Running title: High throughput FT-MIR analysis of sugars and acids in watermelon

High throughput FT-MIR indirect analysis of sugars and acids in watermelon

Raúl MARTÍ[a1], Guadalupe SÁNCHEZ[b1], Mercedes VALCÁRCEL[c], Salvador ROSELLÓ[c], Jaime CEBOLLA-CORNEJO[a]*

aUnidad Mixta de Investigación Mejora de la Calidad Agroalimentaria UJI-UPV. COMAV. Universitat Politècnica de València, Cno. De Vera s/n, 46022 València, Spain. E-mail: raumarre@upvnet.upv.es (R.M) jaicecor@btc.upv.es (J.C.C)

bRijk Zwaan Iberica S.A. Ctra. Viator, Pje. El Mamí, S/N. 04120 La Cañada, Almería, Spain. E-mail: g.sanchez@rijkzwaan.es (G.S.)

cUnidad Mixta de Investigación Mejora de la Calidad Agroalimentaria UJI-UPV. Departament de Ciències Agràries i del Medi Natural, Universitat Jaume I, Avda. Sos Baynat s/n, 12071 Castelló de la Plana, Spain. E-mail: valcarce@uji.es (M.V.), rosello@uji.es (S.R.)

[1]*These authors contributed equally to the work*

*Corresponding author

Unidad Mixta de Investigación Mejora de la Calidad Agroalimentaria UJI-UPV. COMAV. Universitat Politècnica de València, Cno. De Vera s/n, 46022 València, Spain.

Phone: +34 3879423; fax: +34 963879422

E-mail: jaicecor@btc.upv.es

Abstract

Indirect measurements of taste-related compounds are required when a high number of samples has to be analyzed in a short period of time, with a minimum cost. For this purpose, FT-MIR partial least square (PLS) regression models for the prediction of total soluble solids, sugars and organic acids have been developed using three sample sets including breeding lines and commercial varieties of watermelon. Specific models with excellent performance were obtained only for sugars. Nevertheless, a general model supposed a compromise between the best and worse models and offered %RMSEP values of 11.3%, 11.1% and 11.7% for fructose, glucose and sucrose respectively. The model was applied to the selection of high content samples (selection pressure 20% and 30%) obtaining good sensitivity levels and mean percentile of selected samples close to the expected values (100% sensitivity). The robustness of FT-MIR models was assessed with predictions of external assays, obtaining reasonable performances.

Abbreviations

FT-MIR, Fourier transform mid-infrared; NIR, near-infrared; PLS, partial least square; R^2_C, coefficient of determination for calibration; R^2_{CV}, coefficient of determination for cross-validation; R^2_P coefficient of determination for prediction; RMSEC, root mean squared error of calibration; RMSECV, root mean squared error of cross-validation; RMSEP, root mean squared error of prediction; %RMSEP, RMSEP contextualized with the mean; %RMSEP (Maximum) RMSEP contextualized with the maximum value; RPD, ratio of prediction to deviation.

Keywords: *Citrullus lanatus* (Thunb.) Matsum. & Nakai; organoleptic quality; fructose; glucose; sucrose; breeding.

Chemical compounds studied in this article: Fructose (PubChem CID: 5984); Glucose (PubChem CID: 5793); Sucrose (PubChem CID: 5988); Citric acid (PubChem CID: 311); Malic acid (PubChem CID: 525); Glutamic acid (PubChem CID: 611).

Highlights

FT-MIR PLS models were developed for taste-related compounds of watermelon

%RMSEP of 11.3% (fructose), 11.1% (glucose) and 11.7% (sucrose) were obtained

General models offer a compromise between best and worst models

General models are robust and can be efficiently applied to selection programs

1. Introduction

The shape, size, color and gloss of a fruit or vegetable represent the first impact on the consumer and thus, the decision to purchase them. But flavor, taste and aroma, has the largest impact on acceptability and desire to consume it again (Barrett, Beaulieu, & Shewfelt, 2010). Although certain consumer segments are mainly driven by price, it seems that perceived quality would be more important than price in determining consumer choice. In fact, consumers are more willing to pay for quality when they are confident on the premium taste that they are going to obtain. Several examples are available on this sense with model produces, including i.a. orange juice (Lange, Issanchou, & Combris, 2000) or apple (Harker, Gunson, & Jaeger, 2003). As a result, market demand for premium quality characterizes modern consumer societies and, in order to reinstate flavor in fruits and vegetables, it is necessary to turn the improvement of flavor into a main objective in breeding programs (Kyriacou & Rouphael, 2018).

In watermelon, *Citullus lanatus* (Thunb.) Natsum. & Nakai a key factor determining its quality and its commercial value is sweetness (Kyriacou, Leskovar, Colla, & Rouphael, 2018). The compounds responsible of this trait are mainly the soluble sugars fructose, glucose, and sucrose, with a total sugar amount that ranges from 24 to 91.0 mg g^{-1} fresh weight (fw) in watermelon cultivars, with a high influence of the environment and the use of different rootstocks (Yoo, Bang, Lee, Crosby, & Patil, 2012). Fructose is usually the main sugar found in ripe fruits with a concentration ranging from 24 to 51 mg g^{-1} fw, followed by sucrose with mean contents of 10 to 42 mg g^{-1} fw, and finally glucose with contents ranging from 9 to 23 mg g^{-1} fw (Fredes et al., 2017; Yoo et al., 2012). Nevertheless, sugar profiles are variable and watermelon genotypes can be divided into high fructose and glucose and high sucrose genotypes (Yativ, Harary, & Wolf, 2010; Yoo et al., 2012). Crosses between watermelon and wild relatives demonstrated a high genetic variability for sugar profile, and it has been suggested that sucrose accumulation is affected by phloem unloading and sugar metabolism, with a high negative correlation with insoluble acid invertase activity and a positive correlation with sucrose synthase activity (Yativ et al., 2010).

Sugar profiles also vary during the ripening. The concentration of fructose and glucose decrease during the maturation, reaching the lowest level at 50 and 45 days post-anthesis respectively. Sucrose concentration

4

follows an inverse trend, increasing its content and reaching a maximum at 45 days post-anthesis (Soteriou, Kyriacou, Siomos, & Gerasopoulos, 2014).

Traditionally, the sweetness has been indirectly measured as the total soluble solids (TSS) content, because a high percentage of soluble solid content are mono and disaccharides (Maynard, 2001). Despite TSS providing a rapid indirect measure of sugar accumulation, it does not reveal the relative content of each soluble sugar. Considering the unequal contribution to sweetness by each sugar, this gross measurement hinders harnessing the genetic potential for sweetness improvement in watermelon. In fact, in other crops, such as tomato, sucrose equivalents calculated as the weighted sum of sugar concentration using the relative sweetening power of each sugar (1 for sucrose, 1.73 for fructose and 0.74 for glucose) has been more closely related to overall acceptability and sweetness perception than TSS (Baldwin et al., 1998).

Organic acids are also present in watermelon, but to a lesser extent. Among them, malic acid accumulation is considerably higher in this species, followed at a distance by citric acid (Çandir, Yetişir, Karaca, & Üstün, 2013; Fredes et al., 2017). Although acid accumulation in watermelon is usually overlooked due to its insipid acidity, its variations may modulate the perception of sweetness (Soteriou et al., 2014). Apart from a genotypic effect, total acidity and the accumulation of citric and malic acid is affected by grafting in watermelon (Fredes et al., 2017; Soteriou et al., 2014).

A reliable quantification of individual sugar and acid accumulation would be then necessary in order to supply high quality markets, either developing breeding programs considering taste improvement, or assessing the effect of environment or different agricultural practices, such as grafting. Instrumental determinations including high performance liquid chromatography (HPLC) coupled with different detectors and capillary electrophoresis can provide such quantifications (Cebolla-Cornejo, Valcárcel, Herrero-Martínez, Roselló, & Nuez, 2012; Cunha, Fernandes, & Ferreira, 2002; Ma, Sun, Chen, Zhang, & Zhu, 2014; Xu, Liang, & Zhu, 2015). However, the time and cost requirements of these methods are not practical for the development of such programs, as a high number of samples must be analyzed in a short period of time, with a minimum cost.

In this context, the interest of indirect determinations using, for example, infrared spectroscopic methods is undeniable. Flores et al. (2008) used near infrared (NIR) spectroscopy as non-invasive method to measure

5

TSS content in melons and watermelons, successfully classifying melons and watermelons into low, medium and high sweetness levels. However, the classification into different categories do not reflect the concentration of individual compounds. Nevertheless, thick rinds as in the case of watermelon have been reported to interfere with the measurement of internal quality using non-destructive NIR methods (Arendse, Fawole, Magwaza, & Opara, 2018).

In other crops, Fourier-transform mid-infrared (FT-MIR) spectroscopy showed better performances than NIR spectroscopy for the indirect quantification of individual compounds such as fructose, glucose, sucrose, total sugar, and citric acid (de Oliveira, de Castilhos, Renard, & Bureau, 2014). This may be justified as MIR spectra provide information from frequencies of fundamental molecular vibrations and it is less sensitive than NIR to factors influencing light diffusion (de Oliveira et al., 2014). FT-MIR has been successfully applied to the quantification of individual sugars in different crops, including i.a. apricot (Bureau et al., 2009), grapes (Barnaba, Bellincontro, & Mencarelli, 2014), peach (Bureau et al., 2013), tomatoes (Ścibisz et al., 2011; Wilkerson et al., 2013) and passion fruit (de Oliveira et al., 2014).

Little information regarding indirect quantification methods is available in the case of watermelon. This was, in fact, the objective of this work: to assess the performance of partial least square (PLS) regression models using FT-MIR spectra to evaluate the content of soluble sugars (fructose, glucose, and sucrose) and organic acids (malic and citric acids) in watermelon in order to provide a rapid an accurate discrimination and indirect quantification of sugar and acid composition in large number of samples, such as those obtained in breeding programs or production controls targeted to provide high quality fruits.

2. Materials and methods

2.1. Plant material and cultivation

Three sample sets were used in the study. The first two sets included samples of breeding lines grown and provided by Rijk Zwaan Ibérica S.A. In this case, cultivation was performed in greenhouse in Paraje el Mamí, Almeria, Spain (36.851988N;-2421842W) in the spring cycle, with a spacing of 1.4m x 1m and following commercial practices in the area. Ripe fruits were harvested at 20-45 days post pollination. The assays were

6

performed during two consecutive years. In the first year (2017) 59 different breeding lines were grown, representing the variability available in watermelon for the development of new varieties. These varieties represented different fruit sizes: 33 belonged to the 1-3 kg group, 18 to 4-8 kg and 8 of them had fruit weight higher than 8 kg. Different skin types were represented including the "Tiger Stripe", "Sugar Baby", "Charleston Gray" and "Crimson Sweet" types. During the second year (2018), 64 breeding lines were grown, including those of 2017 (except one of them) and five new lines. The same skin types were included and the distribution considering fruit weight was the following: 37 in the 1-3 kg group, 19 in the 4-8 kg group and 8 in the >8 kg group.

A third sample set was obtained in 2018. In this case, 60 fruits belonging to the 4-8 kg group were obtained from local markets in order to represent commercial materials of different varieties and areas of production.

2.2. Sampling

From each breeding line, one representative fruit was sampled. Ripe watermelons were collected and a cross-section of 5 cm was obtained from the equatorial plane of each fruit. The edible part was obtained discarding the pericarp and approximately 2mm of flesh. After removing the seeds (if present) the samples were blended in a crusher until they were completely homogeneous and then stored at -80°C until analysis.

Before the analyses sample supernatants were obtained by centrifuging the defrosted samples at 13.000 rpm during 5 min at 4°C to remove any pulp using a microcentrifuge 5415R with fixed-angle rotor F45-24-11 (Eppendorf, Hamburg, Germany). Resulting supernatants were divided into three equivalent aliquots. The first one of them was used for the determination of total soluble solids content (TSS) using an electronic refractometer PAL-1 (ATAGO, Tokyo, Japan) with 0.1° Brix precision. The rest of aliquots were used for FT-MIR analysis and sugars and acids analysis by capillary electrophoresis (CE).

2.3. Analysis of FT-MIR spectra

A portable Cary 630 FTIR spectrometer (Agilent Technologies, Waldbronn, Germany), equipped with a temperature-stabilized DTGS detector, and a 5-bounce ATR crystal with ZnSe beam splitter was used to record the mid-infrared spectral range of 4000 – 700 cm^{-1}. The effective path length at 1700 cm^{-1} with this ATR is 13.0 µm. Data were acquired operating at a spectral resolution of 4 cm-1 using Microlab FTIR Software

7

B.05.3 (Agilent Technologies, Waldbronn, Germany). To improve the signal-to-noise ratio a total of 64 scans of each measure were averaged.

The spectral acquisition was performed in less than 2 min placing 150 µL of sample directly on the crystal (ATR). Spectra were independently measured twice and the crystal was cleaned between samples with distilled water and cellulose tissue.

To perform the models, only data from the fingerprint spectral region from 1500 – 900 cm-1 was included. This spectral region is associated with C-O and C-C stretching modes and O-C-H, C-C-H, and C-O-H bending vibrational modes (Irudayaraj & Tewari, 2003; Stewart, 2004). Moreover, the selection of this spectral region avoids to include the O-H stretching modes from water (Wilkerson et al., 2013).

2.4. Reagents

Analytical grade standards of organic acids (malic acid and citric acid) and sugars (fructose, glucose and sucrose), 2,6-pyridine dicarboxylic acid (PDC), hexadimethrine bromide (HDM) and sodium dodecyl sulphate (SDS) were purchased from Sigma-Aldrich (Syeinheim, Germany). Sodium hydroxide (NaOH) were from Panreac (Castellar del Vallés, Spain). Ultrapure water was obtained using a Milli-Q water system (Millipore, Molsheim, France).

2.5. Analysis of sugar an acid content

Prior to the capillary electrophoresis analysis, the supernatants were diluted 1:20 with ultrapure water and filtered using a 0.22 µm-Nylon centrifuge tube filter (Costar Spin-X, Corning, NY, USA). The quantification of organic acids and sugars was performed using a 7100 CE system equipped with diode array detector and thermostated sample compartment (Agilent Technologies, Waldbronn, Germany). The procedure described by Cebolla-Cornejo et al. (2012) was followed using uncoated fused silica capillaries of 67 cm total length, 60 cm effective length, 375 µm od, 50 µm id (Polymicro Technologies, Phoenix, AZ, USA). Prior its first use, capillaries were flushed with NaOH 1M during 300 s at 50°C, NaOH 0.1M during 300 s, and water during 600 s. Each working session started with a rinse of the capillary with SDS 58mM during 120 s and running buffer during 300 s. The running buffer was prepared using a 20mM PCA solution with 0.1% (w/v) HDM and adjusted

to pH 12.1. A hydrodynamic injection at 3400 Pa during 10 s was used. The voltage applied for separation was -25 kV at 20°C with indirect detection at 214 nm.

2.6. Data analysis

PLS regression models were obtained for each sample set, as well as a general model with the whole set of samples (183). For each model 75% of the samples were used as a calibration group (calibration and cross-validation). The remaining 25% of samples were included in a validation group to obtain predictions using the model. Samples were randomly included in the calibration and prediction groups of the specific models. In order to enable a comparison between the specific models and the general model, the last included in the calibration and validation groups the same samples of the specific models.

PLS calibration models correlating FT-MIR spectral absorbance data (X matrix) and the concentration of measured compounds (Y vector): TSS, malic acid, citric acid, fructose, glucose, sucrose, total sugars, sucrose equivalents, or citric acid equivalents were obtained. Sucrose equivalents were calculated as the weighted sum of sugar concentration using the relative sweetening power of each sugar: 1 for sucrose, 1.73 for fructose and 0.74 for glucose (Baldwin et al., 1998). Citric acid equivalents were calculated as the weight sum of citric and malic acid considering their relative sourness: 1 for citric acid and 1.14 for malic acid (Stevens, Kader, & Algazi, 1977). All Partial Least Square (PLS) models were performed using Matlab v 9.4 (Mathworks Inc, Natick, MA, USA) and the PLS Toolbox 8.2.1 for Matlab (Eigenvector Research Inc, Wenatchee, WA, USA).

Prior to modelling, data set from the X matrix was transformed using the multiplicative scatter correction (MSC) function, while Y matrices were autoscaled (with mean and standard variation). These pretreatments were selected as they improved the performance of the models compared to other alternatives and they had been used in previous studies dealing with MIR spectra (Wilkerson et al., 2013). The spectral data were then analyzed by PLS using Venetian blinds as cross-validation method. Resulting calibration model performance was evaluated in terms of outlier diagnostics, the number of latent variables (LV), coefficient of determination of calibration (R^2_C), root mean squares error calibration (RMSEC), and root mean squares error

9

of cross-validation (RMSECV). Selection criteria used to choose the most suitable calibration model was focused on minimizing the RMSECV and number of LV and maximizing R^2 values (Wilkerson et al., 2013).

Outlier identification was performed using a graphical evaluation of Q residuals and leverage. Any outlier point that showed a large Q residual or unusual distribution was removed and the model was recalculated. Normalized residuals and leverage parameters were also considered for outlier identification (values <-3 or >3) and elimination in response variables.

Once the final model was constructed, prediction matrices were used to evaluate the expected error when the model was applied to predict new samples (validation group). For that purpose, root mean squared errors of prediction (RMSEP) and correlation coefficient of prediction (R^2_P) were calculated. %RMSEC and %REMSEP were also calculated as a percentage of the mean values of each group in order to contextualize the results. %RMSEP (maximum) was calculated using the maximum value to provide a reference for selection programs considering high content samples.

To evaluate the predictive capacity of the model, residual prediction deviation (RPD) was calculated as the ratio between the standard deviation of the prediction group and RMSEP. This ratio is an adaptation of the original RPD description which uses SEP instead of RMSEP; the use of RMSEP offers lower RPD value, thus representing a more conservative approach. The models are usually considered useful when RPD values are higher than 2 (Fearn, 2002).

3. Results and discussion

3.1. Variation present in the samples used for modelling

The TSS values obtained in the samples of the breeding lines grown in 2017 ranged between 5°Brix and 12.7°Brix and those of breeding lines of 2018 between 9.5°Brix and 14.1°Brix (Table 1). The commercial materials from 2018 ranged from 4.6°Brix to 11.7°Brix. Commercial high-quality watermelons are supposed to be included in the range of 10-14°Brix (Wehner, 2008) though the lower limit can be as low as 4-5% in specific materials (Yoo et al., 2012). Therefore, the usual commercial range of variation was covered with the different sample groups used in the study.

10

Similar mean values were obtained in both years of fructose and sucrose contents in the breeding lines (Table 1). The accumulation of glucose was much lower than the rest of sugars. Although higher contents of fructose and glucose were found in the breeding lines during the 2018 season, the difference was especially important in the case of glucose (Table 1). Samples from the second year tended to have higher ratios of glucose to fructose and glucose to sucrose. Another important difference between both years was that samples from 2018 tended to have higher citric acid (Table 1).

Table 1. Statistical parameters of the sample sets used for the calibration and validation of PLS models. Contents expressed in mg g^{-1} of fresh weight, except Total soluble solids (TSS) expressed in °Brix. N: Number of samples. RSD: Relative standard deviation.

Sample set	Parameters	Calibration				Validation			
		N	Mean	RSD (%)	Range	N	Mean	RSD (%)	Range
Breeding lines 2017	TSS		9.2	19.3	5.0 – 12.7		8.7	19.6	5.0 – 12.2
	Fructose		30.78	25.3	18.78 – 52.40		29.46	24.4	15.41 – 42.52
	Glucose		13.68	43.3	5.48 – 31.03		10.92	33.6	5.77 – 18.09
	Sucrose		31.51	61.0	0.00 – 61.95		25.94	54.3	2.28 – 54.08
	Malic acid	45	1.97	20.5	1.00 – 3.21	14	1.98	21.7	1.04 – 2.71
	Citric acid		0.45	62.0	0.09 – 1.16		0.39	67.2	0.09 – 1.05
	Total sugars		75.97	21.6	42.79 – 102.96		66.33	26.1	39.59 – 102.08
	Sucrose equivalents		94.88	18.7	57.13 – 124.32		84.99	23.7	53.94 – 121.59
	Citric acid equivalents		2.69	20.7	1.56 – 4.42		2.65	20.9	1.46 – 3.76
Breeding lines 2018	TSS		11.7	9.2	9.5 – 14.1		12.0	10.2	8.8 – 14.2
	Fructose		33.62	22.9	17.77 – 55.44		37.65	23.1	24.89 – 52.46
	Glucose		18.57	36.1	7.32 – 37.83		22.14	35.0	9.66 – 35.99
	Sucrose		33.31	44.3	6.17 – 69.65		28.01	43.5	9.18 – 51.64
	Malic acid	48	1.76	24.6	0.77 – 3.43	16	1.73	35.3	0.82 – 2.68
	Citric acid		0.93	38.6	0.32 – 1.71		0.97	44.3	0.28 – 2.01
	Total sugars		85.50	15.0	58.94 – 116.14		87.80	16.1	58.00 – 114.12
	Sucrose equivalents		105.22	14.1	72.72 – 146.77		109.53	16.0	71.57 – 133.40
	Citric acid equivalents		2.94	20.4	1.97 – 5.28		2.94	24.2	1.63 – 4.36
Commercial materials 2018	TSS		9.43	13.9	4.6 – 11.7		9.4	9.7	8.2 – 11.4
	Fructose		24.50	20.0	11.13 – 32.36		23.34	16.1	19.31 – 32.51
	Glucose		21.14	18.7	8.99 – 31.37		20.76	19.0	14.71 – 28.31
	Sucrose		22.38	58.3	0.00 – 53.47		27.12	46.9	9.15 – 53.57
	Malic acid	45	2.07	27.9	1.30 – 3.71	15	1.92	27.7	1.11 – 2.85
	Citric acid		0.69	47.9	0.08 – 1.48		0.75	40.2	0.30 – 1.35
	Total sugars		68.02	22.8	20.12 – 96.11		71.22	19.3	49.02 – 90.90
	Sucrose equivalents		80.41	21.3	25.91 – 108.51		82.86	18.3	58.90 – 103.83
	Citric acid equivalents		3.06	24.0	1.83 – 3.06		2.94	27.0	1.57 – 3.93
General model	TSS		10.2	17.9	4.6 – 14.1		10.1	19.1	5.0 – 14.2
	Fructose		29.72	26.5	11.13 – 55.44		30.33	29.9	15.41 – 52.46
	Glucose		17.81	36.0	5.48 – 37.83		18.19	40.4	5.77 – 35.99
	Sucrose	138	29.16	56.5	0.00 – 69.65	45	27.07	46.9	2.28 – 54.08
	Malic acid		1.93	25.4	0.77 – 3.71		1.87	28.5	082 – 2.85
	Citric acid		0.70	54.4	0.08 – 1.71		0.72	57.3	0.09 – 2.01
	Total sugars		76.69	21.5	20.12 – 116.14		75.59	23.1	39.59 – 114.12
	Sucrose equivalents		93.76	20.7	25.91 – 146.77		93.01	22.9	53.94 – 133.40
	Citric acid equivalents		2.90	22.3	1.56 – 5.28		2.85	24.4	1.46 – 4.36

The range of variation for fructose, glucose and glucose contents of the commercial materials was narrower and, in general, it was included in the range of variation of the breeding lines (Table 1). Though a lower minimum fructose content was registered in the commercial materials. In the case of the acids, commercial materials reached higher accumulation of malic acid and lower accumulation of citric acid.

Considering all the samples, that is the general model, the range of fructose accumulation (11.13-55.44 mg g^{-1}) exceeded the limits of previously reported variation, between 24.0 and 51.0 mg g^{-1} (Fredes et al., 2017; Yoo et al., 2012). A wider difference between the range of variation in the present study and previous reports

was obtained for the ranges of accumulation of glucose (5.48-37.83 mg g^{-1} vs. 9.0-23.0 mg g^{-1}) and sucrose (0.00-69.65 mg g^{-1} vs. 10.0-42.0 mg g^{-1}). In the case of the acid content, the range of variation represented in the general model for citric acid (0.08-1.71 mg g^{-1}) was wider than that observed in previous literature (Çandir et al., 2013; Fredes et al., 2017) while a for malic acid minimum and maximum contents (0.77-3.71 mg g^{-1}) were lower than previously reported contents (1.63-4.57 mg g^{-1}).

As regards the level of variation, for the three sample sets and the general model, the %RSD for fructose was, in general, considerably lower than that for glucose and sucrose, especially in the breeding lines. The highest levels of variation were always detected for sucrose. A similar imbalance in the %RSD applied to the organic acids, with higher levels of variation detected for citric acid, the one with lower mean contents.

The statistical parameters of the samples of the validation group had similar values to those of the calibration group of each sample set.

A correlation analysis was performed among the variables (Supplementary Table 1). As expected, total soluble solids were highly correlated with total sugar accumulation. Among the sugars a higher correlation was observed between TSS and sucrose. Unexpectedly, a high correlation (0.71) was found between citric acid contents and TSS. Consequently, moderate correlations were found between this acid and total sugars. Malic acid showed low or non-significant correlations with the rest of variables. Glucose and fructose showed a positive moderate correlation (0.51), and both of them a negative correlation with sucrose accumulation. This is an expected result. As reviewed by Soteriou et al. (2014), during the initial development of watermelons α-galactosidase and acid invertases keep increasing contents of reducing sugars and restrict sucrose accumulation, while later sucrose accumulates at the expense of fructose and glucose due to an increased activity of sucrose phosphate synthase and sucrose synthase and a reduced activity of soluble acid invertase.

3.2. PLS regression models.

The model obtained for the 2017 breeding lines was excellent (Table 2). Coefficients of determination for calibration ranged from 0.94 for malic acid to 1.00 for sucrose, with %RMSEC values ranging from 1.5% for sucrose to 9.9% for citric acid. RMSECV values obtained in the crossvalidation were on average 2.7 times

13

higher. RMSEP values were in average 3.2 times higher than RMSEC, but still %RMSEP represented 8.5%, 13.9% and 10.1% of fructose, glucose and sucrose mean contents, and R^2_P values ranged from 0.86 for glucose to 0.97 for sucrose. The performance of acid prediction was notably worse, with R^2_P values for prediction lower than 0.32 and %RMSEP values between 17.8% for malic and 54.5% for citric acid. These values led to obtain RPD values higher than 2 for sugars and TSS, thus appropriate for screening purposes (Fearn, 2002), and lower for acids. %RMSEP values considering the maximum value, thus the probable target of a breeding program, were a 40% lower on average.

The performance of the model for breeding lines grown in 2018 was similar (Table 2). R^2_C values for calibration ranged from 0.91 to 1.00 and %RMSEC values from 1% or fructose to 8.6% for citric acid. Cross validation was tougher in this case, and RMSECV values were on average 5.1 times higher than RMSEC. Regarding predictions, the performance of the model was somewhat lower considering R^2_P values for TSS and sugars (0.81-0.91), though much better for acids (0.65-0.73). %RMSEP values for TSS and sugars were similar to those obtained for the first model, ranging from 4.5% for TSS to 14.3% for sucrose. Again, RPD values were higher than 2 for TSS and sugars and slightly lower than 2 for citric and malic acids.

The model for commercial materials offered a similar performance, with better results for TSS and sugars than for acids (Table 2). In this case, though the R^2_P for fructose was low (0.44), the %RMSEP was similar to the previous models. In the case of sugars, %RMSEP ranged from 8.8% for glucose to 11.7% for fructose. RPD values were higher than 2 for all the compounds except for fructose (1.4) and malic acid (1.5).

For the three specific models, a PLS models were also obtained for total sugars, sucrose equivalents and citric acid equivalents. In the case of total sugars and sucrose equivalents, %RMSEP values obtained were similar to the one corresponding to the best individual compound, while in the case of citric acid equivalents %RMSEP values were better than for individual compounds. These results are interesting, as using these values a selection can be performed with an error level lower than the one that would be obtained combining the error of the individual compounds that participate in their calculation.

A general model was calculated to compare the performance of a model with a higher number of data and wider levels of variation with the specific models. R^2_P values obtained with the general model were similar to

those obtained with the best specific model (Table 2). The %RMSEP values were on average a 30% higher than the best model. Specifically, the predictions with the general model were worse for fructose and citric acid compared to the best model. Alternatively, the performance was better than the worst specific model (%RMSEP on average a 12% lower), especially for glucose, sucrose and citric acid. For the general model, RPD values were higher than 2.5 for TSS and sugars and lower for citric and malic acid. The RPD values for total sugars and sucrose equivalents were also higher than 2. It is possible to say then, that a general model would be preferred to specific models, as it represents a compromise between the best and the worst specific models and it would have a wider applicability. The performance of this general model was especially good for sugars, the main target of breeding programs and the commercialization of watermelon, as %RMSEP values for fructose, glucose and sucrose were 11.3%, 11.1% and 11.7% respectively.

In order to try to improve the resulting models, loadings of the PLS model were reviewed (Supplementary Fig. 1), and new models were developed with those wavelengths with higher absolute loadings. With a similar purpose, reverse interval PLS models were also developed. In both cases, the performance of the models obtained were lower and no improvement was achieved with these model variations. Models considering a wider spectra were also considered. Wilkerson et al. (2013) in their work with tomato also considered including the wavelengths between 1800 cm^{-1} and 1500 cm^{-1} for certain models, but the authors already alerted about the high absorption of water in this region. The addition of these wavelengths did not improve the models. Additionally, it was considered the possibility of removing the pretreatment of the spectra with the MSC correction, but this pretreatment increased considerably the performance of the models for organic acids. On the other hand, the possible noise of the CE analytical procedure on the performance of the models was discarded, as mean %RSD for the determinations for citric and malic acids and fructose, glucose and sucrose were respectively 2.4%, 1.6%, 1.2%, 1.3% and 1.0%.

FT-MIR has been applied to the analysis of more than 40 genera of plant species, but most works have dealt with authenticating products and identifying adulteration (Bureau, Cozzolino, & Clark, 2019). Nevertheless, FT-MIR has been applied satisfactorily to predict individual sugar and acid contents in different species. For example, Bureau et al. (2009) developed prediction models for apricot using eight varieties, obtaining R^2_P

15

values ranging from 0.74 for fructose to 0.97 for malic acid, with %RMSEP values about 12% for glucose, malic acid and citric acid, 16% for sucrose, and 18% for fructose. Barnaba et al. (2014) working with a single grape variety "Sangiovese", obtained R^2_P values of 0.93 and 0.92 for glucose and fructose respectively, with %RMSEP values of 4.5% and 5.5%, which represented RPD values around 2.5. In this case, the prediction for malic acid was notably worse, with an RPD value of 1.15. In peach, Bureau et al. (2013), obtained %RMSEP values of 20.2% for sucrose, 13% for glucose, 19% for fructose, 16.4% for malic acid, and 25.7% for citric acid (which had a lower content compared to malic acid). In tomato, excellent models have been obtained for a wide diversity of varieties. Ścibisz et al. (2011) obtained %RMSEP values for glucose and fructose of 5.1% and 6.8% respectively. Working with processing tomatoes, Wilkerson et al. (2013) obtained %RMSEP (%maximum) values of 6.9% for glucose and 6.0% for fructose. In passion fruit de Oliveira et al. (2014) obtained %RMSEP values of 8.7%, 8.3% and 11.7% for glucose, fructose and sucrose respectively. Models with lower performance have also been obtained. For example, with samples of barley malt from six commercial varieties, Huang, Carragher, & Cozzolino (2016) obtained %SEP values of 19.3% for glucose and 31.8% for sucrose, with RPD values of 1.7.

Taken together, the performance of the predictions made in the present work in watermelon can be considered as excellent, considering previous results in other crops. In the case of the acids, the performance decreases, especially in the case of citric acid. But in this case, it should be considered that MIR spectroscopy is generally regarded insensitive to compounds present at concentrations lower than 1 mg g^{-1} (Bureau et al., 2019), as it is the case for this acid in several of the analyzed samples. In fact, other works in tomato using FT-MIR spectra revealed a lower performance in the case of acids, which has been related to their lower concentrations (Ścibisz et al., 2011; Ibañez et al., 2019).

Table 2. Performance of the partial least squares (PLS) regression models predicting taste-related compounds content from ATR FT-MIR spectra. TSS: total soluble solids; LV: number of latent variables; R^2: coefficient of determination; RMSE: root mean squared error; C: calibration; CV: cross-validation; P: prediction; RPD: residual prediction deviation.

Model	Parameters	LV	R^2_C	RMSEC mg g^{-1}	%RMSEC (mean)	RMSECV mg g^{-1}	R^2_P	RMSEP mg g^{-1}	%RMSEP (mean)	%RMSEP (maximum)	RPD
Breeding lines 2017	TSS	5	0.99	0.2	2.0	0.3	0.98	0.3	3.0	2.1	6.6
	Fructose	6	0.97	1.33	4.3	2.68	0.89	2.52	8.5	5.9	2.9
	Glucose	5	0.98	0.86	6.3	1.18	0.86	1.51	13.9	8.4	2.4
	Sucrose	7	1.00	0.48	1.5	2.02	0.97	2.62	10.1	4.9	5.4
	Malic acid	7	0.94	0.10	4.9	0.26	0.29	0.35	17.8	13.0	1.2
	Citric acid	7	0.97	0.04	9.9	0.19	0.32	0.21	54.5	20.5	1.2
	Total sugars	5	0.96	3.40	4.5	4.53	0.88	6.02	9.1	5.9	2.9
	Sucrose equivalents	5	0.95	4.00	4.2	6.10	0.56	7.94	9.4	6.5	2.5
	Citric acid equivalents	8	0.98	0.08	3.1	0.34	0.29	0.46	17.4	12.2	1.2
Breeding lines 2018	TSS	7	0.98	0.2	1.3	0.5	0.81	0.5	4.5	3.8	2.3
	Fructose	10	1.00	0.35	1.0	3.71	0.87	3.04	8.1	5.8	2.9
	Glucose	9	1.00	0.40	2.2	1.47	0.91	2.22	10.1	6.2	3.5
	Sucrose	8	1.00	0.98	2.9	3.05	0.90	4.02	14.3	7.8	3.0
	Malic acid	7	0.91	0.13	7.5	0.39	0.73	0.33	19.1	12.3	1.9
	Citric acid	8	0.95	0.08	8.6	0.15	0.65	0.25	25.7	12.4	1.7
	Total sugars	8	0.97	2.16	2.5	7.16	0.72	7.48	8.5	6.6	1.9
	Sucrose equivalents	8	0.97	2.64	2.5	9.47	0.75	8.87	8.1	6.7	2.0
	Citric acid equivalents	8	0.94	0.14	4.9	0.43	0.69	0.39	13.2	8.9	1.8
Commercial materials 2018	TSS	5	0.96	0.3	2.8	0.5	0.79	0.5	5.0	4.2	1.9
	Fructose	4	0.55	3.27	13.3	3.71	0.44	2.73	11.7	8.4	1.4
	Glucose	5	0.94	0.98	4.7	1.47	0.82	1.82	8.8	6.5	2.2
	Sucrose	4	0.96	2.62	11.7	3.05	0.98	2.53	9.4	4.7	5.0
	Malic acid	6	0.85	0.22	10.6	0.39	0.54	0.36	18.6	12.5	1.5
	Citric acid	6	0.92	0.09	13.1	0.15	0.83	0.13	17.7	9.9	2.3
	Total sugars	4	0.83	6.31	9.3	7.16	0.81	6.16	8.7	6.8	2.2
	Sucrose equivalents	4	0.96	8.33	10.4	9.47	0.76	7.46	9.0	7.2	2.0
	Citric acid equivalents	6	0.88	0.25	8.3	0.43	0.67	0.46	15.7	11.7	1.7
General model	TSS	6	0.98	0.2	2.3	0.3	0.96	0.4	4.0	2.9	4.8
	Fructose	6	0.80	3.50	11.8	3.98	0.85	3.44	11.3	6.6	2.6
	Glucose	6	0.95	1.44	8.1	1.70	0.92	2.02	11.1	5.6	3.6
	Sucrose	6	0.98	2.50	8.6	2.93	0.94	3.78	11.7	5.9	4.0
	Malic acid	7	0.71	0.26	13.6	0.36	0.56	0.35	18.8	12.4	1.5
	Citric acid	8	0.92	0.11	15.2	0.21	0.66	0.24	33.4	12.0	1.7
	Total sugars	5	0.79	7.56	9.9	7.87	0.82	0.75	9.7	6.4	2.4
	Sucrose equivalents	5	0.77	8.95	9.6	9.34	0.81	9.38	10.1	7.0	2.3
	Citric acid equivalents	7	0.76	0.31	10.8	0.43	0.67	0.40	13.9	9.1	1.8

3.3. Applicability of PLS models: selection pressure and external assays.

One of the main applications of indirect prediction of individual compounds in watermelon is related with the development of breeding programs, in which a high number of samples is to be processed in the shorter period of time possible, and with minimum costs. In this case, it would be interesting to know if the models are reliable when applied to select the samples with the highest contents, given a certain selection pressure. For this purpose, the general model was applied to identify the 20% or 30% of the samples of the validation group with the highest values (Table 3). The values of sensitivity (true positive rate) for a selection pressure of 30% were higher than 90% for individual sugars, with specificities (true negative rate) higher than 95% (Table 3). Sensitivities for acids were higher than 70% and for derived variables higher than 75%. More interestingly, the observed mean percentile of selected samples were close to the expected mean percentile considering 100% sensitivity. That means that those samples that were wrongly selected presented high contents close to the threshold of selection. With a selection pressure of 20% the sensitivities were somewhat lower, but probably due to the small number of samples considered with this pressure. It is important to note that the sensitivity for sucrose was 100% and mean percentile for fructose and glucose were 11.4% and 15.6%, again really close to the expected value for 100% sensitivity (11.1%).

In order to check the robustness of FT-MIR predictions, two new general models were calculated using the whole number of samples of one of the sample sets of breeding lines and the set for commercial materials. These new models were then applied to predict the contents of the remaining sample set for breeding lines. This is a really tough test, as it involves the prediction of a high number of samples of an external assay, and it is rarely found in the literature. This test will evaluate if internal calibrations are required to develop specific models for each assay, increasing the cost of analysis, or if reliable general models can be applied without assay-specific calibrations.

Table 3. Values of sensitivity and specificity the selection of samples of the validation group for high content applying a selection pressure of 20% or 30% with the general PLS model. Mean percentiles (observed and expected for 100% sensitivity) of selected plants are also provided.

General model	20% Selection pressure				30% Selection pressure			
	Sensitivity (%)	Specificity (%)	Observed mean percentile (%)	Expected mean percentile (%)	Sensitivity (%)	Specificity (%)	Observed mean percentile (%)	Expected mean percentile (%)
TSS	88.9	97.2	11.9	11.1	92.9	96.8	17.1	16.7
Fructose	88.9	97.2	11.4	11.1	92.9	96.8	18.9	16.7
Glucose	66.7	91.7	15.6	11.1	92.9	96.8	18.1	16.7
Sucrose	100	100	11.1	11.1	92.9	96.8	16.8	16.7
Malic acid	66.7	91.7	17.3	11.1	71.4	87.1	23.5	16.7
Citric acid	66.7	91.7	16.0	11.1	71.4	87.1	22.7	16.7
Total sugars	66.7	91.7	15.3	11.1	85.7	93.5	21.1	16.7
Sucrose equivalents	77.8	94.4	16.3	11.1	85.7	93.5	18.6	16.7
Citric acid equivalents	66.7	91.7	15.8	11.1	78.6	90.3	20.3	16.7

The models developed offered a lower performance compared to the previous general model (Table 4). This result is expected, as a principal component analysis of response variables showed that the range of variation of the three sample sets was rather different, both from the point of view of FT-MIR spectra (Fig. 1) and sugar and acid contents (Fig. 2), especially in the case of breeding lines grown in 2018.

It seems evident, then, the necessity to use a high number of samples representing the widest variability possible to reinforce the robustness of FTIR models. But despite this limitation, the %RMSEP obtained for sugars in the prediction of the breeding lines grown in 2018 ranged from 12.6% for sucrose and 16% for glucose. In the case of the predictions of sugars for 2017, %RMSEP values ranged from 10.1% for sucrose and 28.3% for glucose. These figures prove the reliability of the methodology, considering its indirect nature and the application to assays not included in the calibration.

19

Table 4. Characteristics of the predictions made for one of the samples sets with breeding lines using the remaining set and the set for commercial varieties for calibrating new models. R^2_P: coefficient of determination of the predictions; RMSEP: root mean squared error of prediction; %RMSEP: RMSEP contextualized with mean value of the validation group.

Model	Validation	Parameter	R^2_P	RMSEP	%RMSEP
Breeding lines 2017 and commercial varieties (119 samples)	Breeding lines 2018 (64 samples)	TSS	0.82	0.5	4.3
		Fructose	0.71	43.85	12.7
		Glucose	0.87	31.20	16.0
		Sucrose	0.93	40.23	12.6
		Malic acid	0.48	5.71	32.5
		Citric acid	0.23	3.29	35.0
		Total sugars	0.45	98.30	11.4
		Sucrose equivalents	0.39	123.69	11.6
		Citric acid equivalents	0.50	6.71	22.8
Breeding lines 2018 and commercial varieties (124 samples)	Breeding lines 2017 (59 samples)	TSS	0.98	0.4	3.9
		Fructose	0.78	68.57	22.5
		Glucose	0.92	36.81	28.3
		Sucrose	0.98	30.48	10.1
		Malic acid	0.33	3.97	20.1
		Citric acid	0.57	3.30	75.6
		Total sugars	0.89	122.19	16.6
		Sucrose equivalents	0.84	195.93	21.2
		Citric acid equivalents	0.47	4.50	16.8

4. Conclusions

FT-MIR based PLS regression models in watermelon can offer precise indirect predictions of gross variables such as TSS and of the individual accumulation of specific sugars (fructose, glucose, and sucrose). The error in the prediction of acids, is considerably higher, especially for citric acid, found at low concentrations in this crop. The performance of the models depends on the specific material included in the calibration. Therefore, the use of general models is recommended as they offer a compromise between the performance of the best and worst models and can be applied to a wider range of variation. These general models can be satisfactorily applied to selection programs, with high sensitivity levels. Even those samples wrongly selected would have high contents close to the threshold of selection. General models are robust and can be applied to the prediction of external assays not included in the calibration. Nevertheless, in such cases the use of a wide range of variation is suggested in the development of general models.

Conflict of interest

Acknowledgements

5. References

Arendse, E., Fawole, O. A., Magwaza, L. S., & Opara, U. L. (2018). Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. *Journal of Food Engineering*, *217*, 11–23.

Baldwin, E. A., Scott, J. W., Einstein, M., Malundo, T. M. M., Carr, B. T., Shewfelt, R. L., & Tandon, K. S. (1998). Relationship between sensory and instrumental analysis for tomato flavor. *Journal of the American Society for Horticultural Science*, *123*, 906–915.

Barnaba, F. E., Bellincontro, A., & Mencarelli, F. (2014). Portable NIR-AOTF spectroscopy combined with winery FTIR spectroscopy for an easy, rapid, in-field monitoring of Sangiovese grape quality. *Journal of the Science of Food and Agriculture*, *94*, 1071–1077.

Barrett, D. M., Beaulieu, J. C., & Shewfelt, R. (2010). Color, Flavor, Texture, and Nutritional Quality of Fresh-Cut Fruits and Vegetables: Desirable Levels, Instrumental and Sensory Measurement, and the Effects of Processing. *Critical Review of Food Science and Nutrition*, *50*, 369–389.

Bureau, S., Cozzolino, D., & Clark, C. J. (2019). Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biology and Technology*, *148*, 1–14.

Bureau, S., Quilot-Turion, B., Signoret, V., Renaud, C., Maucourt, M., Bancel, D., & Renard, C. M. G. C. (2013). Determination of the Composition in Sugars and Organic Acids in Peach Using Mid Infrared

Spectroscopy: Comparison of Prediction Results According to Data Sets and Different Reference Methods. *Analytical Chemistry*, *85*, 11312–11318.

Bureau, S., Ruiz, D., Reich, M., Gouble, B., Bertrand, D., Audergon, J., & Renard, C. M. G. C. (2009). Application of ATR-FTIR for a rapid and simultaneous determination of sugars and organic acids in apricot fruit. *Food Chemistry*, *115*, 1133–1140.

Çandir, E., Yetişir, H., Karaca, F., & Üstün, D. (2013). Phytochemical characteristics of grafted watermelon on different bottle gourds (Lagenaria siceraria) collected from the Mediterranean region of Turkey. *Turkish Journal of Agriculture and Forestry*, *37*, 443–456.

Cebolla-Cornejo, J., Valcárcel, M., Herrero-Martínez, J. M., Roselló, S., & Nuez, F. (2012). High efficiency joint CZE determination of sugars and acids in vegetables and fruits. *Electrophoresis*, *33*, 2416–2423.

Cunha, S. C., Fernandes, J. O., & Ferreira, I. M. P. L. V. O. (2002). HPLC/UV determination of organic acids in fruit juices and nectars. *European Food Research and Technology*, *214*, 67–71.

de Oliveira, G. A., de Castilhos, F., Renard, C. M. G. C., & Bureau, S. (2014). Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Research International*, *60*, 154–162.

Fearn, T. (2002). Assessing Calibrations: SEP, RPD, RER and R2. *NIR News*, *13*, 12–13.

Flores, K., Sánchez, M. T., Pérez-Marín, D. C., López, M. D., Guerrero, J. E., & Garrido-Varo, A. (2008). Prediction of total soluble solid content in intact and cut melons and watermelons using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, *16*, 91–98.

Fredes, A., Roselló, S., Beltrán, J., Cebolla-Cornejo, J., Pérez-de-Castro, A., Gisbert, C., & Picó, M. B. (2017). Fruit quality assessment of watermelons grafted onto citron melon rootstock. *Journal of the Science of Food and Agriculture*, *97*, 1646–1655.

Harker, F. R., Gunson, F. A., & Jaeger, S. R. (2003). The case for fruit quality: An interpretive review of consumer attitudes, and preferences for apples. *Postharvest Biology and Technology*, *28*, 333–347.

Huang, Y., Carragher, J., & Cozzolino, D. (2016). Measurement of Fructose, Glucose, Maltose and Sucrose in Barley Malt Using Attenuated Total Reflectance Mid-infrared Spectroscopy. *Food Analytical Methods*,

9, 1079–1085.

Ibáñez, G., Valcárcel, M., Cebolla-Cornejo, J., & Roselló, S. (2019). FT-MIR determination of taste-related compounds in tomato: a high throughput phenotyping analysis for selection programs. Journal of the Science of Food and Agriculture, 99, 5140-5148.

Irudayaraj, J., & Tewari, J. (2003). Simultaneous Monitoring of Organic Acids and Sugars in Fresh and Processed Apple Juice by Fourier Transform Infrared–Attenuated Total Reflection Spectroscopy. *Applied Spectroscopy*, *57*, 1599–1604.

Kyriacou, M. C., Leskovar, D. I., Colla, G., & Rouphael, Y. (2018). Watermelon and melon fruit quality: The genotypic and agro-environmental factors implicated. *Scientia Horticulturae*, *234*, 393–408.

Kyriacou, M. C., & Rouphael, Y. (2018). Towards a new definition of quality for fresh fruits and vegetables. *Scientia Horticulturae*, *234*, 463–469.

Lange, C., Issanchou, S., & Combris, P. (2000). Expected versus experienced quality: trade-off with price. *Food Quality and Preference*, *11*, 289–297.

Ma, C., Sun, Z., Chen, C., Zhang, L., & Zhu, S. (2014). Simultaneous separation and determination of fructose, sorbitol, glucose and sucrose in fruits by HPLC-ELSD. *Food Chemistry*, *145*, 784–788.

Maynard, D. N. (2001). *Watermelons: characteristics, production, and marketing*. Alexandria, VA, USA: American Society for Horticultural Science Press.

Ścibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, MathildeBertrand, D., & Renard, C. M. G. C. (2011). Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. *Food Chemistry*, *125*, 1390–1397.

Soteriou, G. A., Kyriacou, M. C., Siomos, A. S., & Gerasopoulos, D. (2014). Evolution of watermelon fruit physicochemical and phytochemical composition during ripening as affected by grafting. *Food Chemistry*, *165*, 282–289.

Stevens, M. A., Kader, A. A., & Algazi, M. (1977). Genotypic variation for flavor and composition in fresh market tomatoes. *Journal of the American Society for Horticultural Science*.

Stewart, B. (2004). *Infrared Spectroscopy: Fundamentals and Applications*. Sussex, United Kingdom: John

Wiley & Sons, Ltd.

Wehner, T. C. (2008). Watermelon. In *Vegetables I* (pp. 381-418). New York: Springer.

Wilkerson, E. D., Anthon, G. E., Barrett, D. M., Sayajon, G. F. G., Santos, A. M., & Rodriguez-Saona, L. E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *Journal of Agricultural and Food Chemistry*, *61*, 2088–2095.

Xu, W., Liang, L., & Zhu, M. (2015). Determination of sugars in molasses by HPLC following solid-phase extraction. *International Journal of Food Properties*, *18*, 547–557.

Yativ, M., Harary, I., & Wolf, S. (2010). Sucrose accumulation in watermelon fruits: Genetic variation and biochemical analysis. *Journal of Plant Physiology*, *167*, 589–596.

Yoo, K. S., Bang, H., Lee, E. J., Crosby, K., & Patil, B. S. (2012). Variation of carotenoid, sugar, and ascorbic acid concentrations in watermelon genotypes and genetic analysis. *Horticulture Environment and Biotechnology*, *53*, 552–560.
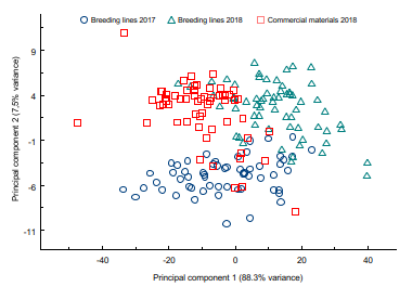
Fig. 1. Scores of populations obtained in the principal component analysis of the FT-MIR spectra.

25

Fig. 2. Biplot of principal component analysis of the response variables.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: