Ibañez, G.; Valcárcel-Germes, M.; Cebolla Cornejo, J.; Rosello Ripolles, 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(11):5140-5148. https://doi.org/10.1002/jsfa.9760

FT-MIR determination of taste-related compounds in tomato: a high throughput phenotyping analysis for selection programs

Running title: FT-MIR determination of taste-related compounds in tomato

Ginés Ibáñez[a], Mercedes Valcárcel [a], Jaime Cebolla-Cornejo[b*], Salvador Roselló[a]

[a] Unidad Mixta de Investigación Mejora de la Calidad Agroalimentaria UJI-UPV. Department de Ciències Agràries i del Medi Natural, Universitat Jaume I, Avda. Sos Baynat s/n, 12071 Castelló de la Plana, Spain

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

*Corresponding author: jaicecor@btc.upv.es; Tel.: +34-963879423

ORCID codes: G. Ibáñez: 0000-0002-1787-8587; M. Valcárcel: 0000-0002-9347-1500; J. Cebolla-Cornejo: 0000-0002-2607-9920; S. Roselló: 0000-0002-7733-4178

Abstract

BACKGROUND: Tomato taste is defined by the accumulation of sugars and organic acids. Individual analyses of these compounds using HPLC or CZE are expensive, time-consuming and are not feasible for large number of samples, justifying the interest of spectroscopic methods such as Fourier transform mid-infrared (FT-MIR). This work analysed the performance of FT-MIR models to determine the accumulation of sugars and acids, considering the efficiency of models obtained with different ranges of variation.

RESULTS: FT-MIR spectra (five-bounce attenuated total reflectance, ATR) were used to obtain PLS models to predict sugar and acid contents in specific sample sets representing different varietal types. A general model was also developed, obtaining R^2 values for prediction higher than 0.84 for main components (SSC, fructose, glucose, and citric acid). Root mean squared error of prediction RMSEP for these components were lower than 15% of the mean contents and lower than 6% of the highest contents. Even more, the model sensitivity and specificity for those variables with a 10% selection pressure was 100%. That means that all samples with the 10% highest content were correctly identified. The model was applied to an external assay and it exhibited, for main components, high sensitivities (>70%) and specificities (>96%). RMSEP values for main compounds were lower than 21% and 13% of the mean and maximum content respectively.

CONCLUSION: The models obtained confirm the effectiveness of FT-MIR models to select samples with high contents of taste-related compounds, even when the calibration has not been performed within the same assay.

Keywords: *Solanum lycopersicum* L., sugars, organic acids, plant breeding, quality, FTIR

INTRODUCTION

Consumer complaints about tomato taste became commonplace during the nineties.[1] This discontent helped to consolidate emerging quality markets associated with tomato landraces with recognized organoleptic quality.[2] Several causes explain the loss of organoleptic quality in modern varieties, including the way tomatoes are produced (early harvesting, high nitrogen fertilization, greenhouse cultivation…), conserved (i.e. refrigeration) or selected in breeding programs.[3–6] This last cause had a strong influence. In part, it was due to negative collateral effects of genes controlling other interesting traits such as *uniform ripening* (u), which leads to reduced sugar contents in ripe fruits,[7] or *rin* and *nor,* which offer long shelf life but affect the production of flavor-related compounds.[8,9] Nevertheless, probably the main negative effect was due to the strong emphasis placed in the selection for high production and biotic stress resistance and the scant attention paid to fruit flavor.

Tomato flavor is determined by the accumulation of taste-related compounds and aroma volatiles. Among the first, taste depends on the accumulation of sugars, mainly fructose and glucose (with traces of sucrose), organic acids, mainly malic, citric and glutamic acids and the ratios between them. On the other hand, volatiles would not only affect aroma perception but also affect the way that tongue receptors perceive sweetness.[6,10]

In order to redirect breeding programs towards high-quality varieties, it would be necessary to determine the individual accumulation of sugars and acids in selection programs. Especially, when these new materials are targeted to cover segmented tomato markets, that demand quality profiles with subtle taste differences.[11] For this purpose, taste-related compounds such as sugars and acids can be determined by precise instrumental techniques such as liquid chromatography, HPLC,[12] or capillary zone electrophoresis, CZE[13]. However, these techniques have a high cost and require qualified personnel for their use. Even more, only a small number of samples can be processed per day considering the duration of analysis.

As an alternative, infrared spectroscopy can provide an indirect quantification of these compounds with several advantages. Its use requires minimum preparation of the sample, it is inexpensive and offers rapid analysis time.[14] Although Near-infrared (NIR) spectroscopy has been successfully used in measuring quality attributes of horticultural produces,[15] but sometimes the goodness of the models obtained retrain their application to general screening purposes[16] and studies in different contexts have evidenced a better performance of Fourier-transform mid-infrared (FT-MIR) spectroscopy in the quantification of sugars and acids.[17]

In tomato, FT-MIR models for the quantification of taste-related compounds are scarce. They have been obtained using a wide range of varieties[18] or a specific varietal type, such as processing tomato.[19,20] Although different FT-MIR methodologies have been tested, Wilkerson et al.[19] concluded that little difference was observed comparing the use of triple bounce attenuated total reflectance (ATR) and transmission. In their study, Wilkerson et al.[19] concluded that increasing the number and diversity of the samples would reduce the impact of irrelevant spectral-variations (noise) in the calibration model, thus resulting in partial least squared regression (PLSR) models with higher correlation coefficients.

In the present study, this premise is addressed: does a higher amplitude of samples and environments really increase the efficiency of FT-MIR models in the prediction of the concentration of taste-related compounds in tomato? At the same time, the feasibility of the use of five-bounce ATR FT-MIR robust models in selection procedures of fresh and processing tomato cultivars is studied. For that purpose, the prediction models obtained were tested with

an external assay, including 111 samples, representing the variation produced by changes in the environmental growing conditions.

MATERIALS AND METHODS

Plant Material

Three sets of samples were used to develop prediction models. Set number 1 included 108 samples of processing tomato of eight varieties grown with different water and fertilization regimes in Extremadura (Spain). Set number 2 included 107 samples of medium sized tomatoes from 25 varieties including beef, rounded, plum and cluster tomatoes from commercial and traditional varieties. Set number 3 included 115 samples of 32 varieties of cherry and cocktail tomato. Samples from set 1 were obtained during the development of different agronomical studies.[21,22] Samples from sets 2 and 3 were purchased from local markets.

Each specific sample set and a general set grouping the total 330 tomato samples were used for the construction of models predicting sugar and acid contents from FT-MIR spectra. In all cases, fully ripe fruits were sampled.

A fourth sample set from a different external assay was also obtained. It contained 111 samples of processing tomato representing the same varieties and water regimes of sample set 1, but grown in Navarra (Spain) with different environmental conditions (lower radiation and temperature). Growing conditions were described with higher detail by Martí et al.[22]. These samples were not included in the calculation of the general model and were only used to test the robustness and true prediction capabilities of the model.

Sample preparation

The tomatoes were crushed, homogenized and stored at -80ºC until analysis. Subsequently, the samples were thawed and centrifuged at 15680g for 10 minutes, following the conditions reported by Wilkerson et al. for FT-MIR analysis[19]. Three aliquots of the supernatant were obtained. One of them was used to determine soluble solids content (SSC) with a Pocket PAL-α digital refractometer (Atago, Tokyo, Japan). The remaining two were used to obtain the FT-MIR spectra and sugar and acid contents via capillary zone electrophoresis (CZE).

Infrared spectroscopy analysis

The absorbance measurements of the FT-MIR spectrum were carried out using a portable Cary 630 FT-MIR spectrometer (Agilent Technologies, Waldbronn, Germany), equipped with a DTGS (Deuterated triglycine sulfate) detector and a five-bounce ZnSe ATR. Microlab FT-MIR Software v. B.05.3. (Agilent Technologies, Waldbronn, Germany) was used to acquire the data, selecting a spectral resolution of 4 cm^{-1} in a range of the average infrared spectrum of 4000-650 cm^{-1}. After reviewing the spectra, and considering previous literature[19], only the 1500-900 cm^{-1} spectra were used for chemometrics.

Two independent spectral measurements (average of 64 consecutive scans) were performed in each sample. Measurements of the reference spectrum (background) were obtained between samples to correct uncontrolled variations in the spectral measurements due to variations in environmental conditions. Between the different measurements, the glass was carefully cleaned with distilled water and dried with cellulose tissues.

CZE analysis

The quantifications of the main reducing sugars (fructose and glucose) and organic acids (citric, malic and glutamic) were performed by capillary zone electrophoresis (CZE) with an Agilent 7100 equipment (Agilent Technologies, Waldbronn, Germany) following the method described by Cebolla-Cornejo et al.[13]

Data analysis

Three specific prediction models for sample sets 1 to 3 and an additional general model with all these samples was calculated. Each model was calculated using 75% of the sample set (calibration group) to develop the calibration and cross-validation procedures. The remaining 25% of the samples (validation group) were not included in the models and were used to obtain an accurate estimate of the error committed predicting the contents of new samples. Both the calibration and validation groups were randomly selected. In the case of the general model the calibration and validation groups were again randomly selected, independently of the selection made for each specific model.

FT-MIR spectra were pre-treated to eliminate multiplicative signal interferences with a Multiplicative Scatter Correction (MSC) and response variables were autoscaled using the mean and standard deviation. The predictive models were then obtained by least squares partial regression, PLS.[23] In order to choose the optimal number of latent variables, Venetian blinds cross-validation procedure was applied. In order to check the validity of the model root mean squared errors of calibration (RMSEC) and cross-validation (RMSECV) were calculated. RMSECV values were used as one of the selection criteria for the number of latent variables to be included in the model. In this sense, new latent variables were not included if they did not lead to a reduction of RMSECV higher than 2%. The second criterion used was to select the lowest possible number of latent variables.

Outliers in the FT-MIR spectra and response variables were identified and removed. The values of the Hotelling T2 statistics and the Q residues were considered for the former and the values of the normalized residuals (<-3 or >3) and leverage parameters were considered for the response variables. After calculating the PLS regression models, the FT-MIR spectra of the samples of the validation group were used to make predictions and root mean squared errors of prediction (RMSEP) were then calculated.

Correlation coefficients (R^2) were calculated for the calibration (R^2_C), cross-validation (R^2_{CV}) and prediction data (R^2_P). RMSEP values were also contextualized using the mean (%mean) and maximum (%maximum) values. Additionally, the predictive capacity of the models was assessed using the dimensionless parameter residual prediction deviation (RPD), which represents the ratio between the standard deviation of the validation and RMSEP, and the range error ratio (RER), which is the ratio between the range in the composition values of the validation samples and the RMSEP. RPD and RER enable a better comparison between models obtained with different samples, especially the former, as RER values are highly dependent on rare high contents. Usually, RPD values should be higher than 2 in order to represent useful models for classification or quantification.[24,25] On the other hand, Williams and Norris[26] suggested that RER > 10 highly useful models while they would have limited to good application for values between 3 and 10. Additionally, the true prediction performance was assessed with external samples. In order to determine the effectiveness of the general model for screening, sensitivity (true positive rate) and specificity (true negative rate) values were calculated with the general model when a 10% or 20% selection pressure was applied.

The samples of the external model were used to determine if the obtained general model had a reliable performance not only with samples grown in the same environmental conditions (validation group), but also with those harvested in completely different conditions (external sample set).

Spectra pre-treatment, PLS regression models, detection of outliers, error parameters and goodness of fit for each model were performed in Matlab v 9.4 environment (Mathworks Inc, Natick, MA, USA) using the PLS_Toolbox v 8.2.1 module (Eigenvector Research Inc, Wenatchee, WA, USA). A heatmap of the correlations between statistical parameters of the samples and the performance of the PLS regression models was obtained with *Heatmapper* (http://www.heatmapper.ca/pairwise/).

RESULTS

The spectra obtained for the tomato samples were characterized by two high absorption areas corresponding to 3700-3000 cm^{-1} and 1750-1500 cm^{-1} (Fig. 1). In the first area, considerable differences were detected among samples, while in the second these differences were limited. Although absorbance peak in the area 1500-900 cm^{-1} was not as high, important differences in this area were detected among samples, especially in the area 1150-950 cm^{-1}.

Characterization of sample sets

SSC values obtained were in agreement with the expected values considering the varietal types used. Fresh mid-sized tomatoes had the lowest value, followed by processing tomato and cherry and cocktail tomato (Table 1). Accordingly, fructose and glucose mean contents followed the same distribution. The contents of the acids (citric, malic and glutamic) had a different distribution. Cherry and cocktail maintained higher contents, but fresh mid-sized tomatoes had higher levels of acids compared to processing tomato. For all the models, the samples selected for the validation group, that were used to predict values using the calculated PLS models, represented a similar range of variation (Table 1).

The range of variation for each variable differed among sample sets (Table 1). In general, the range of variation for sugars was lower than for acids, with the lowest variation being present for SSC. Processing tomatoes had the lowest range of variation for the accumulation of sugars, followed by fresh mid-sized tomatoes. Cherry and cocktail tomatoes represented the highest levels of variation, not only for sugars but also for glutamic acid. Regarding the rest of acids, mid-sized fresh tomatoes included higher levels of variation. The general model included a high range of composition for all the compounds.

Regarding the external assay, samples grown in Navarra had similar SSC, but lower sugar and higher acid contents than those obtained in the sample set 1, corresponding to Extremadura growing conditions (Table 1).

FT-MIR prediction models

In the three specific sample sets R^2 values for calibration were satisfactory, especially in the case of sugars and SSC, ranging from 0.82 for fructose in the model with mid-sized tomatoes to 0.98 for SCC in the model of cherry and cocktail tomato (Table 2). The R^2 values for calibration of the acids tended to be lower, especially in the mid-sized and cherry tomatoes models. For all the models, R^2 cross-validation values decreased considerably and, at the same time, RMSECV

values considerably increased over those obtained for the calibration. Indeed, the cross-validation seemed to be rather tough, as in the three models the R^2 prediction values were higher and RMSEP values were similar or lower than RMSECV.

In processing tomatoes RMSEP (%mean) values for SSC, citric fructose and glucose were lower than 10% and slightly higher (<15%) for glutamic and malic acid (Table 2). The error was equally distributed, and it was similar for samples with low, medium or high contents (fructose predictions are shown as an example, Fig. 2). Consequently, the relative error tended to decrease for increasing contents. When RMSEP was contextualized using the maximum value (%maximum), the percentages of error remained below 9% for all the compounds. In mid-sized tomatoes, RMSEP (%mean) values were lower than 18% for SSC, glucose, fructose, and citric and lower than 25% in the rest of compounds. RMSEP (%maximum) values were lower than 10% for SSC, fructose, glucose and citric and lower than 13% for malic and glutamic acids. In the case of cherry and cocktail tomatoes RMSEP (%mean) values were lower than 15% for all the compounds except for glutamic acid (27.5%) and RMSEP (%maximum) values were lower than 10% for all the compounds.

Considering the range of variation represented in the samples, for the three models RPD values were, in general, close to or higher than 2 (Table 2). On the other hand, RER values were higher than 6 for all the variables in processing and mid-sized-tomatoes, and even higher than 10 for most variables in cherry and cocktail tomatoes.

A general model was obtained with the entire set of samples. R^2 values for calibration ranged from 0.65 for malic acid to 0.96 for SSC. R^2 and RMSECV values for cross-validation were similar, to those of the calibration, thus revealing a low impact of cross-validation. R^2 values for the validation group were almost identical to those of the calibration, except for glutamic acid (0.58 vs. 0.75). RMSEP values even improved RMSEC values (mean decrease of 6%). RMSEP %mean and %maximum values for main compounds were lower than 15% and 6% respectively. RPD values were higher than 2.5 for main compounds and RER values were higher than 11 for all the compounds, except for malic acid (8.5).

It seemed that increasing the number of samples did not necessarily lead to a better performance of the model. In fact, after studying different correlations between model and prediction parameters and the description of the sample, it resulted that the highest absolute correlations between relative errors (%mean) and the characteristics of the samples were found with the coefficient of variation and minimum values of the samples (Supp. Figure 1). Specifically, the correlation between the coefficients of variation of the sample sets and relative RMSEC, RMSECV and RMSEP values were 0.75, 0.69 and 0.79 respectively. An inverse relationship was observed between the minimum value of the samples and RMSEC, RMSECV and RMSEP values (R=-0.46).

Sensitivity (true positive rate) and specificity (true negative rate) values were calculated with the general model to predict the results when a 10 or 20% selection pressure was applied. Sensitivity for a 10% selection pressure was 100% for SSC, glucose, fructose and citric acid and 75% for malic and glutamic acid (Table 4). That means that when 10% of samples with the highest predicted content were selected, all of them had the highest (10%) measured (CZE) values for the main components. Specificity was 100% for main components and higher than 97% for malic and glutamic acids, meaning that within the rejected samples only 3% had high malic or glutamic content. Even for malic and glutamic acid, the mean percentile of CZE measured values of the selected samples was close to 5%, meaning that false positives also had high contents, close to

the limit of selection. When a lower selection pressure was applied (selecting the 20% of samples), sensitivity decreased, but it was still higher than 78% for main components. Specificity was higher than 90% in all cases. Mean percentile of CE measured values was below 18% for all the variables.

Prediction of an external assay

The obtained general model was applied to the prediction of a new assay (111 samples) not included in its development. R^2 values were moderately high in the case of SSC (0.74) and glucose (0.63) and considerably lower in the rest of the cases (Table 3). Nevertheless, RMSEP values (%mean) were lower than 15% for SSC and glucose, lower than 20% for fructose and citric acid and lower than 40% for malic and glutamic acids. RMSEP values (%maximum) were lower than 21% in all cases. The specific model for sample set 1, representing the same varieties but grown in different environmental conditions was also used for predictions, but it showed a worse performance than the general model (Table 3).

When a 10% selection pressure was applied to select material using the general model, sensitivities for main components were higher than 70% and specificities higher than 96% (Table 4). The mean percentile of CZE measured values of the selected samples was lower than 8%. Thus, the false positives included within the selected samples had high values close to the limit of selection. With a lower selection pressure (20%) sensitivities decreased, with good values for SSC and glucose, but lower for fructose (60%) and citric acid (52.4%).

DISCUSSION

Tomato is a complex crop, with specific characteristics associated with different varietal types. For example, it has been proven that the breeding history of processing tomato led to a notable divergence of the genome from fresh tomato varieties.[27] This divergence is in fact accentuated in chromosome 5, where several QTL for high soluble solid content can be found. This is in part due to the strong selection pressure in processing tomato breeding programs for high SSC. Within fresh tomato, a clear distinction is observed between cherry and other types, such as round or beef tomatoes, with important differences in the metabolic profile.[28] In this sense, cherry tomatoes tend to have higher levels of sugars and citric acid and lower levels of malic acid, as it was observed in the present study.

With this level of differentiation, it seems probable that indirect quantification methods for the accumulation of taste-related compounds should be developed specifically for each varietal type. In fact, previous works using FT-MIR prediction for sugar and acid content have been either focussed on processing tomato,[19,20] or using a limited number of varieties. That is the case of Beullens et al.[29] that applied their model to four varieties or Vermeir et al.[25] with sets of six and three varieties. Nevertheless, some attempts have been made to include a higher level of diversity. It is the case of the study performed by Scibisz et al.[18], that included 39 commercial and traditional varieties of, i.a., cherry, cocktail, Marmande, and processing tomato.

Consequently, the results obtained in these studies are not easily compared. Usually, cross-validation performance is provided to assess the robustness of the models. In the present study cross-validation using Venetian blinds proved to be rather tough, as the values for R^2 and RMSE in the prediction improved those of the cross-validation. Nevertheless, the performance of the model predicting values for samples not included in the calibration gives a better idea of its robustness, thus comparisons should be made with prediction performance results. But even then, not always mean values and coefficients of variation are provided to contextualize RMSEP

values, and other values such as the dimensionless RPD and RER values are not usually provided. RPD can be interpreted as the ratio of natural variation in the samples to the size of likely prediction errors and RER is similar in spirit, but in this case, it uses the range of variation of the samples instead of the standard deviation.[24]

Despite these limitations, some comparisons can be made. In the case of processing tomato, Ayvaz et al.[20] obtained RMSEP values for glucose of 1.4 g L^{-1} and fructose 1.46 g L^{-1} and SSC of 0.12°Brix, values that contextualized with the mean result in 16%, 17% and 2% for ranges 1.1-20.2 g L^{-1}, 0.1-20.4 g L^{-1} and 3.8-7.2°Brix respectively. Also with processing tomato, Wilkerson et al.[19] obtained values of RMSEP with a triple-bounce ATR of 1.47 g L^{-1}, 1.23 g L^{-1} and 0.23 g L^{-1} for glucose, fructose and SSC. In this case, the means were not provided, but samples used for prediction ranged 10.0-21.4 g L^{-1} for glucose, 11-20.6 g L^{-1} for fructose and 4.2-6.7 °Brix for SSC. These values are similar to those obtained in the present work in the model restricted to processing tomato (RMSEP values of 1.2 g kg^{-1}, 1.4 g kg^{-1} and 0.2 °Brix with ranges of 10.01-19.69 g kg^{-1}, 9.03-20.04 g kg^{-1} and 3.5-5.8 °Brix respectively).

Regarding studies with a wider diversity, Scibisz et al.[18] obtained RMSEP values (contextualized with the mean) for glucose, fructose and SSC of 5.1%, 6.8%, and 2.9% respectively with coefficients of variation of prediction samples of 25%, 20% and 20% respectively. While in our case with the general model RMSEP values were 12.1%, 14.3% and 4.9% with much higher variability in the prediction samples (coefficients of variation of 43.2%, 39.9%, and 26.7% respectively).

Few studies are available where the authors have applied a PLS regression model to samples obtained in an external assay. Among them, Scibisz et al.[18] obtained models for individual compounds, but the external assay was applied only to predict SSC, total acidity and dry matter. In that case, they obtained an RMSEP (%mean) value for SSC of 4.5%. In the present work, RMSEP was higher 8.4%, but it should be considered that Scibisz et al.[18] included two varieties and the present work included 8 varieties grown with different water and fertilization regimes.[21,22] It could be argued that the external sample set included varieties already considered in the model and that the samples were obtained in the same years (those studies were performed with samples obtained during two years). It should be considered though, that the general model included not only these eight varieties, but much more. Additionally, previous studies with these materials proved that the site of cultivation had a similar effect or even higher than the year effect on the metabolic profile of tomato.[21,22] Even considering the limitation of the external sample set, it seems clear that the potential of using FT-MIR models with external samples would be a real.

The models developed in this work have proved to be robust and reliable for main components (SSC, glucose, fructose and citric acid), with worse models obtained for malic and glutamic acid. Vermeir et al.[25] also found lower predictivity levels for malic and glutamic acids compared to citric acid. In that case, they justified that the range of variation for these compounds and their concentration was small. Scibisz et al.[18] also found that the model for malic acid was weak, and also stated that it was probably due to the low concentration found in tomato, in agreement with the results of Rudnitskaya et al.[30] in apple, pointing out that FT-MIR models were not suitable for minor components. On the other hand, Wilkerson et al.[19] did not include malic acid in their models, but the characteristics of the glutamic model were similar to those of the citric acid. On the other hand, most published literature shows that models for SSC are much better than those for individual components[18–20], as in the present work. One plausible explanation is that SSC involves the signal of individual components and it is obviously higher. Nevertheless,

SSC indirect prediction has the only advantage of providing an estimate to contextualize new results with previous experiences were no data for individual compounds is available.

Apart from different comparisons, it seems clear that the obtained general model will be highly valuable for screening purposes. RPD values higher than 2 and RER values higher than 10, confirm this point, as these values are considered a threshold to define useful models.[24–26] In addition, RMSEP (%maximum) values for SSC, fructose, glucose and citric acid were lower than 6%. This evidences the goodness of the indirect FT-MIR determinations in the selection of materials with outstanding contents of taste-related compounds.

It also seems clear that a general model is most robust than specific models for each varietal type. In our case, this seems to be due to the higher variability included rather than to the use of a higher number of samples, as the performance of the models represented in the RMSE (%mean) values, had a high positive correlation with the coefficient of variation of the sample sets used during the calibration. It also seems that low minimum values of the variables in the samples would interfere with the calculation of good models, as moderate negative correlations were obtained with RMSE (%mean) values. Thus, answering the initial question proposed in this work, a higher amplitude of samples and environments really increases the efficiency of FT-MIR models in the prediction of the concentration of taste-related compounds in tomato.

The feasibility of the use of FT-MIR models in screening activities during breeding programs is clear. The benefits of the use of indirect quantification methods are evident when it is considered that in screening programs for a high content of specific sugars and acids, a notable contribution of genotype x environment interaction has been described, forcing to develop multi-environmental trials.[31] Even more, the considerable amount of variation present within accessions, especially in wild germplasm, leads to the need to analyze individually a high number of plants per accession in order to identify sources of variation. And again, a high number of plants are to be analyzed in segregant populations during the introgression of the trait.

The sensitivity of the PLS regression model when applying a 10% selection pressure is 100%. That means that in order to select the best materials in screening programs, time-consuming and expensive HPLC or CZE determinations can be replaced by rapid and cheap ATR FT-MIR analysis. This is of utmost utility for the development of breeding programs, which still rely on the calculation of gross measures such as SSC due to the cost of implementing selections based on individual compounds.

In indirect spectroscopic analysis, it is usually necessary to perform a calibration for each assay. But in the present work, it has been proven that even applying the general model to an external assay not included in the calibration, as it was possible to obtain sensitivities higher than 70% with 10% selection pressure for main taste-related compounds. Additionally, those false positives corresponded to samples with high contents close to the limit of selection, as the mean percentiles of selected samples were close to the ideal 5%. Thus, even if these samples were selected, their effect on the development of a breeding program would be minimum. Even for quantification purposes, RMSEP values contextualized with maximum values were lower than 13% for main compounds, which is a notable performance, considering that it has been obtained with samples not included in calibration models.

CONCLUSIONS

Our results support the applicability of ATR FT-MIR as a tool to select samples with high contents of SSC, glucose, fructose and citric acid. For that purpose, it would be advisable to use pools of samples with the highest variability to develop robust general models, rather than specific models. Preliminary results using an external sample set with materials grown in a different environment suggest that even though a calibration for each assay is advisable, it would not be necessary for screening programs if a later more precise determination is planned. Even if no further analysis were performed, most selected samples would be close to the selection objective.

ACKNOWLEDGMENTS

This research was performed despite the lack of direct public funding for its development and thanks to the enthusiasm of the authors. The authors thank Dr. Lahoz and Dr. Campillo for providing samples of processing tomato. G. Ibañez thanks Universitat Jaume I for funding his pre-doctoral grant (PREDOC/2015/45).

REFERENCES

1. Bruhn CM, Feldman N, Garlitz C, Harwood J, Ivans E, Marshall M, et al. Consumer perceptions of quality: apricots, cantaloupes, peaches, pears, strawberries,. *J Food Qual* **14**(3):187–195 (1991).https://doi.org/10.1111/j.1745-4557.1991.tb00060.x.

2. Cebolla-Cornejo J, Soler S, Nuez F. Genetic erosion of traditional varieties of vegetable crops in Europe: tomato cultivation in Valencia (Spain) as a case Study. *Int J Plant Prod* **1**(2):113–128 (2007).https://doi.org/10.22069/IJPP.2012.531.

3. Davies JN, Hobson GE. The constituents of tomato fruit — the influence of environment, nutrition, and genotype. *C R C Crit Rev Food Sci Nutr* **15**(3):205–280 (1981).https://doi.org/10.1080/10408398109527317.

4. Díaz de León-Sánchez F, Pelayo-Zaldívar C, Rivera-Cabrera F, Ponce-Valadez M, Ávila-Alejandre X, Fernández FJ, et al. Effect of refrigerated storage on aroma and alcohol dehydrogenase activity in tomato fruit. *Postharvest Biol Technol* **54**(2):93–100 (2009).https://doi.org/10.1016/j.postharvbio.2009.07.003.

5. Cebolla-Cornejo J, Roselló S, Valcárcel M, Serrano E, Beltrán J, Nuez F. Evaluation of Genotype and Environment Effects on Taste and Aroma Flavor Components of Spanish Fresh Tomato Varieties. *J Agric Food Chem* **59**(6):2440–2450 (2011).https://doi.org/10.1021/jf1045427.

6. Tieman D, Zhu G, Resende MFR, Lin T, Nguyen C, Bies D, et al. A chemical genetic roadmap to improved tomato flavor. *Science (80-)* **355**:391–394 (2017).

7. Powell ALT, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, et al. Uniform ripening Encodes a Golden. *Science (80-)* **336**(29):1711–1715 (2012).https://doi.org/10.1126/science.1222218.

8. Osorio S, Alba R, Damasceno CM, Lopez-Casado G, Lohse M, Zanor MI, et al. Systems Biology of Tomato Fruit Development: Combined Transcript, Protein, and Metabolite Analysis of Tomato Transcription Factor (nor, rin) and Ethylene Receptor (Nr) Mutants Reveals Novel Regulatory Interactions. *Plant Physiol* **157**(1):405–425 (2011).https://doi.org/10.1104/pp.111.175463.

9. Qin G, Wang Y, Cao B, Wang W, Tian S. Unraveling the regulatory network of the MADS box transcription factor RIN in fruit ripening. *Plant J* **70**(2):243–255

(2012).https://doi.org/10.1111/j.1365-313X.2011.04861.x.

10. Baldwin EA, Scott JW, Einstein M, Malundo TMM, Carr BT, Shewfelt RL, et al. Relationship between sensory and instrumental analysis for tomato flavor. *J Am Soc Hortic Sci* **123**(5):906–915 (1998).

11. Causse M, Friguet C, Coiret C, LéPicier M, Navez B, Lee M, et al. Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. *J Food Sci* **75**(9):531–541 (2010).https://doi.org/10.1111/j.1750-3841.2010.01841.x.

12. Agius C, von Tucher S, Poppenberger B, Rozhon W. Quantification of sugars and organic acids in tomato fruits. *MethodsX* **5**(May):537–550 (2018).https://doi.org/10.1016/j.mex.2018.05.014.

13. Cebolla-Cornejo J, Valcárcel M, Herrero-Martínez JM, Roselló S, Nuez F. High efficiency joint CZE determination of sugars and acids in vegetables and fruits. *Electrophoresis* **33**(15):2416–2423 (2012).https://doi.org/10.1002/elps.201100640.

14. Bureau S, Cozzolino D, Clark CJ. Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biol Technol* **148**:1–14 (2019).https://doi.org/10.1016/j.postharvbio.2018.10.003.

15. Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol Technol* **46**(2):99–118 (2007).https://doi.org/10.1016/j.postharvbio.2007.06.024.

16. García-Martínez, S., Gálvez-Sola, L. N., Alonso, A., Agulló, E., Rubio, F., Ruiz, J. J., & Moral, R. Quality assessment of tomato landraces and virus-resistant breeding lines: quick estimation by near infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, **92**(6), 1178-1185 (2012). https://doi.org/10.1002/jsfa.4661.

17. de Oliveira GA, de Castilhos F, Renard CMGC, Bureau S. Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Res Int* **60**:154–162 (2014).https://doi.org/10.1016/j.foodres.2013.10.051.

18. Scibisz I, Reich M, Bureau S, Gouble B, Causse M, Bertrand D, et al. Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. *Food Chem* **125**(4):1390–1397 (2011).https://doi.org/10.1016/j.foodchem.2010.10.012.

19. Wilkerson ED, Anthon GE, Barrett DM, Sayajon GFG, Santos AM, Rodriguez-Saona LE. Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *J Agric Food Chem* **61**(9):2088–2095 (2013).https://doi.org/10.1021/jf304968f.

20. Ayvaz H, Sierra-Cadavid A, Aykas DP, Mulqueeney B, Sullivan S, Rodriguez-Saona LE. Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. *Food Control* **66**:79–86 (2016).https://doi.org/http://dx.doi.org/10.1016/j.foodcont.2016.01.031.

21. Lahoz I, Leiva-Brondo M, Martí R, Macua JI, Campillo C, Roselló S, et al. Influence of high lycopene varieties and organic farming on the production and quality of processing tomato. *Sci Hortic (Amsterdam)* **204**:128–137 (2016).https://doi.org/10.1016/j.scienta.2016.03.042.

22. Martí R, Valcárcel M, Leiva-Brondo M, Lahoz I, Campillo C, Roselló S, et al. Influence of controlled deficit irrigation on tomato functional value. *Food Chem* **252**:250–257 (2018).https://doi.org/10.1016/j.foodchem.2018.01.098.

23. Naes T, Isaksson T, Fearn T, Davies T. A User-Friendly Guide to Multivariate Calibration and Classificationtion. NIR Publications; 2002.

24. Fearn T. Assessing calibrations: SEP, RPD, RER and R2. *NIR news* **13**(6):12–14

(2002).https://doi.org/10.1255/nirn.689.

25. Vermeir S, Beullens K, Mészáros P, Polshin E, Nicolaï BM, Lammertyn J. Sequential injection ATR-FTIR spectroscopy for taste analysis in tomato. *Sensors Actuators, B Chem* **137**(2):715–721 (2009).https://doi.org/10.1016/j.snb.2009.01.056.

26. Williams P, Norris K. Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, Inc.; 1987.

27. Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, et al. Genomic analyses provide insights into the history of tomato breeding. *Nat Genet* **46**(11):1220–1226 (2014).https://doi.org/10.1038/ng.3117.

28. Ursem R, Tikunov Y, Bovy A, Van Berloo R, Van Eeuwijk F. A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica* **161**(1–2):181–193 (2008).https://doi.org/10.1007/s10681-008-9672-y.

29. Beullens K, Kirsanov D, Irudayaraj J, Rudnitskaya A, Legin A, Nicolaï BM, et al. The electronic tongue and ATR-FTIR for rapid detection of sugars and acids in tomatoes. *Sensors Actuators, B Chem* **116**(1–2):107–115 (2006).https://doi.org/10.1016/j.snb.2005.11.084.

30. Rudnitskaya A, Kirsanov D, Legin A, Beullens K, Lammertyn J, Nicolaï BM, et al. Analysis of apples varieties – comparison of electronic tongue with different analytical techniques. *Sensors Actuators B Chem* **116**(1–2):23–28 (2006).https://doi.org/10.1016/j.snb.2005.11.069.

31. Galiana-Balaguer L, Ibáñez G, Cebolla-Cornejo J, Roselló S. Evaluation of germplasm in Solanum section Lycopersicon for tomato taste improvement. *Turkish J Agric For* **42**(5):309–321 (2018).https://doi.org/10.3906/tar-1712-61.

Table 1. Statistical parameters of the samples regarding the accumulation of soluble solids content (SSC), sugars and acids contents using refractometry and capillary zone electrophoresis. For each sample set the characteristics of the samples used for calibration (n_c) and cross-validation (n_{cv}), and those used for prediction are indicated. An external assay was included only to predict values using the general model.

		Calibration group samples					Validation group samples				
	Parameters	Mean	Standard deviation	Coefficient of variation (%)	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)	Minimum	Maximum
Sample set 1 Processing tomato Extremadura (n_c =81; n_v = 27)	SSC[1]	4.55	0.42	9.2	3.80	5.55	4.54	0.60	13.1	3.50	5.80
	Glucose[2]	14.16	2.08	14.7	9.84	19.48	14.57	2.50	17.2	10.01	19.69
	Fructose[2]	14.83	2.46	16.6	9.72	19.94	15.14	2.97	19.6	9.03	20.04
	Citric[2]	3.53	0.57	16.0	2.25	5.34	3.62	0.60	16.7	2.53	5.22
	Malic[2]	1.23	0.30	24.1	0.79	1.83	1.25	0.30	23.7	0.87	2.00
	Glutamic[2]	0.96	0.43	44.7	0.36	2.35	1.03	0.41	39.7	0.42	1.96
Sample set 2 Fresh mid-sized tomato (n_c =80; n_v = 27)	SSC[1]	4.27	0.55	13.0	2.95	5.40	4.25	0.52	12.3	3.50	5.65
	Glucose[2]	12.10	3.40	28.0	6.73	22.49	11.50	3.31	28.8	6.98	22.72
	Fructose[2]	13.77	3.67	26.7	8.25	25.63	12.99	3.47	26.7	8.15	24.18
	Citric[2]	5.84	1.93	33.0	3.06	14.03	5.43	1.74	32.1	2.70	9.75
	Malic[2]	1.76	0.63	35.6	0.56	4.02	1.73	0.69	39.7	0.56	3.75
	Glutamic[2]	1.70	0.76	44.8	0.71	4.25	1.68	0.68	40.3	0.70	3.25
Sample set 3 Cherry&cocktail tomato (n_c =86; n_v = 29)	SSC[1]	5.77	1.44	24.9	3.50	11.35	6.00	1.84	30.7	4.00	12.40
	Glucose[2]	18.14	7.73	42.6	6.73	46.09	19.33	9.13	47.2	10.03	50.42
	Fructose[2]	20.81	7.77	37.3	8.44	51.99	21.57	9.15	42.4	12.41	53.43
	Citric[2]	8.42	1.65	19.6	4.67	12.38	8.36	1.60	19.1	5.20	11.66
	Malic[2]	1.47	0.47	31.6	0.79	3.44	1.50	0.49	32.9	0.96	3.01
	Glutamic[2]	1.60	1.18	73.4	0.32	6.77	1.71	1.47	86.1	0.28	6.43
General model (n_c =245; n_v = 87)	SSC[1]	4.91	1.18	23.9	2.95	11.35	4.90	1.31	26.7	3.50	12.40
	Glucose[2]	14.97	5.70	38.1	6.73	46.09	15.17	6.55	43.2	6.73	43.18
	Fructose[2]	16.66	6.17	37.0	8.15	51.99	16.71	6.67	39.9	8.25	53.43
	Citric[2]	5.90	2.43	41.2	2.36	12.38	6.13	2.58	42.1	2.25	14.03
	Malic[2]	1.51	0.54	35.9	0.56	4.02	1.45	0.47	32.4	0.56	3.01
	Glutamic[2]	1.47	0.93	63.1	0.28	6.77	1.43	0.92	64.1	0.32	5.12
External assay Processing tomato Navarra (n_v = 111)	SSC[1]						4.66	0.58	12.5	3.45	6.10
	Glucose[2]						12.38	2.82	22.8	6.10	20.87
	Fructose[2]						13.13	3.14	24.6	5.84	22.42
	Citric[2]						4.21	0.85	20.2	2.13	7.06
	Malic[2]						0.84	0.24	28.1	0.32	1.49
	Glutamic[2]						1.70	0.42	24.6	0.81	2.82

[1]°Brix; [2] g kg^{-1}

Table 2. Performance of the partial least squares (PLS) regression models predicting taste-related compounds content from ATR FT-MIR spectra. SSC: soluble solids content; R^2 correlation coefficient; RMSE: root mean squared error; C: calibration; CV: cross-validation; P: prediction; RPD: residual prediction deviation; RER: range error ratio.

Model	Parameters	R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP	RMSEP (%Maximum)	RMSEP (%Mean)	RPD	RER
Sample set 1 Processing tomato Extremadura (n_c[3] =81; n_v[4] = 27)	SSC[1]	0.85	0.2	0.50	0.3	0.81	0.2	4.2	5.3	2.5	9.5
	Glucose[2]	0.88	0.7	0.63	1.3	0.79	1.2	5.9	8.0	2.2	8.3
	Fructose[2]	0.88	0.9	0.45	1.9	0.77	1.4	6.8	9.0	2.2	8.1
	Citric[2]	0.90	0.2	0.51	0.4	0.68	0.3	5.6	8.0	2.1	9.3
	Malic[2]	0.82	0.1	0.52	0.2	0.58	0.2	8.5	13.6	1.8	6.7
	Glutamic[2]	0.94	0.1	0.70	0.2	0.83	0.1	7.1	13.6	2.9	11.0
Sample set 2 Fresh mid-sized tomato (n_c =80; n_v = 27)	SSC[1]	0.93	0.1	0.78	0.3	0.91	0.2	3.3	4.4	2.8	11.6
	Glucose[2]	0.83	1.4	0.39	2.7	0.57	2.0	8.8	17.4	1.7	7.9
	Fructose[2]	0.82	1.5	0.30	3.1	0.70	1.6	6.7	12.4	2.2	10.0
	Citric[2]	0.58	1.1	0.39	1.3	0.75	0.9	9.6	17.1	1.9	7.6
	Malic[2]	0.63	0.4	0.09	0.6	0.49	0.4	10.7	23.1	1.7	8.0
	Glutamic[2]	0.54	0.5	0.38	0.6	0.65	0.4	12.6	24.4	1.7	6.2
Sample set 3 Cherry & cocktail tomato (n_c =86; n_v = 29)	SSC[1]	0.98	0.2	0.97	0.3	0.98	0.3	2.6	5.3	5.8	26.5
	Glucose[2]	0.99	0.9	0.94	1.9	0.99	1.1	2.1	5.4	8.7	38.5
	Fructose[2]	0.95	1.8	0.92	2.2	0.93	2.4	4.5	11.2	3.8	17.0
	Citric[2]	0.76	0.8	0.60	1.0	0.57	1.1	9.0	12.6	1.5	6.1
	Malic[2]	0.92	0.1	0.62	0.3	0.90	0.2	5.3	10.7	3.1	12.8
	Glutamic[2]	0.88	0.4	0.82	0.5	0.89	0.5	7.3	27.5	3.1	13.1
General model (n_c =245; n_v = 87)	SSC[1]	0.96	0.2	0.94	0.3	0.95	0.2	1.9	4.9	5.5	37.1
	Glucose[2]	0.92	1.6	0.89	1.8	0.89	1.8	4.3	12.1	3.6	19.8
	Fructose[2]	0.89	2.1	0.84	2.5	0.87	2.4	4.5	14.3	2.8	18.9
	Citric[2]	0.95	0.6	0.87	0.9	0.90	0.8	5.6	12.7	3.3	15.1
	Malic[2]	0.65	0.3	0.54	0.3	0.54	0.3	9.6	20.0	1.6	8.5
	Glutamic[2]	0.85	0.4	0.75	0.5	0.58	0.4	8.0	28.7	2.2	11.7

[1]RMSE values expressed as °Brix; [2]RMSE values expressed as g kg^{-1}; [3]number of samples in calibration set; [4]number of samples in validation set

Table 3. Performance of the predictions made for the external assay (processing tomato grown in Navarra; 111 samples) obtained with the general model and the specific model of sample set 1 (processing tomato grown in Extremadura). R^2_P correlation coefficient of the prediction; RMSEP: root mean squared error of the prediction (°Brix for SSC and g kg^{-1} for individual compounds).

Model	Parameter	R^2_P	RMSEP	RMSEP (%Mean)	RMSEP (%Maximum)
General model predicting external assay (processing tomato Navarra)	SSC	0.74	0.39	8.4	6.4
	Glucose	0.63	1.74	14.1	8.3
	Fructose	0.34	2.51	19.1	11.2
	Citric	0.25	0.86	20.4	12.2
	Malic	0.18	0.31	36.9	20.8
	Glutamic	0.19	0.49	28.6	17.2
Sample set 1 (processing tomato Extremadura) model predicting external assay (processing tomato Navarra)	SSC	0.56	0.68	14.6	11.12
	Glucose	0.49	2.40	19.4	11.5
	Fructose	0.17	3.66	27.9	16.3
	Citric	0.05	1.07	25.4	15.2
	Malic	0.21	0.24	28.8	16.23
	Glutamic	0.11	0.45	26.2	15.8

Table 4. Values of sensitivity and specificity obtained using the general PLS model in plant selection for high content of taste-related constituents when applying a selection pressure of 10% or 20%. Samples from the general model validation set group and the external assay (processing tomato grown in Navarra) were evaluated. Mean percentiles of selected plants are also provided.

	10% Selection pressure			20% Selection pressure		
General model	Sensitivity (%)	Specificity (%)	Mean percentile (%)	Sensitivity (%)	Specificity (%)	Mean percentile (%)
SSC	100.0	100.0	5.8	86.7	96.8	11.3
Glucose	100.0	100.0	5.6	87.5	96.9	11.7
Fructose	100.0	100.0	5.4	82.4	92.4	16.0
Citric	100.0	100.0	5.8	78.6	94.6	12.3
Malic	75.0	97.1	5.3	62.5	90.3	13.3
Glutamic	75.0	97.2	6.9	68.8	92.2	17.7

	10% Selection pressure			20% Selection pressure		
External assay	Sensitivity (%)	Specificity (%)	Mean percentile (%)	Sensitivity (%)	Specificity (%)	Mean percentile (%)
SSC	88.9	98.8	5.4	73.7	93.4	13.2
Glucose	90.0	98.9	5.8	70.0	92.5	14.0
Fructose	70.0	96.7	7.0	60.0	90.1	21.9
Citric	70.0	96.8	8.0	52.4	88.0	22.0
Malic	40.0	93.6	8.8	47.6	86.6	19.3
Glutamic	60.0	95.7	10.0	42.9	85.4	37.1

FIGURE CAPTIONS

Figure 1. FT-MIR spectra of tomato samples in the 4000-650 cm^{-1} region (a) and in the 1500-900 cm^{-1} region (b) used for chemometric analysis.

Figure 2. Predicted (ATR FT-MIR) vs. measured (CZE) glucose contents using the partial least squares (PLS) regression model for sample set 1 (processing tomato in Extremadura). Grey dots: samples used to calibrate the PLS model; red diamonds: samples used to predict using the PLS model.