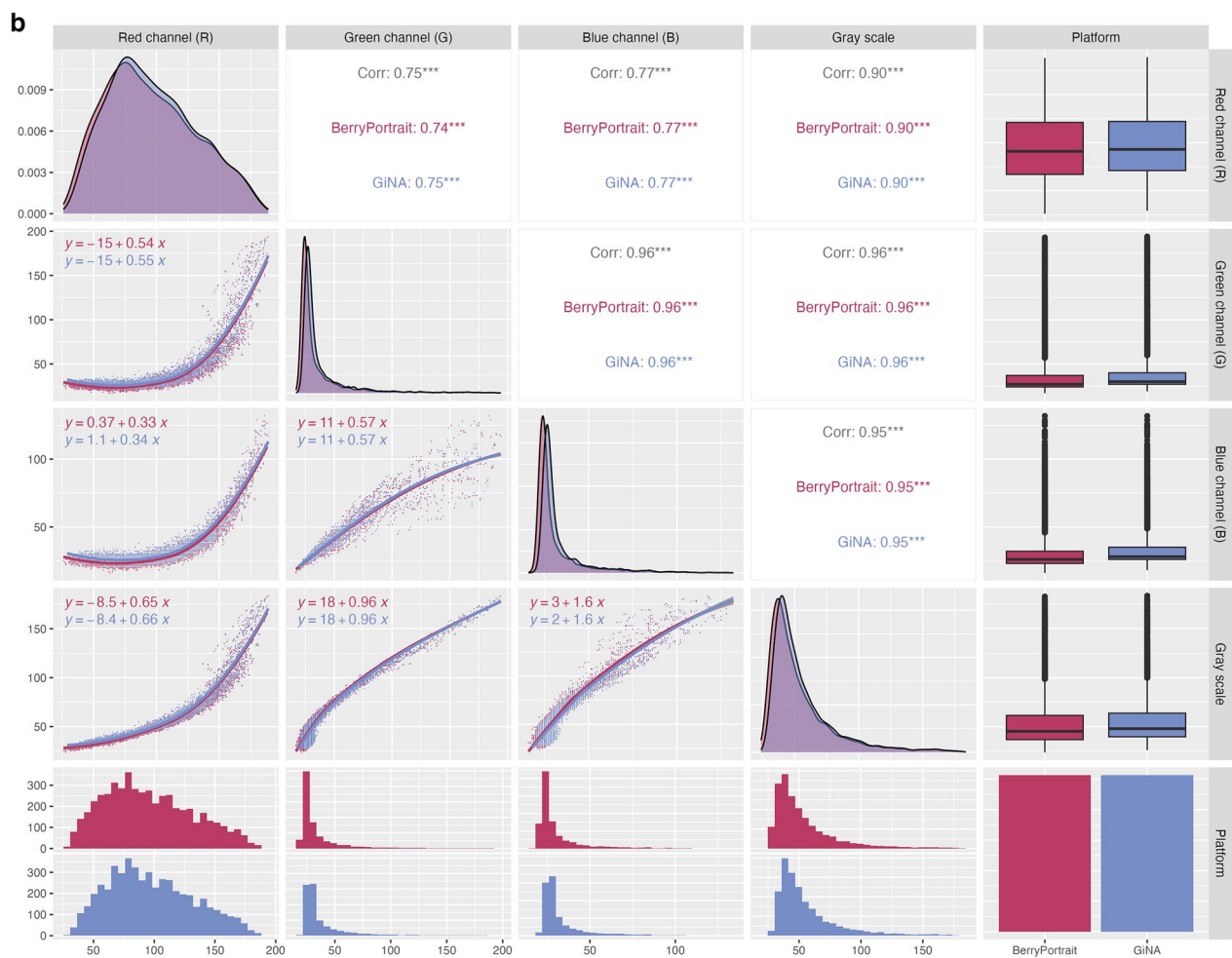


Fig. 2 continued

High correlation does not necessarily suggest equivalence. Though perimeter and volume are highly correlated between the GiNA and BerryPortraits (Table 3), density plots suggest a lack of agreement (Fig. 1a—non-overlapping density plots). This discrepancy is not surprising given that each platform had a different method for calculating volume. Berry volume and surface area are not easily measurable or ground-truthed with hand measurements—as such, these trait values were by necessity calculated from length and width. GiNA estimates surface area and volume by assuming a perfectly elliptical object, which does not work well for irregular shaped fruit (e.g., pyriform, spindle, or ovate shaped fruit). BerryPortraits has made substantial improvements compared with GiNA by addressing the technical challenge of measuring irregularly shaped fruit, by adjusting the method for calculating volume such that it is not limited by the assumption of perfect ellipses. We thus expect to produce more accurate volume measurements

for imperfect ellipsoids and by extension to improve our ability to detect significant marker associations for shape and size characteristics in cranberry and other irregularly shaped fruit crops. It is noteworthy that GiNA's role in the successful identification of three significant multi-year genetic markers (PVE < 10%) associated with image-based fruit size in blueberry—which correlates highly with fruit weight ($r > 0.97$), a very important factor in fruit quality [55]—lends credibility to the use of image-based fruit size in future fruit studies.

Color metrics

Since cranberries deepen in pigment dramatically as they ripen and accumulate anthocyanins, berry color can be a proxy for total anthocyanin content (TAc), which is useful since anthocyanin quantification is costly and laborious. Color metrics do not need to be complex in order to be useful—for instance, mean berry grayscale values as measured by GiNA were correlated

with TAcy ($r=0.83$) [8]. Both grayscale and RGB values have limitations, however. Grayscale values cannot distinguish between shades of green and red with similar grayscale values, which are often seen during the middle stage of ripening. Differences in RGB channel values are not perceptually uniform; two color differences of the same numerical value may look very different in magnitude to the human eye [74]. GiNA only measures

color with RGB, in which all three color channels are confounded with overall brightness—GiNA outputs grayscale values, but green/yellow (under-ripe) and red (ripe) berries can have very similar grayscale values (Fig. 3). The red channel (R) has counterintuitive color interpretation, as green fruit have highest R values and dark red fruit have lowest R values. A user selecting varieties with high mean R values would be selecting

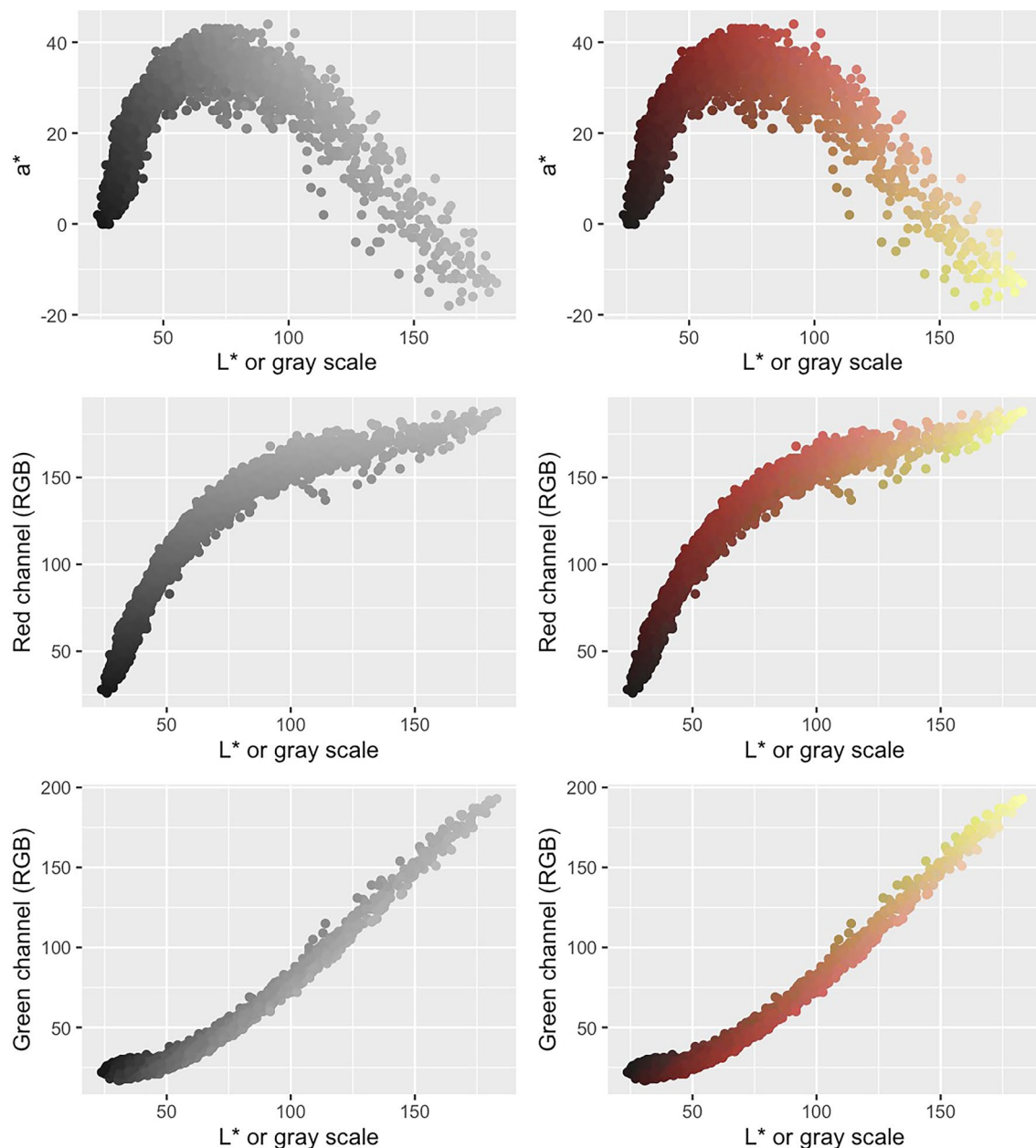


Fig. 3 In RGB color space, red channel and green channel confound berry color intensity (y axis) with berry lightness or darkness (x-axis). Berries with intense red pigment are also perceived as darker, while yellow berries are perceived as lighter. In CIELAB color space, color intensity is independent of perceived lightness (L^*). The y-axis (a^*) represents the red-green color axis from the CIELAB color space

for those berries that are brightest, i.e. those that are least ripe and with the *lowest* TAcY.

While BerryPortraits reports RGB and HSV values for completeness, the primary color metrics intended for downstream use are per-berry values in the $L^*a^*b^*$ (also referred to as CIELAB) color space. This has several advantages over the RGB color space. $L^*a^*b^*$ separates color into a lightness axis L^* and two color axes a^* and b^* , avoiding the issues of color channels being highly confounded with brightness (Fig. 3). The a^* dimension represents a color's position along roughly a red-green axis, making it an obvious a priori choice for fruits which change from red to green during ripening. Finally, the scale was designed so that differences in color would be perceptually uniform; if two objects differ in color, and their colors in $L^*a^*b^*$ coordinates when viewed under standardized conditions differ by a given Euclidean distance ΔE , this difference looks roughly the same to the human eye no matter in the color space the two objects lie [74]. Mean/median a^* value on a per-berry basis is a useful proxy for complete color development for red berries. Mean/median L^* is a simple proxy for overall berry brightness (ripeness): unripe berries are lighter in color and have higher L^* , while ripe berries are darker in color and have a lower L^* . It is highly correlated to grayscale value but offers slightly more resolution on darker berries. L^* in isolation is not suited for identifying berries in an acceptable window along a ripeness gradient. A breeder looking to quickly summarize which fruits are fully ripe, but not overripe, could not do so using L^* data alone. Low variance of image L^* indicates synchronous, uniform brightness and ripening across berries (berries are similar in ripeness to one another), while low variance of individual berry L^* indicates uniform individual berry color development. Low variance of image a^* indicates synchronous, uniform color development across berries (berries are similar in color to one another), while low variance of individual berry a^* indicates uniform individual berry color development.

We recommend that users rely on three metrics to understand the color of their imaged samples: mean or median L^* as a proxy for sample brightness (ripeness) (stabilizing selection); standard deviation of L^* as a proxy for uniform sample ripeness (low values); mean/median a^* as a measure of sample redness (high values); and standard deviation of a^* as a proxy for sample color uniformity (low values).

Significance in berry crop breeding

Over 95% of the cranberries produced are machine-harvested. The predominant market for the last 30 years has been the sweetened dried cranberry (SDC) processing market, which has driven the value and demand for high

quality cranberries. The most recent influx of cranberry cultivars in the market were bred for color intensity, size and shape, and early fruit maturation [8, 9, 75, 76], which are the highest priority fruit quality traits for processing efficiency by the commercial cranberry industry in the U.S. and Canada [77]. Thus the current cranberry market demands cultivars with inter-berry and intra-berry color and size uniformity (personal communication: Nicole Hansen). One of BerryPortraits's unique features is its ability to quantify developmental uniformity among berries in a sample, perceptually uniform ripeness in size (Fig. 4a) and color (Fig. 4b), within and between berries. Size uniformity increases harvest efficiency, as mechanical harvesters do not discriminate between ripe and unripe berries. Small berries and berries with incomplete

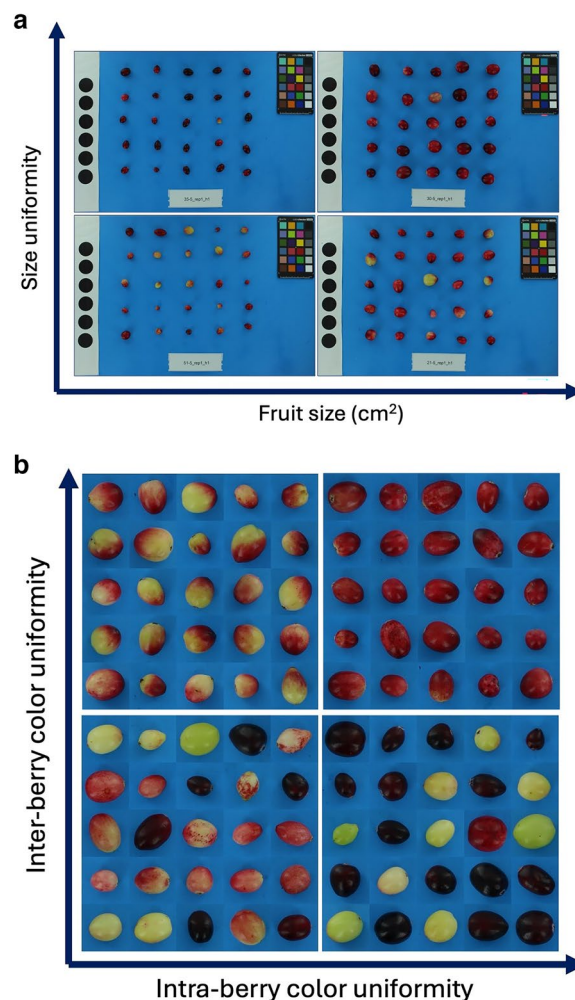


Fig. 4 **a** Inter-berry size uniformity and fruit size are important metrics for cranberry processing efficiency. **b** Inter-berry color uniformity and intra-berry color uniformity are important metrics for cranberry processing efficiency

color development are culled during the sorting and packing process. As such, cultivars with a high proportion of uniformly red berries are highly preferred. Furthermore, uniform size increases processing efficiency, as fruit with extreme size variability do not process well during SDC production, and thus are sorted out and culled prior to processing. Identifying breeding lines that produce uniformly sized fruit at harvest will reduce losses for growers and processors. BerryPortraits is also capable of distinguishing between green and white berries, which is important for both breeding and sorting. A fraction of white berries are usable for SDC, as during the infusion step of processing, white berries can absorb the syrup and develop into marketable red or pale red fruit (personal communication: Amaya Atucha). However, green berries become brown after processing, resulting in complete loss of commercial quality. Thus, breeders prioritize advancement of varieties with synchronous ripening and complete color development, which reduces losses for growers due to underdeveloped berries. With the data output from BerryPortraits, breeders can use selection indices to identify, select, and advance cranberry genotypes that meet the needs for cranberry growers and processors (Fig. 4a, b). To this end, selecting genotypes with a high a^* value, low a^* variance, high L^* value, and low L^* variance will yield lines that are uniformly red and uniformly bright.

Applications

Though developed for plant breeders, BerryPortraits offers innumerable possible research applications for food scientists, plant physiologists, plant pathologists, and horticultural researchers. Repeated measurements over the fruit ripening period to estimate the rate of ripening for different genotypes and identify genotypes suited to seasonal harvest windows, or genotypes with concentrated uniform ripening for machine harvest. Plant geneticists can use this high throughput and accurate phenotype output to identify trait-associated markers in linkage mapping populations or diverse populations. Plant physiologists and horticultural researchers can evaluate correlation of plant pigment accumulation with other morphometric characteristics, or relationships among visual ripening cues and sensory ripening cues. Food science researchers can efficiently and accurately quantify fruit quality under various storage conditions, insect/pathogen damage to fruit, and uniformity metrics provide an updated method for evaluating grades and standards in USDA's agricultural marketing service. The integration of BerryPortraits with BreedBase [59] and Fieldbook [60] enables efficient phenotype data collection.

Imaging recommendations

Capturing consistent high-quality images is the most important factor in obtaining accurate, reproducible results from BerryPortraits. While a full treatment of photography methods is outside the scope of this paper, below are several basic recommendations for best practices.

Choose a background that is durable and uniform, in a solid color dissimilar to that of the target berries across all stages of fruit ripening and color development. The white background used for some of the images in this study had poor contrast against pale unripe berries, which could cause issues with segmentation. We recommend users test out gray backgrounds available from photography supply vendors. For repeatability and accuracy, the same background should be used for all images you wish to compare or analyze as a single dataset.

Keep image layout consistent, with all berries separated and minimal non-target objects visible in frame. Berries do not need to be aligned into a grid, but this can help with analyzing data and tracing any potential issues.

Always include size markers and a color reference card, ideally a standard X-Rite ColorChecker Classic. Doing appropriate color correction with a proper color reference card is challenging; without one, it is impossible. Similarly, berries cannot be calibrated to real units without appropriate size markers.

Use diffuse lighting from a single controlled light source with a high color rendering index (CRI). Do not rely on ambient daylight, as this can change dramatically within and between days. On berries with reflective epicuticular waxes, direct light without a diffuser or reflector will lead to bright spots of specular reflection, which will throw off color analyses. While diffusers are ideal for waxy berries, reflectors are also useful; either is preferable to direct, non-diffuse light.

When capturing images with a DSLR camera, set white balance before imaging and capture images in RAW format in addition to JPG format. Post hoc color correction on compressed JPG images may be useful, but color correction and white balancing on RAW images is preferable. The ideal protocol is to eliminate or reduce the need for any post hoc color correction by capturing images with a uniform approach.

Learning and creating in BerryPortraits

Although this program was built to be intuitive and friendly, we encourage users to approach with patience, a growth-mindset, playfulness, and willingness to learn. We advise beginners to acquire some basic familiarity with Python and to continue to grow and enhance their programming skills, which will substantially improve

their experience using and creatively building upon BerryPortraits.

Learning a programming language can be both frustrating and rewarding, and thankfully there are abundant free community-based resources for supporting Python learners. Finally, we recommend pilot testing phenotypically representative berry subsamples before launching into a full-scale experimentation.

Limitations

Though the human health benefits of anthocyanin are well studied and widely known in a variety of crops [8, 9, 78–84], the targeted trait in cranberry breeding programs is the appearance of the color, and the color of interest to the market (bright red) is not as dark or anthocyanin-rich as the darkest berries. BerryPortraits differentiates between bright and dark berries, however researchers may need to troubleshoot to achieve the conditions required to distinguish among the most deeply pigmented berries.

Quantifying berry color is a complex process. Variations in lighting and background can affect mean color values to a drastic extent, and color correction is not a panacea. When comparing two image sets captured under different conditions, any comparison of berry color values is best done with a high degree of caution.

It is not currently known how these visual ripeness traits correlate with indicators of ripeness (such as degrees Brix, volatiles, or other physiological ripeness traits). An important caveat in our grower-friendly definition of “ripeness” is that fully physiologically ripe fruit can have incomplete (non-red) color development.

Finally, we advise that researchers of climacteric fruit minimize the time between harvest and imaging, as fruit color may continue to develop and change in the interim.

Conclusions

BerryPortraits is an open source, Python-based, high throughput phenotyping platform that accurately measures berry color, size, and shape, saving an abundance of time, labor, and cost in horticultural research programs. There are innumerable possible creative applications for this software by food scientists, plant physiologists, horticultural researchers, and plant breeders. BerryPortraits has strong potential to be adapted for phenotyping other berry crops, such as grape, blueberry, lingonberry, caneberry, and any crops in which color, size, shape, and uniformity are important for fruit quality, production, or marketability.

Abbreviations

CIELAB Perceptually uniform color space that expresses color as three values based on human perception from the visible spectrum: $L^*a^*b^*$

CIE (from CIELAB)	<i>Commission Internationale de l'Eclairage</i>
L^* (from CIELAB)	L^* for luminance (perceptual lightness)
a^* (from CIELAB)	A^* for red/green color coordinates
b^* (from CIELAB)	B^* for blue/yellow color coordinates
DSLR	Digital single lens reflex (camera)
MLP	Multilayer perceptron
QTL	Quantitative trait locus/loci
RGB	Red/green/blue channel
SDC	Sweetened dried cranberries
TAcy	Total anthocyanin
YOLOv8	You Only Look Once (version 8)

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01285-1>.

Supplementary Material 1. Supp. Fig. 1: Cranberry samples from a single genotype staged within an enclosed lightbox on a blue background with black 2.54-cm scale markers, color card, and sample label. Camera is placed atop the lightbox and the lens dropped through an opening in the lightbox's ceiling. Black light curtain with reflective interior is replaced over lightbox before image capture. The imaging field is displayed on the computer monitor. A homemade wooden block with 5x5 bored holes facilitates rapid berry placement. Supp. Fig. 2: Example of berry imaging photographs inside the enclosed light box for controlled and standardized lighting conditions. Sample of background 30 fruit with 2.54-cm diameter black scale markers. Photograph from digital SLR. Supp. Fig. 3: Examples of cranberry images used for training set with YOLOv8. Supp. Fig. 4: Volume approximated as a series of stacked cylinders, 1 pixel high; surface area approximated as a series of stacked truncated cones, 1 pixel high. Supp. Fig. 5: Solidity comparison between BerryPortraits and GiNA. BerryPortraits had fewer berries below the 0.95 threshold than GiNA. While both platforms demonstrate excellent ability to segment fruit, BerryPortraits has higher overall performance for solidity. GiNA's segmenting ability appears to decrease with pink or light red fruit, while in BerryPortraits, solidity appears to decrease with yellow fruit. Supp. Fig. 6: Range of colors in cranberry population under study. $L^*a^*b^*$ values converted to hexcode for visualization

Acknowledgements

The authors would like to extend gratitude to Nicholi Vorsa for his care in creating and stewarding the CNJ02-1 population and for sharing propagules for establishment in Wisconsin; AJ Ackerman, Josue Chinchilla-Vargas, Edwin Reidel, and Tyler Slonecki (and the entire Breeding Insight science team); David Meidlinger, (and the entire Breeding Insight development team); the BI leadership, program coordinators, and technical steering board; Jim Spoden, Kristy Morrow, Ree Klumb, Randey Lewis, Gustavo Ivan Hernandez-Sierra, and Eric Wiesman for their diligent assistance in the lab; and Nicole Hansen and Bill Hatch (Cranberry Creek Inc.) for effusively sharing their knowledge of the commercial cranberry industry, which helped us to identify relevant traits for cranberry growers. JZ would like to express his gratitude through PS 136:1. The authors thank Collins Wakholi and Devin Rippner for their generous consultation and sharing their knowledge of digital image color correction.

Author contributions

Conceptualization (JL, TWH, HLM, AFM, LDG, JN, CTB, MJS, JZ) Design (JL, TWH, HLM, AFM, CFS, CTB, MJS, JZ) Data collection (JL, HLM, JN) Analysis (JL, TWH, AFM, ML, MATM) Software testing (JL, TWH, HLM, AFM, MATM, JN, GMS) Preparation of the manuscript (all co-authors).

Funding

This project was supported by the Agriculture and Food Research Initiative (AFRI) grant no. 2022-67012-37202 from the USDA National Institute of Food and Agriculture (NIFA); USDA NIFA, award number 2019-51181-30015, project “VacciniumCAP: Leveraging genetic and genomic resources to enable development of blueberry and cranberry cultivars with improved fruit quality attributes”; United States Department of Agriculture (USDA) Agricultural Research Service (ARS), project no. 5090-21220-007-00-D, Ocean Spray Cranberries, Inc.,

Wisconsin Cranberry Growers Association, Cranberry Institute, Mariani Packing Co. Breeding Insight is funded by the U.S. Department of Agriculture, under agreement numbers [8062-21000-043-004-A, 8062-21000-052-002-A, and 8062-21000-052-003-A].

Availability of data and materials

BerryPortraits software is available at <https://github.com/Breeding-Insight/BerryPortraits/>. The datasets used and/or images analyzed are available at <https://cggl.horticulture.wisc.edu/> or from Juan Zalapa (juan.zalapa@usda.gov) on reasonable request.

Availability and requirements: Project name: BerryPortraits: Phenotyping Of Ripening Traits with Rapid Automated Imaging Tools. Project home page: <https://github.com/Breeding-Insight/BerryPortraits/>. Operating system(s): Platform independent. Programming language: Python. Other requirements: Users of BerryPortraits web app will require an ORCID ID for authentication. Users of BerryPortraits code will require a GitHub account to fork or download the code.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Plant and Agroecosystem Sciences, University of Wisconsin-Madison, Madison, WI, USA. ²United States Department of Agriculture-Agricultural Research Service, Vegetable Crops Research Unit, Madison, WI, USA. ³Cornell University-Breeding Insight, Ithaca, NY, USA. ⁴Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA. ⁵Department of Viticulture and Enology, University of California-Davis, Davis, CA, USA. ⁶Phillip E. Marucci Center for Blueberry and Cranberry Research & Extension, Chatsworth, NJ, USA. ⁷United States Department of Agriculture-Agricultural Research Service, Genetic Improvement of Fruits & Vegetables Laboratory, Beltsville, MD, USA. ⁸Rutgers University-Department of Plant Biology, New Brunswick, NJ, USA. ⁹Department of Horticultural Science, North Carolina State University, Raleigh, NC, USA. ¹⁰Plant Human Health Institute, North Carolina State University, Raleigh, NC, USA.

Received: 15 June 2024 Accepted: 2 October 2024

Published: 13 November 2024

References

- Regan BC, Julliot C, Simmen B, Viénot F, Charles-Dominique P, Mollon JD. Fruits, foliage and the evolution of primate colour vision. *Philos Trans R Soc Lond B Biol Sci*. 2001;356(1407):229–83.
- Goldman IL, Wang Y, Alfaro AV, Brainard S, Oravec MW, McGregor CE, et al. Form and contour: breeding and genetics of organ shape from wild relatives to modern vegetable crops. *Front Plant Sci*. 2023;14:1257707.
- Furbank RT, Tester M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci*. 2011;16(12):635–44.
- Manco R, Basile B, Capuozzo C, Scognamiglio P, Forlani M, Rao R, et al. Molecular and phenotypic diversity of traditional European plum (*Prunus domestica* L.) germplasm of Southern Italy. *Sustainability*. 2019;11(15):4112.
- Basile B, Mataffo A, Forlani M, Corrado G. Diversity in morphometric, pomological, and fruit-quality traits of apricot (*Prunus armeniaca*) traditional varieties: implications for landrace differentiation at regional scale. *Diversity*. 2022;14(8):608.
- Wang B, Chen X, Huang S, Tan J, Zhang H, Wang J, et al. Bulk segregant analysis sequencing and RNA-Seq analyses reveal candidate genes associated with sepal color phenotype of eggplant (*Solanum melongena* L.). *Plants*. 2024;13(10):1385.
- Kuehni RG, editor. Color: an introduction to practice and principles. John Wiley & Sons; 2012.
- Díaz-García L, Schlautman B, Covarrubias-Pazarán G, Maule A, Johnson-Cicalese J, Grygleski E, et al. Massive phenotyping of multiple cranberry populations reveals novel QTLs for fruit anthocyanin content and other important chemical traits. *Mol Genet Genomics*. 2018;293(6):1379–92.
- Vorsa N, Zalapa J. Domestication, genetics, and genomics of the american cranberry. In: Goldman I, editor. Plant breeding reviews. 1st ed. Wiley; 2019. p. 279–315. <https://doi.org/10.1002/9781119616801.ch8>.
- Díaz-García L, Covarrubias-Pazarán G, Schlautman B, Grygleski E, Zalapa J. Image-based phenotyping for identification of QTL determining fruit shape and size in American cranberry (*Vaccinium macrocarpon* L.). *PeerJ*. 2018;6:e5461.
- Owens KW, Bliss FA, Peterson CE. Genetic analysis of fruit length and weight in two cucumber populations using the inbred backcross line method. *J Amer Soc Hort Sci*. 1985;110(3):431–6.
- Lopez-Moreno H, Phillips M, Díaz-García L, Torres-Meraz MA, De La Torre F, Berro I, et al. A survey of key methods, traits, parameters, and conditions for measuring texture in cranberry (*Vaccinium macrocarpon* Ait.). *Horticulturae*. 2023;9(4):479.
- Díaz-García L, Rodríguez-Bonilla L, Phillips M, Lopez-Hernández A, Grygleski E, Atucha A, et al. Comprehensive analysis of the internal structure and firmness in American cranberry (*Vaccinium macrocarpon* Ait.) fruit. *PLoS ONE*. 2019;14(9):e0222451.
- Gehan MA, Kellogg EA. High-throughput phenotyping. *Am J Bot*. 2017;104(4):505–8.
- Arya S, Sandhu KS, Singh J, Kumar S. Deep learning: as the new frontier in high-throughput plant phenotyping. *Euphytica*. 2022;218(4):47.
- Dang LM, Nadeem M, Nguyen TN, Park HY, Lee ON, Song HK, et al. VPBR: an automatic and low-cost vision-based biophysical properties recognition pipeline for pumpkin. *Plants*. 2023;12(14):2647.
- Fontein H, Afonso M, Lensink D, Mooij M, Faber N, Vroegop A, et al. Automatic phenotyping of tomatoes in production greenhouses using robotics and computer vision: from theory to practice. *Agronomy*. 2021;11(8):1599.
- Afonso M, Fontein H, Fiorentin FS, Lensink D, Mooij M, Faber N, et al. Tomato fruit detection and counting in greenhouses using deep learning. *Front Plant Sci*. 2020. <https://doi.org/10.3389/fpls.2020.571299/full>.
- Van Der Heijden G, Song Y, Horgan G, Polder G, Dieleman A, Bink M, et al. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. *Functional Plant Biol*. 2012;39(11):870.
- Ampatzidis Y, Partel V. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*. 2019;11(4):410.
- Sudars K, Namatēvs I, Judvaitis J, Balašs R, Ņikūlins A, Peter A, et al. YOLOv5 Deep Neural Network for Quince and Raspberry Detection on RGB Images. In: 2022 workshop on microwave theory and techniques in wireless communications (MTTW). 2022. p. 19–22. <https://ieeexplore-ieee.org.ezproxy.library.wisc.edu/abstract/document/9942550>. Accessed 8 Jun 2024.
- Li H, Lee WS, Wang K. Identifying blueberry fruit of different growth stages using natural outdoor color images. *Comput Electron Agric*. 2014;106:91–101.
- Ni X, Li C, Jiang H, Takeda F. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. *Hortic Res*. 2020;7(1):110.
- Gonzalez S, Arellano C, Tapia JE. Deepblueberry: quantification of blueberries in the wild using instance segmentation. *IEEE Access*. 2019;7:105776–88.
- Aquino A, Barrio I, Diago MP, Millán B, Tardaguila J. VitisBerry: an Android-smartphone application to early evaluate the number of grapevine berries by means of image analysis. *Comput Electron Agric*. 2018;148:19–28.
- Ariza-Sentís M, Baja H, Vélez S, Valente J. Object detection and tracking on UAV RGB videos for early extraction of grape phenotypic traits. *Comput Electron Agric*. 2023;211:108051.
- Daviet B, Fournier C, Cabrera-Bosquet L, Simonneau T, Cafier M, Romieu C. Ripening dynamics revisited: an automated method to track the development of asynchronous berries on time-lapse images. *Plant Methods*. 2023;19(1):146.
- Haucke T, Herzog K, Barré P, Höfle R, Töpfer R, Steinhage V. Improved optical phenotyping of the grape berry surface using light-separation and

- automated RGB image analysis. *VITIS J Grapevine Res.* 2021. <https://doi.org/10.5073/vitis.2021.60.1-10>.
29. Yoon S, Shin M, Kim JH, Jeong HJ, Park J, Ahn TI. Computer vision approach for phenotypic characterization of horticultural crops. *J Bio-Env Con.* 2024;33(1):63–70.
 30. Abebe AM, Kim Y, Kim J, Kim SL, Baek J. Image-based high-throughput phenotyping in horticultural crops. *Plants.* 2023;12(10):2061.
 31. James KMF, Sargent DJ, Whitehouse A, Cielniak G. High-throughput phenotyping for breeding targets—current status and future directions of strawberry trait automation. *Plants People Planet.* 2022;4(5):432–43.
 32. Gibbons A. Modular computer vision pipelines for monitoring lettuce growth.
 33. Kim C, van Iersel MW. Image-based phenotyping to estimate anthocyanin concentrations in lettuce. *Front Plant Sci.* 2023. <https://doi.org/10.3389/fpls.2023.1155722/full>.
 34. Brainard SH, Bustamante JA, Dawson JC, Spalding EP, Goldman IL. A digital image-based phenotyping platform for analyzing root shape attributes in carrot. *Front Plant Sci.* 2021;12: 690031.
 35. Turner SD, Ellison SL, Senalik DA, Simon PW, Spalding EP, Miller ND. An automated image analysis pipeline enables genetic studies of shoot and root morphology in carrot (*Daucus carota* L.). *Front Plant Sci.* 2018;9:1703.
 36. Junior M, Gomes RS, De Almeida CF, Carneiro VQ, De Oliveira RL, Laurindo RDF, et al. Research Article Digital phenotyping of winter squash fruits. *Genet Mol Res.* 2020. <https://doi.org/10.4238/gmr18646>.
 37. Feldmann MJ, Tabb A. Cost-effective, high-throughput phenotyping system for 3D reconstruction of fruit form. *Plant Phenome J.* 2022;5(1): e20029.
 38. Bouillon P, Fanciullino AL, Belin E, Bréard D, Boisard S, Bonnet B, et al. Image analysis and polyphenol profiling unveil red-flesh apple phenotype complexity. *Plant Methods.* 2024;20(1):71.
 39. Kirchgessner N, Hodel M, Studer B, Patocchi A, Broggini GA. FruitPhenoBox—a device for rapid and automated fruit phenotyping of small sample sizes. In Review; 2023. <https://www.researchsquare.com/article/rs-2632704/v1>. Accessed 1 Mar 2024.
 40. Xiao B, Nguyen M, Yan WQ. Fruit ripeness identification using YOLOv8 model. *Multimed Tools Appl.* 2024;83(9):28039–56.
 41. Diaz-Garcia L, Covarrubias-Pazarán G, Schlautman B, Zalapa J. GiNA, an efficient and high-throughput software for horticultural phenotyping. *PLoS ONE.* 2016;11(8): e0160439.
 42. Chizk TM, Lee JA, Clark JR, Worthington ML. ShinyFruit: interactive fruit phenotyping software and its application in blackberry. *Front Plant Sci.* 2023;14:1182819.
 43. Strautiņa S, Kalniņa I, Kaufmane E, Sudars K, Namatēvs I, Judvaitis J, et al. Initial results of the development of intelligent non-invasive phenotyping of raspberries using machine learning and 3D imaging. *Acta Hortic.* 2023;1381:101–8.
 44. Zingaretti LM, Monfort A, Pérez-Enciso M. Automatic fruit morphology phenome and genetic analysis: an application in the octoploid strawberry. *Plant Phenomics.* 2021. <https://doi.org/10.34133/2021/9812910>.
 45. Nagamatsu S, Tsubone M, Wada T, Oku K, Mori M, Hirata C, et al. Strawberry fruit shape: quantification by image analysis and QTL detection by genome-wide association analysis. *Breed Sci.* 2021;71(2):167–75.
 46. He JQ, Harrison RJ, Li B. A novel 3D imaging system for strawberry phenotyping. *Plant Methods.* 2017;13(1):93.
 47. Li B, Cockerton HM, Johnson AW, Karlström A, Stavridou E, Deakin G, et al. Defining strawberry shape uniformity using 3D imaging and genetic mapping. *Hortic Res.* 2020;7(1):115.
 48. Hayashi A, Tanabata T, Wada T, Isobe S. A proposal of image analysis system for measuring strawberries. *Hort J.* 2017;16(Suppl 1):446.
 49. Leiva-Valenzuela GA, Aguilera JM. Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality. *Food Control.* 2013;33(1):166–73.
 50. Matiacevich S, Celis Cofré D, Silva P, Enrione J, Osorio F. Quality parameters of six cultivars of blueberry using computer vision. *Int J Food Sci.* 2013;2013(1): 419535.
 51. Matiacevich S, Silva P, Enrione J, Osorio F. Quality assessment of blueberries by computer vision. *Procedia Food Science.* 2011;1:421–5.
 52. Park B, Shin TS, Cho JS, Lim JH, Park KJ. Characterizing hyperspectral microscope imagery for classification of blueberry firmness with deep learning methods. *Agronomy.* 2022;12(1):85.
 53. Barré P, Herzog K, Höfle R, Hullin MB, Töpfer R, Steinhage V. Automated phenotyping of epicuticular waxes of grapevine berries using light separation and convolutional neural networks. *Comput Electron Agric.* 2019;156:263–74.
 54. Mengist MF, Grace MH, Xiong J, Kay CD, Bassil N, Hummer K, et al. Diversity in metabolites and fruit quality traits in blueberry enables ploidy and species differentiation and establishes a strategy for future genetic studies. *Front Plant Sci.* 2020;11:370.
 55. Mengist MF, Bostan H, Young E, Kay KL, Gillitt N, Ballington J, et al. High-density linkage map construction and identification of loci regulating fruit quality traits in blueberry. *Hortic Res.* 2021;8(1):169.
 56. Reinert S, Van Tassel DL, Schlautman B, Kane NC, Hulke BS. Assessment of the biogeographical variation of seed size and seed oil traits in wild *Silphium integrifolium* Michx. genotypes. *Plant Genet Resour.* 2019;17(5):427–36.
 57. Schlautman B, Diaz-Garcia L, Barriball S. Reprint of: Morphometric approaches to promote the use of exotic germplasm for improved food security and resilience to climate change: a kura clover example. *Plant Sci.* 2020;295: 110415.
 58. Maule AF, Loarca J, Diaz-Garcia L, Lopez-Moreno H, Johnson-Cicalese J, Vorsa N, et al. Of buds and bits: a meta-QTL study identifies stable QTL for berry quality and yield traits in cranberry mapping populations (*Vaccinium macrocarpon* Ait.). *Front Plant Sci.* 2024. <https://doi.org/10.3389/fpls.2024.1294570/full>.
 59. Morales N, Ogbonna AC, Ellerbrock BJ, Bauchet GJ, Tantikanjana T, Teclé IY, et al. Breedbase: a digital ecosystem for modern plant breeding. *G3 Genes Genomes Genet.* 2022;12(7):078.
 60. Rife TW, Poland JA. Field book: an open-source application for field data collection on android. *Crop Sci.* 2014;54(4):1624–7.
 61. Schlautman B, Covarrubias-Pazarán G, Diaz-Garcia LA, Johnson-Cicalese J, Iorizzo M, Rodríguez-Bonilla L, et al. Development of a high-density cranberry SSR linkage map for comparative genetic analysis and trait detection. *Mol Breeding.* 2015;35(8):177.
 62. Schlautman B, Bolivar-Medina J, Hodapp S, Zalapa J. Cranberry SSR multiplexing panels for DNA horticultural fingerprinting and genetic studies. *Sci Hortic.* 2017;219:280–6.
 63. Covarrubias-Pazarán G, Schlautman B, Diaz-Garcia L, Grygleski E, Polashock J, Johnson-Cicalese J, et al. Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* Ait. *Front Plant Sci.* 2018;9:1310.
 64. Diaz-Garcia L, Garcia-Ortega LF, González-Rodríguez M, Delaye L, Iorizzo M, Zalapa J. Chromosome-level genome assembly of the american cranberry (*Vaccinium macrocarpon* Ait.) and its wild relative *Vaccinium microcarpum*. *Front Plant Sci.* 2021;12: 633310.
 65. Maule AF, Loarca J, Diaz-Garcia L, Lopez-Moreno H, Vorsa N, Iorizzo M, et al. Of buds and bits: a metaQTL study on traditional upright traits and modern plot phenotyping in cranberry (*Vaccinium macrocarpon* Ait.). *Front Plant Sci.* 2024;15:1294570.
 66. Jocher G, Chaurasia A, Qui J. Ultralytics YOLO; 2023. <https://github.com/ultralytics/ultralytics>
 67. Wang CY, Yeh IH, Liao HYM. YOLOv9: learning what you want to learn using programmable gradient information. arXiv; 2024. <http://arxiv.org/abs/2402.13616>. Accessed 12 Sept 2024.
 68. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. YOLOv10: real-time end-to-end object detection. arXiv; 2024. <http://arxiv.org/abs/2405.14458>. Accessed 12 Sept 2024.
 69. Aharon S, Louis-Dupont, oferbaratz, Masad O, Yurkova K, Fridman L, et al. Deci-AI/super-gradients: 3.0.8. Zenodo; 2023. <https://zenodo.org/records/7789328>. Accessed 12 Sept 2024.
 70. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE; 2023. p. 3992–4003. <https://ieeexplore.ieee.org/document/10378323/>. Accessed 26 May 2024.
 71. Joblove GH, Greenberg D. Color spaces for computer graphics.
 72. R Core Team. R: a language and environment for statistical computing [computer software]. Vienna: R Foundation for Statistical Computing; 2021.
 73. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund R, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K,

- Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the tidyverse. *J Open Sour Softw*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>.
74. Schanda J, editor. *Colorimetry: understanding the CIE system*. John Wiley & Sons; 2007.
 75. Hancock JF, Lyrene P, Finn CE, Vorsa N, Lobos GA. Blueberries and cranberries. In: Hancock JF, editor. *Temperate fruit crop breeding*. Dordrecht: Springer; 2008. p. 115–50. https://doi.org/10.1007/978-1-4020-6907-9_4.
 76. Vorsa N, Polashock J, Cunningham D, Roderick R, Howell A. Evaluation of fruit chemistry in cranberry germplasm: potential for breeding varieties with enhanced health constituents. *Acta Hort*. 2002;574:215–9.
 77. Gallardo RK, Klingthong P, Zhang Q, Polashock J, Atucha A, Zalapa J, et al. Breeding trait priorities of the cranberry industry in the United States and Canada. *Horts*. 2018;53(10):1467–74.
 78. Koo H, Duarte S, Murata RM, Scott-Anne K, Gregoire S, Watson GE, et al. Influence of cranberry proanthocyanidins on formation of biofilms by streptococcus mutans on saliva-coated apatitic surface and on dental caries development in vivo. *Caries Res*. 2010;44(2):116–26.
 79. Moyer RA, Hummer KE, Finn CE, Frei B, Wrolstad RE. Anthocyanins, phenolics, and antioxidant capacity in diverse small fruits: vaccinium, rubus, and ribes. *J Agric Food Chem*. 2002;50(3):519–25.
 80. Bagchi D, Bagchi M, Stohs SJ, Das DK, Ray SD, Kuszynski CA, et al. Free radicals and grape seed proanthocyanidin extract: importance in human health and disease prevention. *Toxicology*. 2000;148(2):187–97.
 81. Pappas E, Schaich KM. Phytochemicals of cranberries and cranberry products: characterization, potential health effects, and processing stability. *Crit Rev Food Sci Nutr*. 2009;49(9):741–81.
 82. Vvedenskaya IO, Vorsa N. Flavonoid composition over fruit development and maturation in American cranberry *Vaccinium macrocarpon* Ait. *Plant Sci*. 2004;167(5):1043–54.
 83. Wang Y, Johnson-Cicalese J, Singh AP, Vorsa N. Characterization and quantification of flavonoids and organic acids over fruit development in American cranberry (*Vaccinium macrocarpon*) cultivars using HPLC and APCI-MS/MS. *Plant Sci*. 2017;262:91–102.
 84. Wang Y, Han A, Chen E, Singh RK, Chichester CO, Moore RG, et al. The cranberry flavonoids PAC DP-9 and quercetin aglycone induce cytotoxicity and cell cycle arrest and increase cisplatin sensitivity in ovarian cancer cells Corrigendum in. *Int J Oncol*. 2015;46(5):1924–34.

Publisher's Note

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