Agric Res (September 2013) 2(3):229–235
DOI 10.1007/s40003-013-0068-4

FULL-LENGTH RESEARCH ARTICLE

Authentication of Mango Varieties Using Near-Infrared Spectroscopy

Shyam N. Jha · Pranita Jaiswal · Kairam Narsaiah · Ramesh Kumar · Rajiv Sharma · Mansha Gupta · Rishi Bhardwaj · Ashish K. Singh

Received: 21 August 2012 / Accepted: 31 May 2013 / Published online: 23 June 2013

Abstract Mango cultivars are presently identified phenotypically. In the present study, near-infrared spectroscopy in the wavelength range of 1,200–2,200 nm, in combination with chemometrics was evaluated for the purpose. Principal component analysis and partial least square (PLS) discriminate analysis were applied using characteristic variables as 0 and 1 for cv Alphonso and Banganapalli, respectively, in one group and for Dasheri and Malda, respectively, in the other group. PLS was carried out with and without pre-processing of spectral data. Wavelength range of 1,600–1,674 nm was found to be suitable for the classification of Alphonso and Banganapalli after detrending of data, while Dasheri and Malda could be classified in the wavelength range 1,200–2,200 nm after computing Norris-Gap derivative. Alphonso–Banganapalli could be differentiated from each other with 99 % accuracy, whereas, accuracy for Dasheri and Malda was about 94 at 5 % level of significance.

Keywords Classification · Characteristic variables · Cultivars · Discriminate analysis · Pre-processing

Introduction

Mango (*Mangifera indica* L.) is one of the most popular, delicious, nutritionally rich fruit with health-promoting qualities. Mango is mainly a tropical fruit with a few representatives in temperate region, accounting for approximately half of all tropical fruits produced worldwide. India is the main producer contributing over 50 % of world production [23]. Many authors consider India as the centre of origin [14] of mango due to its high degree of diversity with over 1,000 recognized varieties [24], most of them selected from naturally occurring open-pollinated seedlings [7]. These varieties are available in a range of colours, sizes and shapes. *Alphonso*, *Benishaan* or *Benisha* (*Banganapalli* in Telugu and other south Indian languages) mango varieties are the most popular varieties in southern states of India, while Kesar, *Chausa*, *Dasheri*, *Langra* and *Malda* are popular in the northern states.

The physical attributes of fruits such as weight, peel content, length, seed, pulp, gloss are highly variable with variety [11]. Mango fruits contain protective nutrients and are rich in minerals such as magnesium, potassium, sulphur, phosphorus and vitamin A, C and D. Good mango varieties contain 20 % of total soluble sugars [22]. The acid content of ripe fruit varies from 0.2 to 0.5 % and protein content is about 1 %. The nutrient percentage depends upon the variety and maturity of the fruit. When mango is green, vitamin C is higher. As it ripens, the amount of beta carotene (vitamin A) increases. Based on the physico-chemical quality parameters, different mango varieties are used for different kinds of processing. Besides, these

S. N. Jha (✉) · P. Jaiswal · K. Narsaiah · R. Sharma · M. Gupta · R. Bhardwaj · A. K. Singh
Agricultural Structures and Environmental Control Division, Central Institute of Postharvest Engineering &s Technology (CIPHET), Ludhiana 141004, India
e-mail: snjha_ciphet@yahoo.co.in

R. Kumar
Horticulture Crop Processing Division, Central Institute of Postharvest Engineering &s Technology (CIPHET), Ludhiana 141004, India

parameters also govern the sensory appeal of the varieties and hence consumer preference. Shafqat et al. [26] reported that *Alphonso* and *Fajri* had the largest fruit and *Banganapalli* had the maximum pulp (79.4 %). Kumar et al. [20] evaluated 20 mango varieties and among them, *Dasheri* excelled in pulp percentage. Sugar content was reported maximum in *Langra* (20 %) and *Malda* (20 %) [17]. Maximum peel weight was observed in variety *Fajri* (84.99 g), followed by *Alphonso* (71.33 g). Among the varieties exported from India, *Alphonso*, *Dashehari*, *Banganapalli* and *Kesar* have high demand at the international market [23]. Every year 13,000 MT of Alphonso variety is exported to Middle East, UK and the Netherlands [22]. *Alphonso* and *Banganapalli* are harvested almost concurrently in Southern India and same is the case with *Dasheri* and *Malda* in Northern India. Since *Alphonso* and *Banganapalli* occur concurrently in India and are similar in size and shape, the probability of mixing the two varieties is huge, if someone is not aware of their skin and surface colour, thereby escalating the need for a system to differentiate between the two varieties non-destructively before export.

Mangoes are sold by naming its varieties and are identified on the basis of morphological traits which usually become difficult for a common man as they are not well conversed with the varietal traits. Sub-varieties which neither have the sweetness, nor the flavour of original one but are morphologically similar and can fetch a good price at the cost of lack of experience of customer. Traders/sellers therefore mislead consumers. Variety identification based exclusively on phenotypic traits is inaccurate due to the influence of the environment and the limiting number of discriminating traits. Thus, molecular identification of mango varieties has been tried with different molecular systems such as isozymes [2], minisatellites [1], ISSRs [4], AFLPs [3, 19] and RAPDs [5, 6, 18, 21, 24, 25]. All these techniques, however, are of destructive nature in which sample cannot be sold after testing. The objective of this study was therefore to explore the potential of near infrared (NIRS) to classify two pairs of major mango variety, each one available concurrently in the markets for developing instrumental method to classify the mango non-destructively.

Materials and Methods

Sample Collection

Two pairs of mango varieties viz. *Alphonso* and *Banganapalli; Dasheri* and *Malda*, selected for this study were harvested manually for two consecutive years (2009 and 2010) following randomised block design [14, 15] from the commercial orchards of Indian states (*Maharashtra, Karnataka, Odisha, Punjab and Bihar*). The fruits were picked from five different positions viz. north, south, east, west and centre of tree canopy at three different dates called first, second and third harvest in the morning hours with intact stalk of 8–10 mm. The first harvesting was at relatively immature stage, second at about 100 % maturity and third at over mature stage based on growers experience and previous findings [12, 13, 15]. The fruits were transported to the laboratory within 48 h of harvesting in ventilated corrugated fibre board boxes containing partially frozen gel packs at the bottom and top of mangoes' layer to minimize the quality loss. Thereafter mangoes of each variety free from external injury/blemish and of visually similar in size and colour were immediately sorted, destalked and desapped. Five fresh fruits of each variety were used for experiment (0 day) and the remaining samples were stored under ambient conditions (26–30 °C, RH 60–70 %) for natural ripening. Further three fruits of each variety were used from stored samples at an interval of 2 days (2, 4, 6, 8 and 10 days) to have samples of varying physicochemical properties.

NIR Spectra Acquisition

The acquisition of spectra was performed using a portable NIR spectrophotometer (Luminar 5030, Brimrose Corp., Maryland, USA), equipped with a diffuse reflectance optical configuration and an InGaAs (1,200–2,200 nm) detector. The spectrometer included a measurement unit connected to computer similar to the one used by Jha et al. [16]. Each fruit was scanned 50 times in the sampled region and the readings were automatically averaged to one spectrum signal. A total of four spectra from two equidistant points along the equatorial region of both sides of fruit were acquired, averaged and stored for analysis.

Chemometric Analysis

Spectral data were imported to unscrambler (CAMO AS, Trondheim, Norway, version 10.1) software and data were plotted (Fig. 1) to inspect the nature of the spectra. Visually odd spectral curves among the groups were identified as outliers and were deleted. For preliminary classification, principal component analysis (PCA) was performed on the selected spectra and score plots between different principal components (PCs) were plotted to see the formation of different groups of samples based on their varieties. After ascertaining the group,

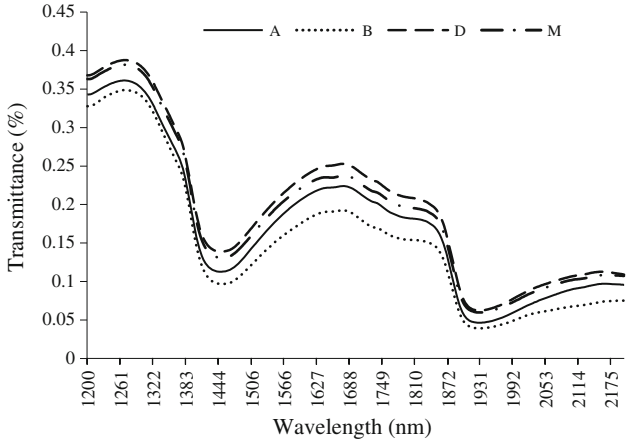

Fig. 1 Typical NIR spectra of mango in wavelength range of 1,200–2,200 nm

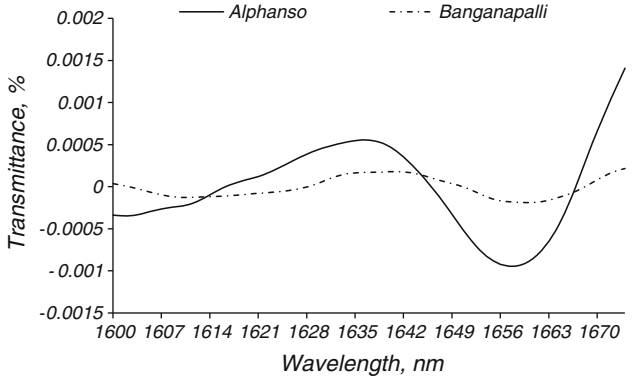

Fig. 2 Detrended NIR spectra of *Alphonso* and *Banganapalli* cultivars of mango in the wavelength range of 1,600–1,674

final classification was carried out using partial least square (PLS) discriminate analysis. PLS regression and PCA projection methods were employed thereafter for classification of samples. Altogether 1,310 samples were used for classification of mango samples.

Spectral Region Selection

Selection of spectral region plays vital role in building calibration model using latent variables methods such as PCA and PLS. The absorbing regions (wave number where maximum light is absorbed) of spectra decide the number of input variables in multivariate statistical model. The spectral range should include information on varietal variation of the samples monitored and other matrix constituents while excluding regions dominated by noise or other artifacts that might be incorporated into the model. Suitable regions may be identified by plotting the correlation spectra or by simply plotting the raw spectra and looking in differences in various peaks.

Classification of Samples

Preliminary classification of samples was performed using PCA. Score plots between different PCs were drawn to study the distribution depicting similarities among samples. Thereafter PLS discrimination analysis was performed. This method is based on modelling the differences between several classes with PLS. In this method both the varieties under analysis were assigned indicator variable values of 0 and 1, i.e., a binary variable with value 1 for members of that class, 0 for non-members. In current experiment cv *Alphonso* was assigned 0 and *Banganapalli* 1 value. Similarly in case of *Dasheri* and *Malda*, *Dasheri* was assigned 0 and *Malda* 1 value. Thereafter by building a PLS model with all indicator variables as Y, class membership was directly predicted from the *X*-variables describing the samples. Discrimination was carried out on the raw data and on the selected peak regions of the spectra with and without applying data pre-treatment. A predicted value of 0.5 was selected as the cutoff for assigning samples to either variety category.

To further verify the above classification, PLS prediction method of classification was used. For this purpose, the samples were grouped into different varieties manually. Separate PLS for each category was performed and optimum number of PCs (same for both categories of varieties) was selected based on *x*-explained variance. These PLS models were used to develop classification table using the classification module available in the Unscrambler software to assign each sample to respective variety and the percentage of correctly classified samples was computed.

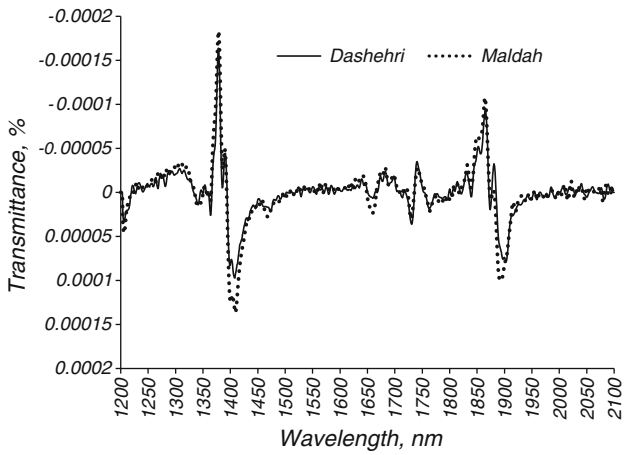

Fig. 3 Second-order derivatives (Norris-Gap) of *Dasheri* and *Malda* cultivar in the wavelength range of 1,200–2,100 nm

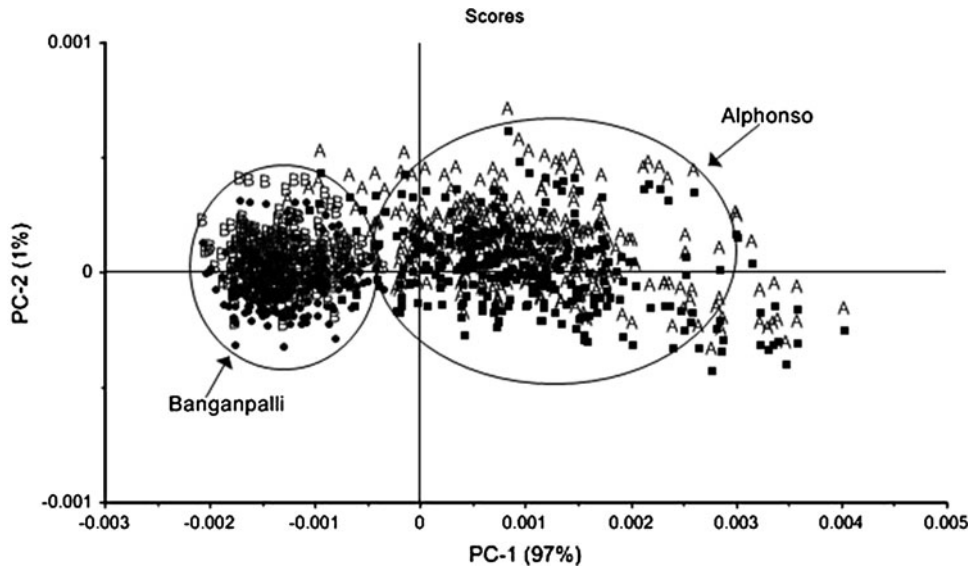

Fig. 4 Principal component scores plot of 564 samples of *Alphonso and Banganapalli* in the wavelength range of 1,600–1,674 nm, *ellipses* show grouping of samples according to variety (colour figure online)

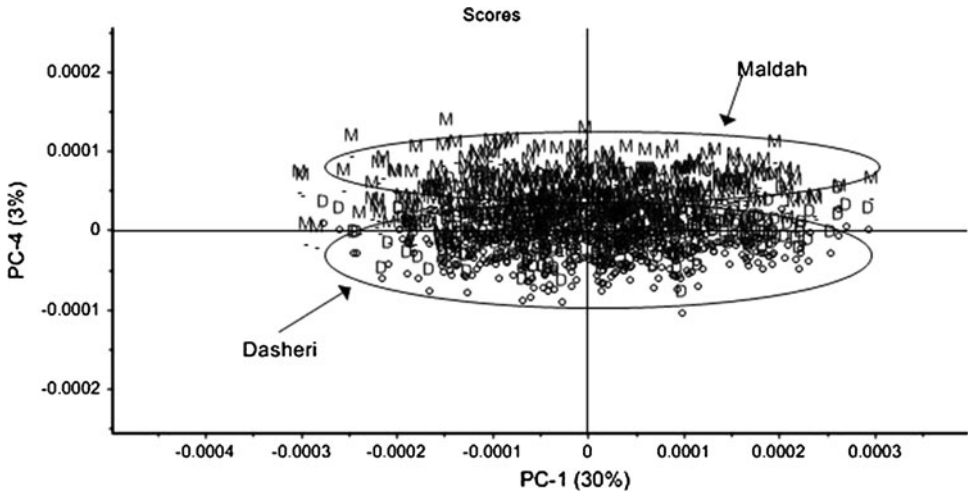

Fig. 5 Principal component scores plot of 746 samples *Dasheri and Malda* in the wavelength range of 1,200–2,100 nm, *ellipses* show grouping of samples according to variety (colour figure online)

Results and Discussion

Preliminary Analysis

A typical spectral curve (Fig. 1) of all four variety samples (*Alphonso*, *Banganapalli*, *Dasheri and Malda*) did not show any major differences in peaks and depressions. The peaks and depressions in spectra show the strong and weak transmittance characteristics of the mangoes, respectively, within the range of study. The relative values in other region of spectra, however, varied from sample to sample. Such variations have been reported due to change in surface texture and moisture content of the fruits [8, 9].

To improve the efficacy of raw spectral data for classifying mango samples, pre-processing of spectra such as baseline correction, detrending, smoothing (second order), full multiplicative scatter correction (MSC) and Norris-Gap derivatisation was performed [8, 16]. The spectral curve obtained following pre-processing was critically studied in different wavelength, for any visible varietal difference. The number of wavelengths was also sequentially minimized to obtain small groups and best performing wavelengths for distinguishing *Alphonso* from *Banganapalli* and *Dasheri* from *Malda*. Similar methodology was employed by Jha and Gunasekaran [10]. The spectral curve for *Alphonso* and *Banganapalli* differed in wavelength range of 1,600–1,674 nm following application of detrending (Fig. 2). The Alphonso variety, unlike *Banganapalli*, showed sharp depression in wavelength range of 1,644–1,664 nm and a peak in 1,616–1,644 nm.

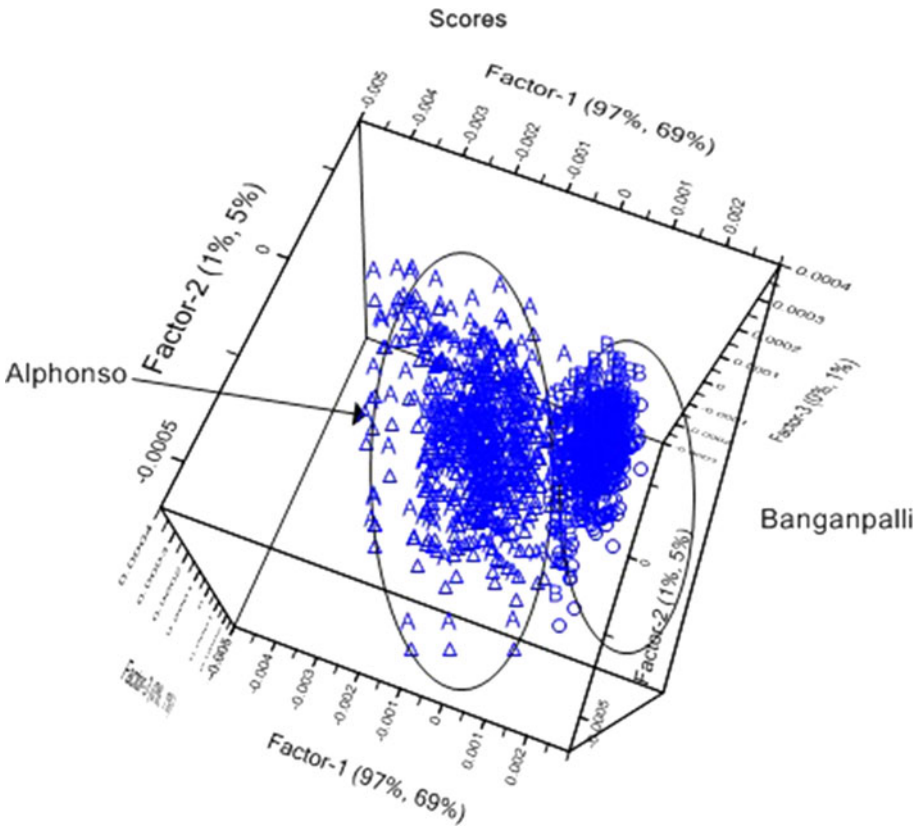

Fig. 6 Partial least square scores plot of 564 samples of *Alphonso and Banganapalli* in the wavelength range of 1,600–1,674 nm, *ellipses* show grouping of samples according to variety (colour figure online)

Dasheri and *Malda* also showed distinguishable peaks in 1,200–2,200 nm (Fig. 3) after pre-processing (Norris-Gap derivative).

These wavelength ranges were further used to build respective score plots following PCA. The score plots for *Alphonso* and *Banganapalli* between PCs 1 and 2, which respectively described 97 and 1 % of the variance in the spectral collection, roughly showed two groups based on the type of variety (Fig. 4). Similarly for second pair (*Dasheri–Malda*), the score plots between components 1 and 4 showed two groups based on the type of variety with the variance of 30 and 3 % (Fig. 5), indicating that the classification of mango varieties is possible, with a significance level of 5 %. Further the score plots developed after PLS regression on same wavelength range showed clear separation of *Alphonso* from *Banganapalli* (Fig. 6) and *Dasheri* from *Malda* (Fig. 7). In both the pairs, sum of explained variances was found to be highest by the factor PCs 1 and 2. In first pair, the sum of explained variances by factors 1 and 2 was very high, 98 and 74 % for *X* and *Y* components, respectively. This indicates that classification of *Alphonso* and *Banganapalli* is possible using NIR spectroscopy with a reasonable degree of certainty. For the second pair, the sum of explained variances by factors 1 and 2 was comparatively lower, 34 and 60 % for *X* and *Y* components, respectively.

Classification of Samples

PLS discriminate analysis was performed using raw, Norris-Gap derivative, detrended spectra for variety based classification of mango samples. It is evident from the summary of analyses that in case of *Alphonso–Banganapalli* (Table 1), 99.07 % samples of *Alphonso* and 99.58 % samples of *Banganapalli* could be classified correctly, with predicted category variables <0.5 for *Alphonso* and higher than the same for *Banganapalli* varieties. Similarly, in case of *Dasheri–Malda* variety, almost all the samples (98.37 %) of *Dasheri* have predicted category variables <0.5, whereas, for *Malda,* predicted category variables were higher than 0.5 for 945 of samples (Table 1). PLS model developed for *Alphonso–Banganapalli* after detrending of data in the range of 1,600–1,674 nm therefore correctly discriminated the two varieties. Similarly PLS model of treated (Norris-Gap derivative) spectral data in the range of 1,200–2,200 nm could correctly discriminate *Dasheri–Malda*. Jha and Gunasekaran [10] also demonstrated the use of PLS discrimination and PCA as a tool to classify samples of mango juice with and without added sugar content.

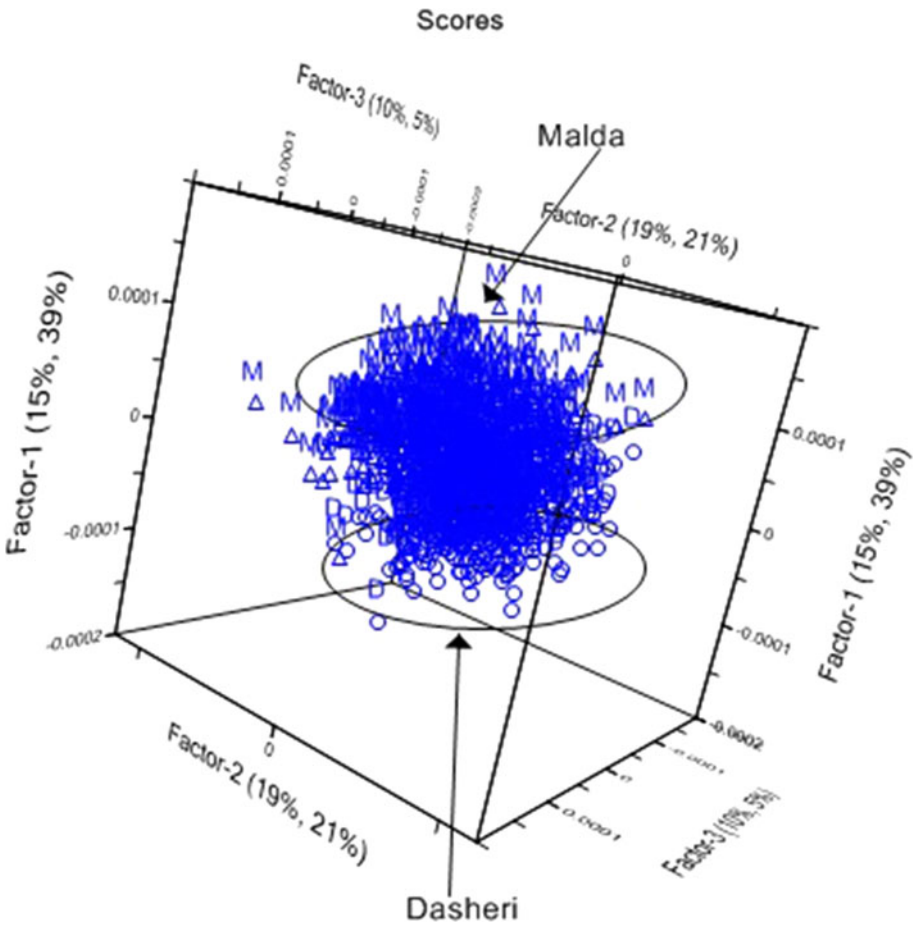

Fig. 7 Partial least square scores plot of 746 samples of *Dasheri and Malda* in wavelength range of 1,200–2,100 nm, *ellipses* show grouping of samples according to variety (colour figure online)

Table 1 PLS discriminate analysis for cultivar based classification of mango samples

Data treatment	Varieties	No. of samples	Wavelength range (nm)	Number of samples with computed category variables				Correctly classified (%) in validation set
				≥0.5		≤0.5		
				Calibration	Validation	Calibration	Validation	
De-trending	*Alphonso*	325	1,600–1,674	10	3	315	322	99.07
	Banganapalli	239		238	238	1	1	99.58
Norris Gap Derivative	*Dasheri*	431	1,200–2,100	3	7	428	424	98.37
	Malda	317		315	298	2	19	94.00

Conclusions

Near infrared spectroscopy in combination with chemometric analyses showed potential in identifying specific mango variety under study. The *Alphonso* and *Banganapalli* cv at 5 % significance level were predicted with accuracy of 99.07 and 99.58 %, respectively, whereas the *Dasheri* and *Malda* variety can be identified at an accuracy level of 98.37 and 94 %, respectively.

Acknowledgments This research was supported by the National Agricultural Innovation Project (NAIP), Indian Council of Agricultural Research (ICAR) through its subproject entitled "Development of non-destructive systems for evaluation of microbial and physico-chemical quality parameters of mango" Code number "1030".

References

1. Adato A, Sharon D, Lavi U (1995) Application of DNA fingerprinting for identification and genetics analysis of mango

(*Mangifera indica* L.) genotypes. J Am Soc Hortic Sci 120(2): 259–264
2. Degani C, El-Batsri R, Gazit S (1990) Enzyme polymorphism in mango. J Am Soc Hortic Sci 115(5):844–847
3. Eiadthong W, Yonemori K, Sugiura A, Utsunomiya N, Subhadrabandhu S (1999) Analysis of phylogenetic relationship in Mangifera by restriction site analysis of an amplified region of cpDNA. Sci Hortic 80:145–155
4. Eiadthong W, Yonemori K, Kanzaki S, Sugiura A (2000) Amplified fragment length polymorphism analysis for studying genetic relationship among *Mangifera* species in Thailand. J Am Soc Hortic Sci 125:160–164
5. Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis in bread wheat. Euphytica 113:163–185
6. Hemanth KNV, Narayanswamy P, Theertha PD, Mukunda GK, Sondur SN (2001) Estimation of genetic diversity of commercial mango varieties using RAPD markers. J Hortic Sci Biotechnol 76:529–533
7. Iyer CPA, Degani C (1997) Classical breeding and genetics. In: Litz RE (ed) The Mango, botany, production and uses. CAB International, Wallingford
8. Jaiswal P, Jha SN, Bhardwaj R (2012) Non-destructive prediction of quality of intact banana using spectroscopy. Sci Hortic 135: 14–22
9. Jha SN, Garg R (2010) Non destructive prediction of quality of intact apple using near infrared spectroscopy. J Food Sci Technol 47:207–213
10. Jha SN, Gunasekaran S (2010) Authentication of sweetness of mango juice using Fourier transform infrared attenuated total reflection spectroscopy. J Food Eng 101(3):337–342
11. Jha SN, Matsuoka T (2002) Surface gloss and weight of eggplant during storage. Biosyst Eng 81(4):407–412
12. Jha SN, Chopra S, Kingsly ARP (2006) Non-destructive determination of firmness and yellowness of mango during growth and storage using visual spectroscopy. Biosyst Eng 94:397–402
13. Jha SN, Chopra S, Kingsly ARP (2007) Modeling of colour values for nondestructive evaluation of maturity of mango. J Food Eng 78:22–26
14. Jha SN, Jaiswal P, Narsaiah K, Bhardwaj R, Sharma R, Basedia A (2010) Postharvest micro-flora of major varieties of Indian mangoes. Sci Hortic 125(4):617–621
15. Jha SN, Jaiswal P, Narsaiah K, Singh AK, Kaur PP, Sharma R, Kumar R, Bhardwaj R (2011) Prediction of sensory profile of mango using textural attributes during ripening. Food Bioprocess Technol. doi:10.1007/s11947-011-0720-6
16. Jha SN, Jaiswal P, Narsaiah K, Gupta M, Bhardwaj R, Singh AK (2012) Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy. Sci Hortic 138:171–175
17. Jilani MS, Bibi F, Waseem K, Khan MA (2010) Evaluation of physico-chemical characteristics of mango (*Mangifera indica* L.) varieties grown in D. I. Khan. J Agric Res 48(2):201–207
18. Karihaloo JL, Dwivedi YK, Archak S, Gaikwad AB (2003) Analysis of genetic diversity of Indian mango varieties using RAPD markers. J Hortic Sci Biotech 78:285–289
19. Kashkush K, Jinggui F, Tomer E, Hillel J, Lavi U (2001) Variety identification and genetic map of mango (*Mangifera indica*). Euphytica 122:129–136
20. Kumar R, Sharma RM, Kher R (2002) Performance of some mango varieties under sub-tropical rainfed region of Jammu. Haryana J Hortic Sci 31(1–2):8–10
21. Lopez-Valenzuela JA, Martinez O, Paredes-Lopez O (1997) Geographic differentiation and embryo type identification in *Mangifera indica* L. varieties using RAPD markers. Hortic Sci 32:1105–1108
22. National Horticulture Board (NHB) (2003) Database National Horticulture Board. Ministry of Agriculture, Government of India, India. www.nhb.gov.in. Accessed 21 June 2012
23. National Horticulture Board (NHB) (2009) Database National Horticulture Board. Ministry of Agriculture, Government of India, India. www.nhb.gov.in. Accessed 21 March 2012
24. Ravishankar KV, Lalitha A, Dinesh MR (2000) Assessment of genetic relatedness among mango varieties of India using RAPD markers. J Hortic Sci Biotechnol 75(2):198–201
25. Schnell RJ, Ronning CM, Knight RJ (1995) Identification of varieties and validation of genetic relationships in *Mangifera indica* L. using RAPD markers. Theor Appl Genet 90:269–274
26. Shafqat A, Haq CA, Hussain S (1992) Physico-chemical studies of varieties of mango grown at Shujabad. Pak J Agric Sci 13(4): 350–355