Supporting information

Real-time Food Authentication Using a Miniature Mass Spectrometer

Stefanie Gerbig[#], Stephan Neese[#], Alexander Penner, Bernhard Spengler, Sabine Schulz*

[#]equally contributing

Institute of Inorganic and Analytical Chemistry, Justus Liebig University Giessen, 35392 Giessen, Germany

*Corresponding Author: Sabine Schulz
Sabine.schulz@anorg.chemie.uni-giessen.de
Tel: 0641-9934807
Fax: 0641-9934809

Abstract

Detailed views of the graphical user interface of the *MS Classifier* software are depicted in Figures S-1, S-2, S-3 and S-4. Figure S-5 shows two PCA score plots of data points derived from pure milk species and their 50:50 (v:v) mixture. Figure S-6 shows three different views of the same PCA score plot of pure cow and soy milk and mixtures of different ratios of those milk types. Figure S-7 depicts the processing time and memory usage for different bin sizes and compares the QR and NIPALS algorithm of PCA analysis. Figure S-8 demonstrates the processing time and memory usage for the classification of hundreds of samples and compares the three available classifiers (Euclidean distance, Mahalanobis distance and Linear Discriminant analysis). Figure S-9 shows the PCA score plot of milk samples. In this analysis, several brands of the same milk type and more than 1000 spectra were included. Figure S-10 shows the PCA score plot of around 6600 milk spectra that were recorded in the time course of over one year. Figure S-11 shows the graphical user interface during a live classification experiment and also depicts the classification window. Figure S-12 depicts averaged mass spectra of fish samples obtained with desorption electrospray ionization and the miniature mass spectrometer. Figure S-13 depicts additional orientations of the PCA score plot of fish samples. The same data as shown in Fig. 2 in the main manuscript was used. In Figure S-14, low temperature plasma-MS spectra of the analysed coffee beans obtained with the miniature mass spectrometer are depicted.

Table S-1 provides information about cross-validation results of the milk data set depending on different bin sizes, PCA algorithms, scaling of the data and the number of principal components. Table S-2 lists cross-validation results for the analysis of coffee beans using different settings for bin size, data scaling and algorithm. Table S-3 provides information about cross-validation results for the fish sample set.

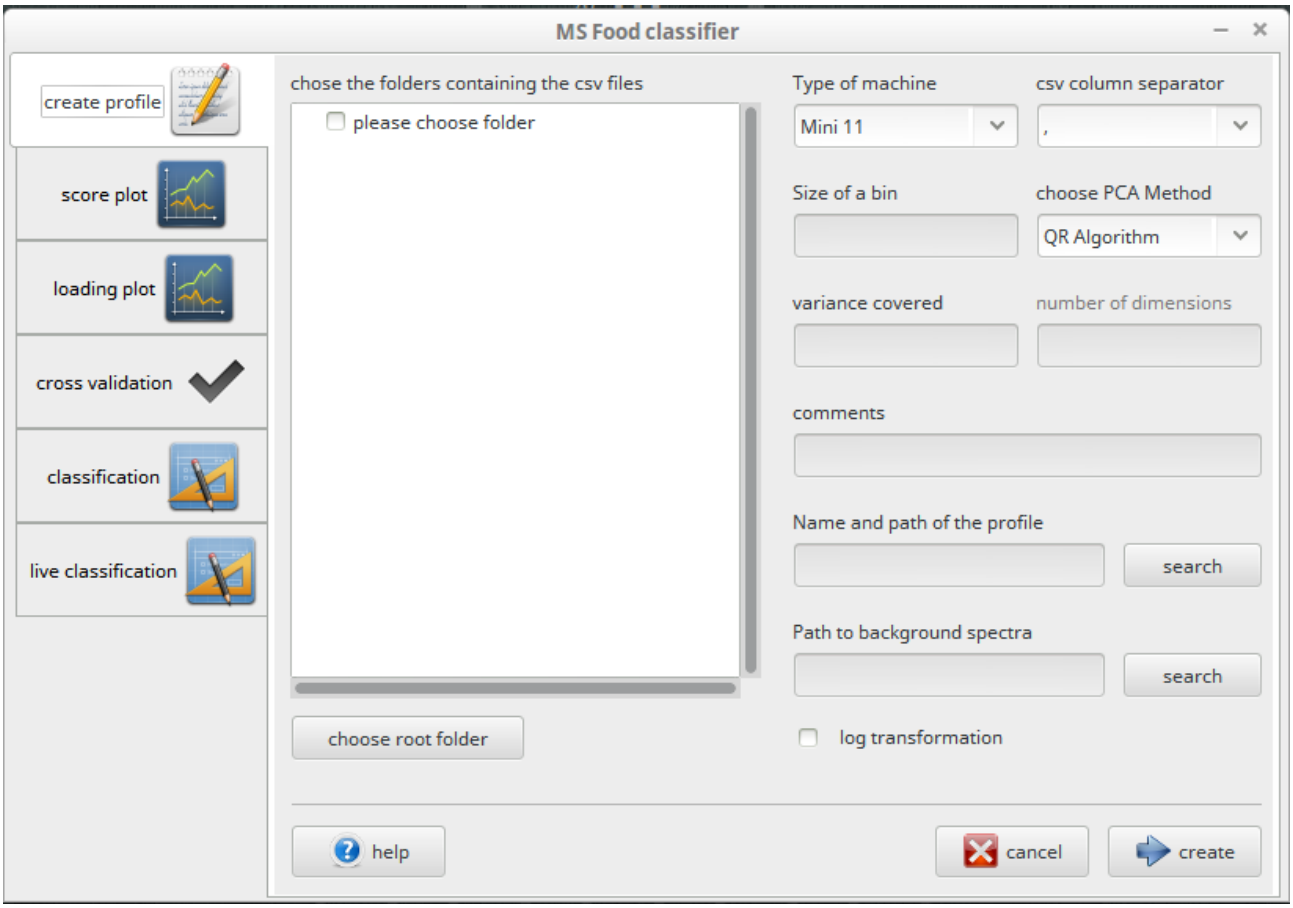

Figure S-1: User interface of the program: The starting screen of the program showing the tab for the creation of a new profile. The program is divided into six tabs (left side): creation of a new profile, creation of a score plot for an existing profile, generation of a loading plot for an existing profile, cross-validation, classification of new food samples and (live) classification of new food samples during an MS experiment

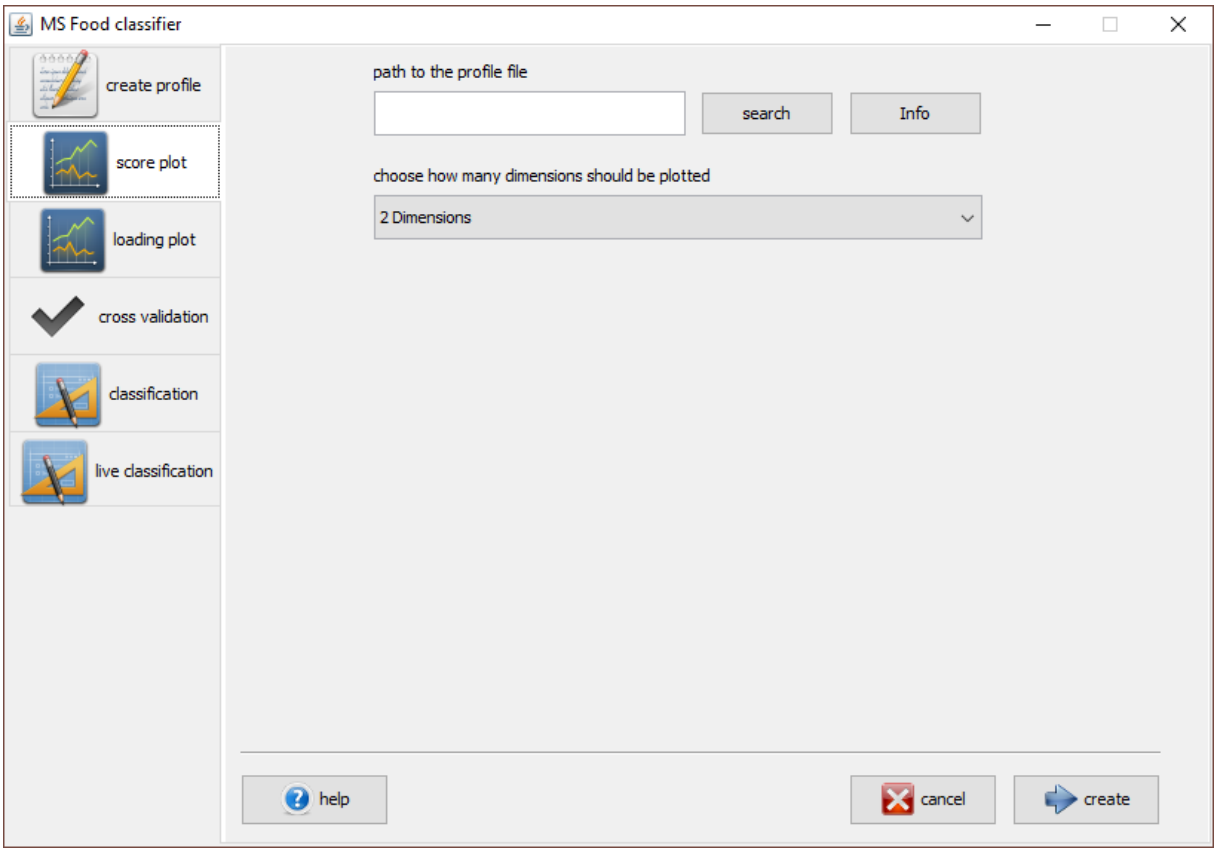

Figure S-2: Score plot tab of the program. Loading plot tab looks the same.

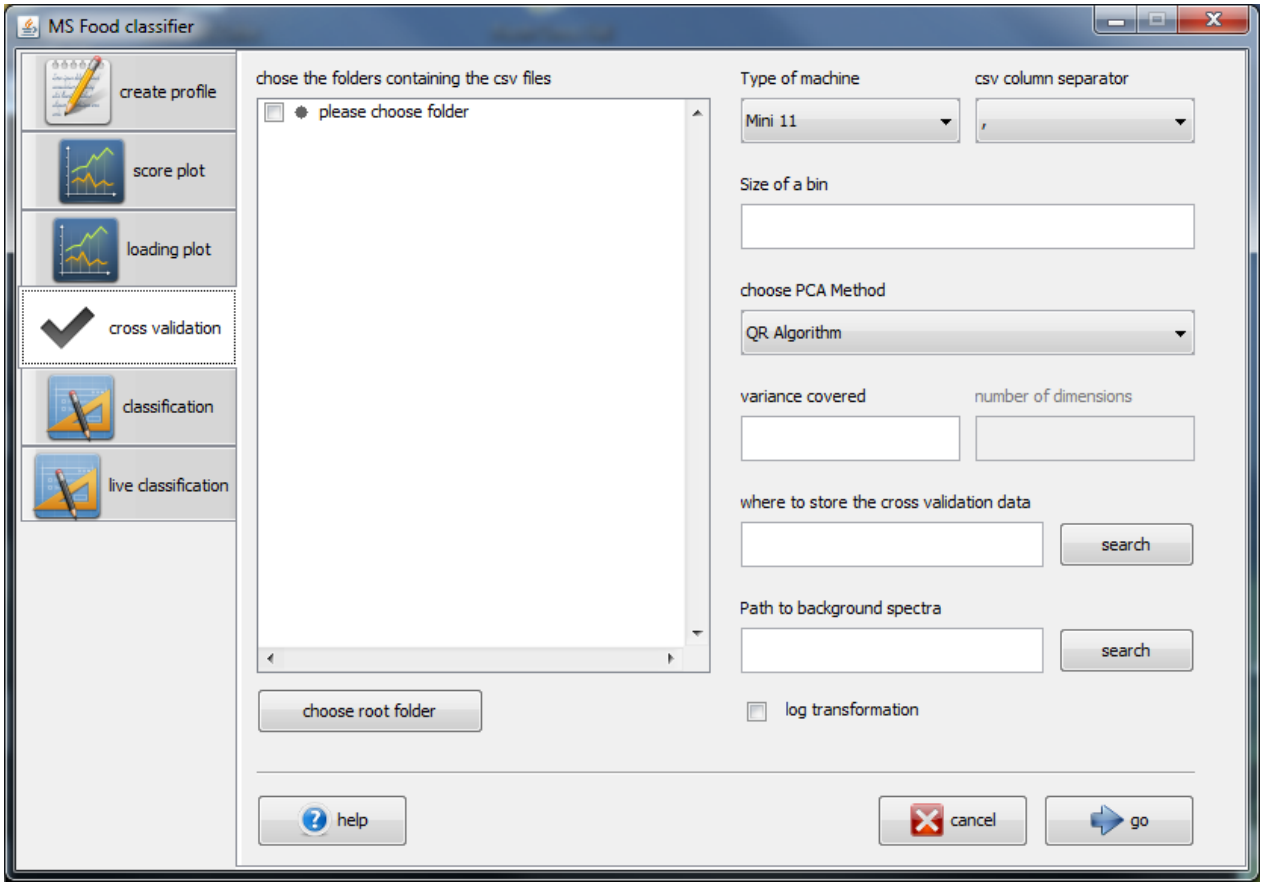

Figure S-3: Cross-validation tab of the program.

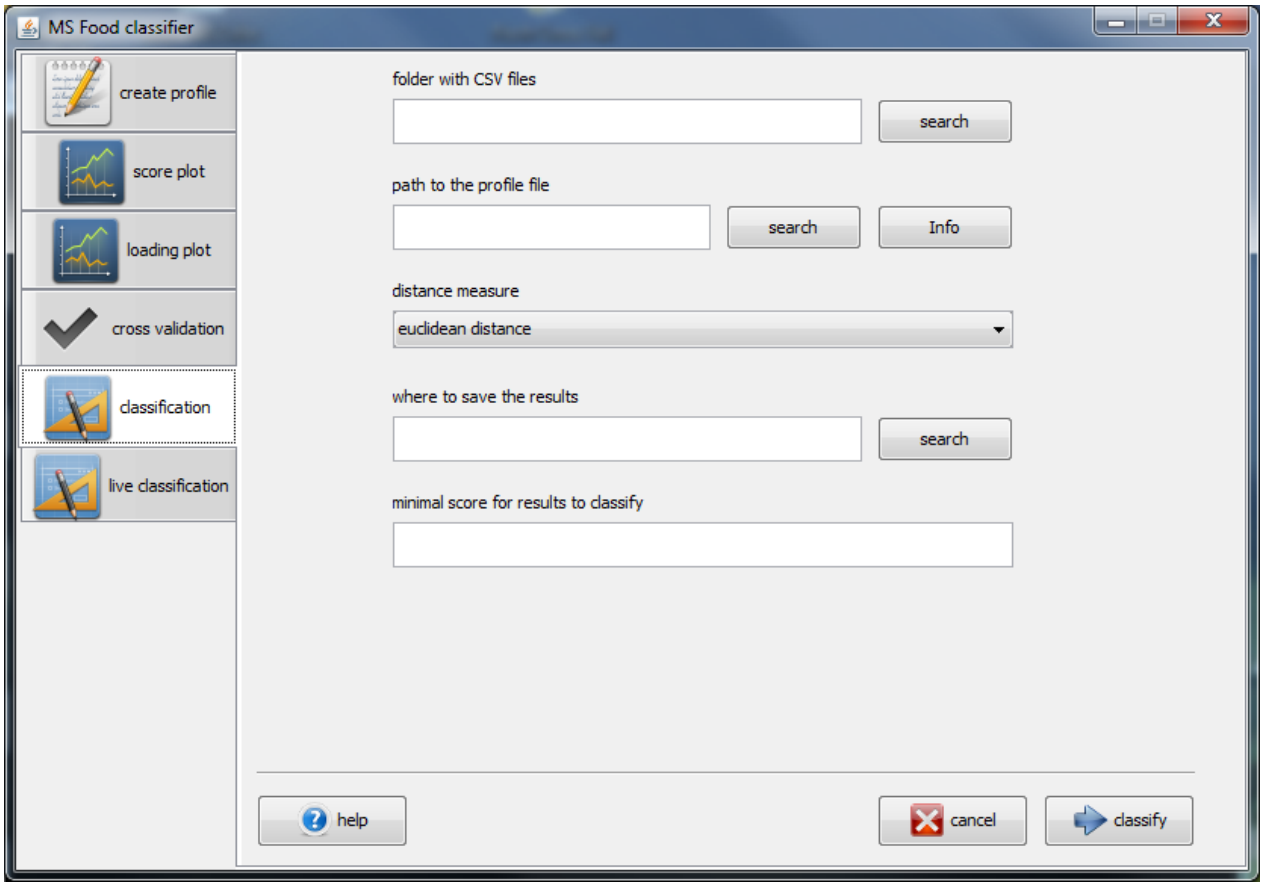

Figure S-4: Classification tab of the program. Live classification tab looks the same.

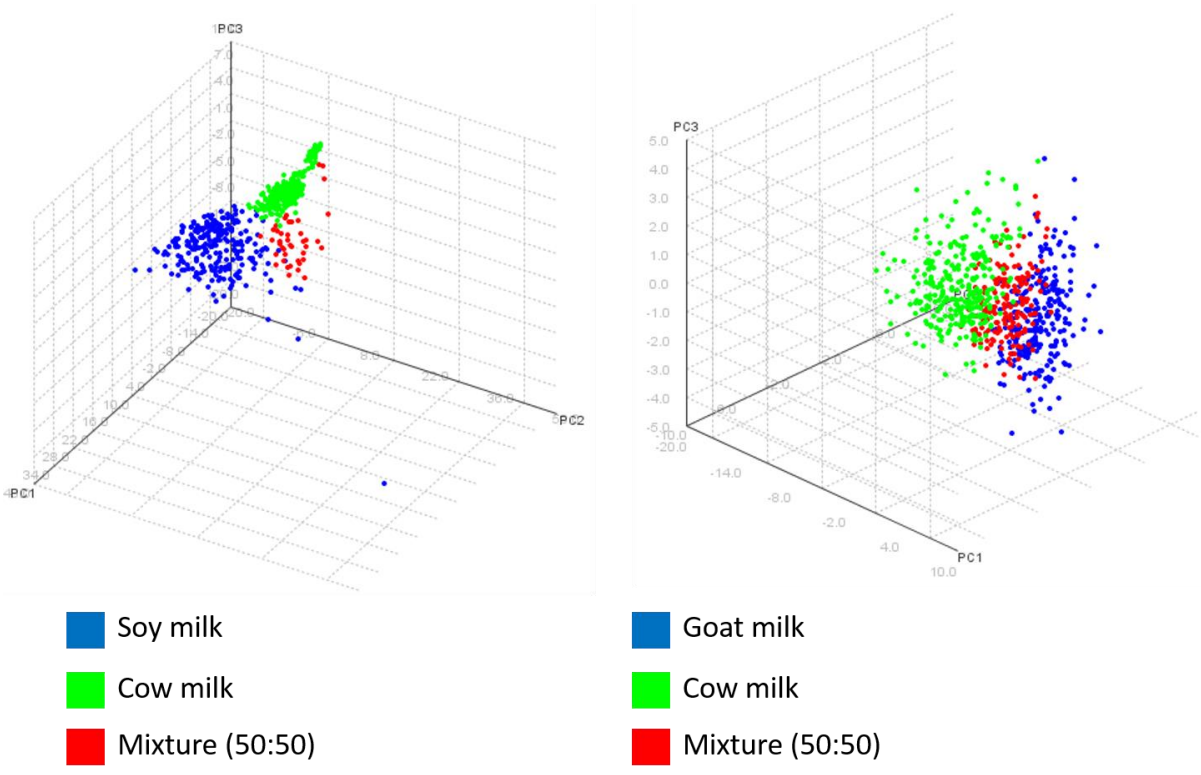

Figure S-5: PCA score plot of soy milk, cow milk and a 50:50 (v:v) mixture (left). PCA score plot of goat milk, cow milk and a 50:50 (v:v) mixture of both (right). In both cases, data points of the mixture are located in between the data points of the pure milk samples.

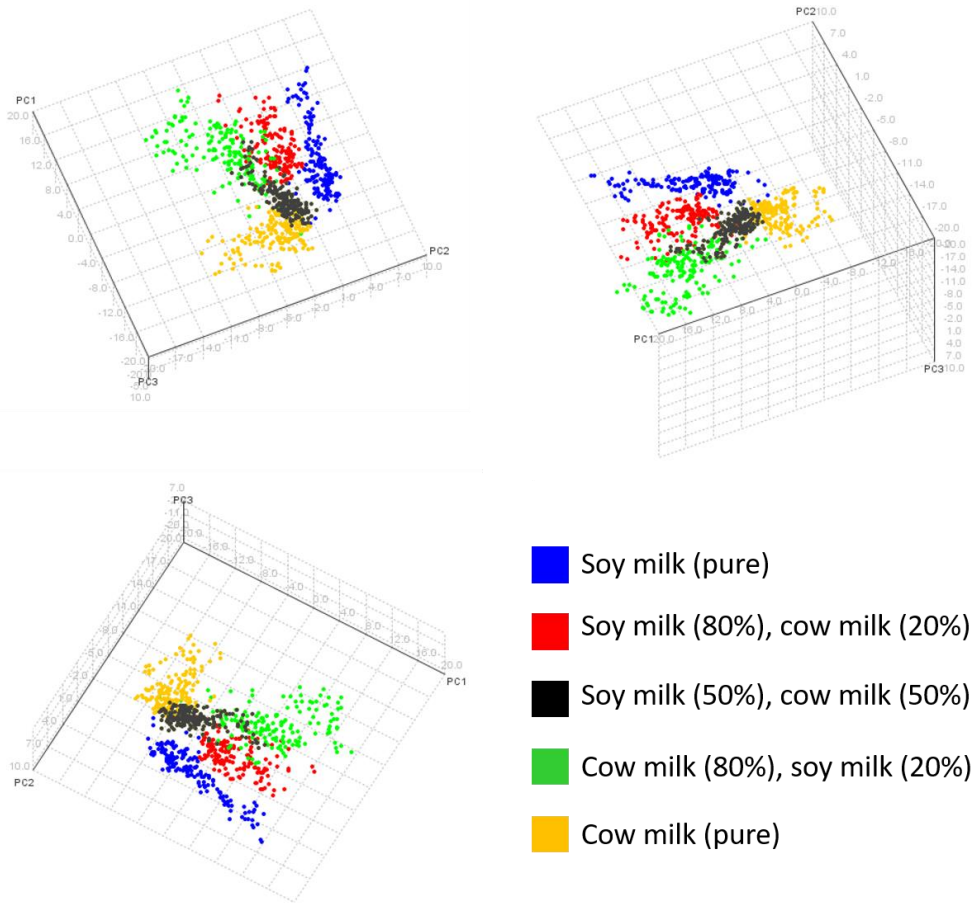

Figure S-6: PCA score plots of mixtures of cow and soy milk analyzed using ESI and the Mini 11. The pictures depict three different angles of the same 3D score plot. Around 50 spectra were recorded for each type of milk and subjected to PCA.

Processing time and memory usage of the program for the PCA transformation showed high dependence on the used bin size and PCA algorithm (Supporting Figure 7). The smaller the bin size, the higher the processing time and memory usage. This is true for both PCA algorithms, however processing time and memory usage was always higher for NIPALS algorithm compared to QR algorithm. While the mean processing times were 11.8 s, 84.6 s and 548.2 s for the bin sizes of 2, 0.5 and 0.25 u using the QR algorithm, NIPALS algorithm (60 PCs) needed 6, 3.6 and 1.4 times longer. Memory usage increased from 174.4 to 322.6 MB when decreasing bin size from 2 to 0.25 u using the QR algorithm, memory usage of NIPLAS algorithm was only 24 % and 3 % higher. Processing time for the classification showed no dependency on the used PCA algorithm or the used classifier, but on the bin size (Supporting Figure 8). Processing time for the classification of about 600 spectra increased on average from 11.6 to 13.5 to 17.6 s when decreasing the bin size from 2 to 0.5 to 0.25 u. Memory usage tended to increase with decreasing bin size for both algorithms and all classifiers ranging for QR algorithm from 138.6 MB (Euclidean distance) for a bin size of 2.0 u to 349 MB (Mahalanobis distance) for a bin size of 0.25 u. The range was smaller for NIPALS algorithm, but differences in memory usage for the three classifiers and two algorithms were non-significant in most cases. Absolute values on processing time and memory usage are given here to demonstrate the performance of the program. They strongly depend on factors such as available RAM memory of the system or frequency of garbage collection (non-deterministic in Java), but results show that the processing of reference data sets can be performed in short time even on a normal notebook.

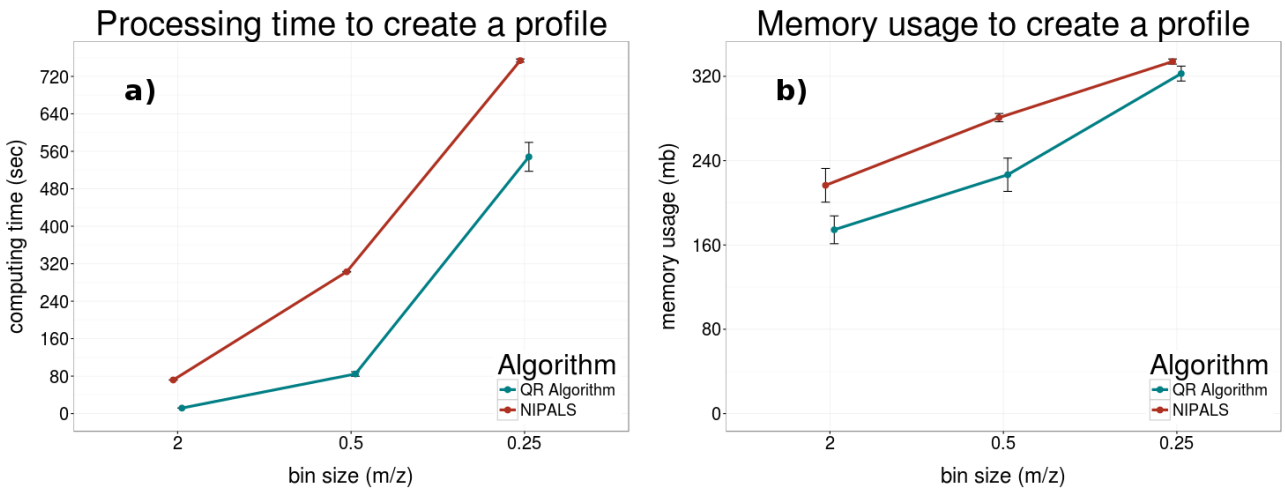

Figure S-7: Resource usage for creating a profile with bin size 2.0 *m/z*, 0.5 *m/z* and 0.25 *m/z*. a) shows the mean computing time in seconds for both QR algorithm (blue) and NIPALS algorithm (red) calculated over five iterations. b) shows the mean memory usage for creating a profile using both algorithms calculated over five iterations. For each mean value the error bars show the standard errors.

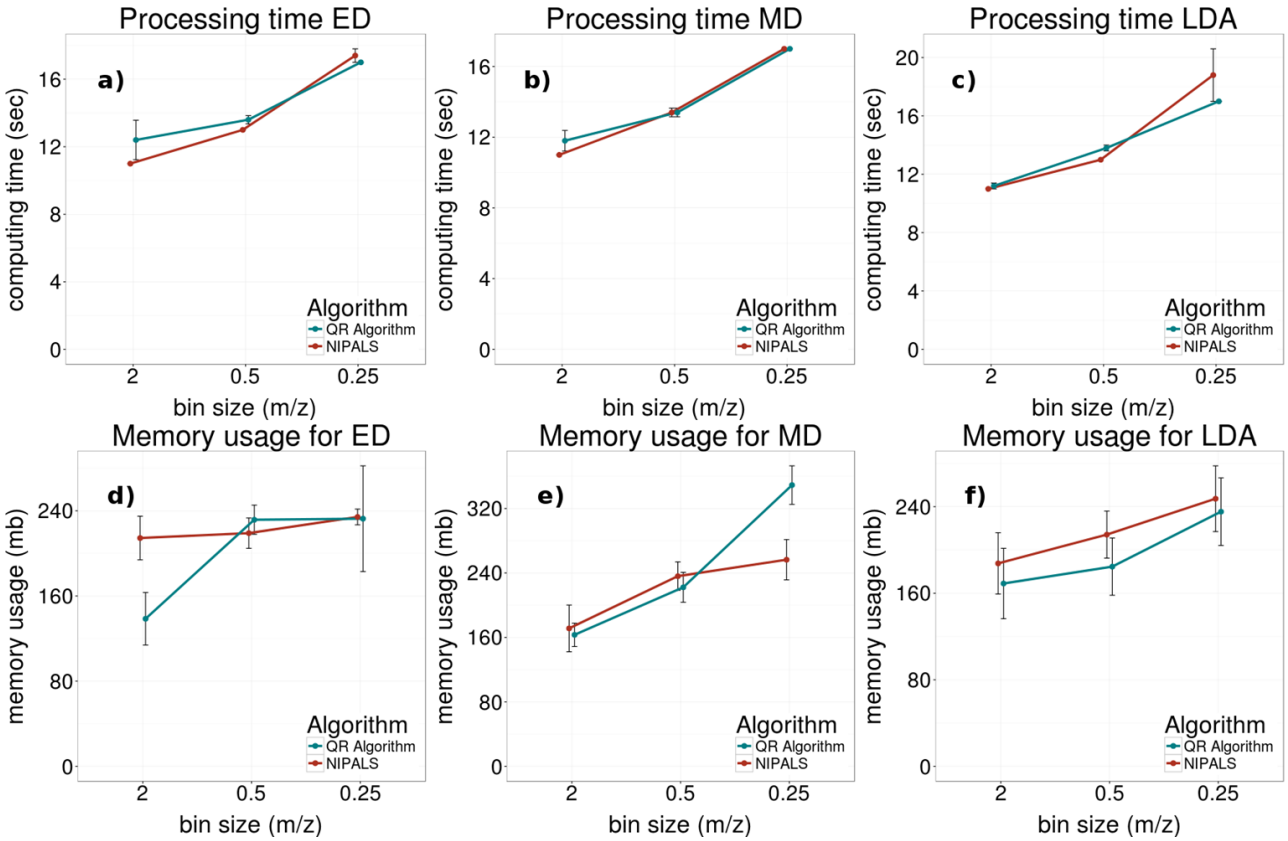

Figure S-8: Time consumption (upper row) and memory usage (lower row) to classify 657 samples. a) and d) show time consumption and memory usage of a classification using the Euclidean distance (ED), b) and e) show the same information using the Mahalanobis distance (MD) and c) and f) show the same information using the Linear Discriminant analysis (LDA) as algorithm.

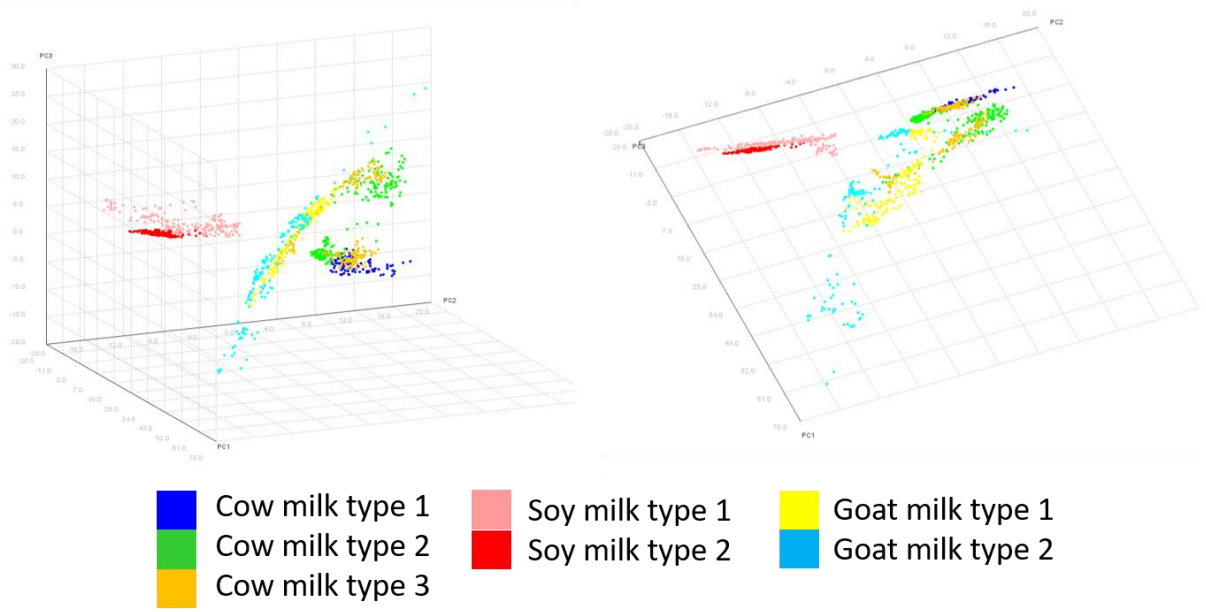

Figure S-9: PCA score plot of milk measurement of seven different samples derived from three species of milk. Different brands were purchased at the local supermarket and measured according to the parameters stated in the Experimental section. The score plot visualizes a separation between the three milk species.

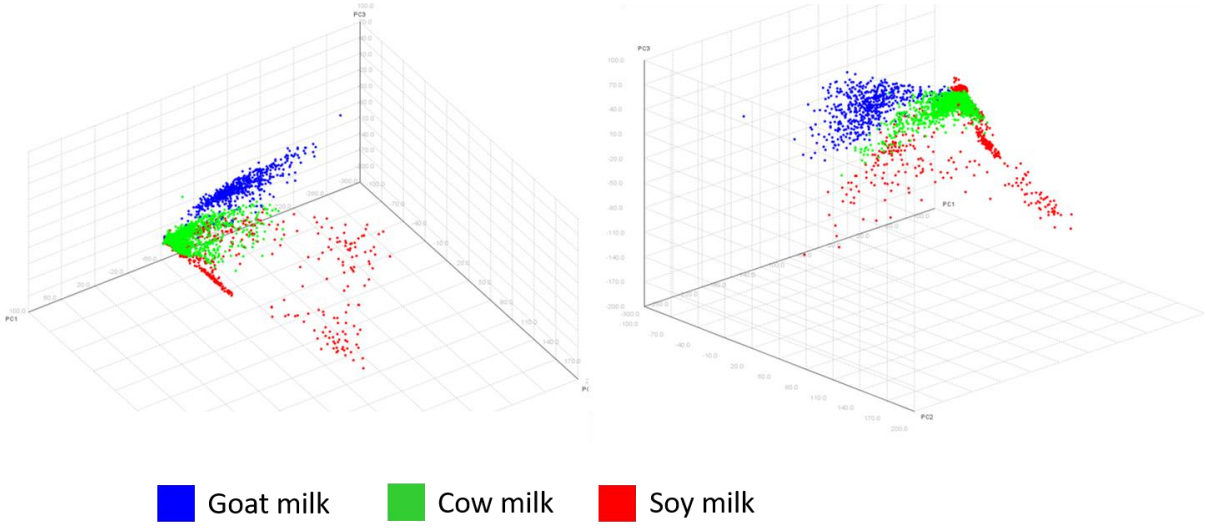

Figure S-10: PCA score plot containing spectra derived from all milk measurements performed so far. The score plot comprises around 6600 spectra and displays the separation of the three main groups of milk. This separation was achieved despite different geometrical measurement settings and a time difference of more than one year between the first and last analysis. Cross-validation yielded 54 % correct classification using Euclidean distance, 91 % using Mahalanobis distance and 86 % using LDA.

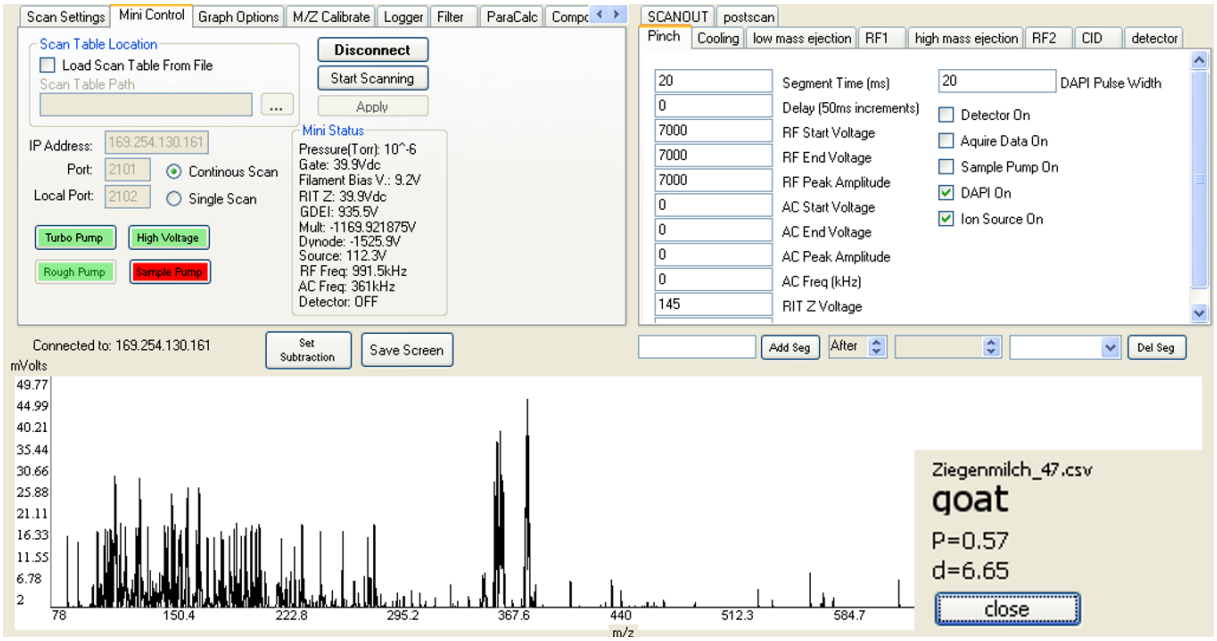

Figure S-11: Screenshot of live classification of a new mass spectrum of goat milk recorded with the Mini 11. The control software of the Mini 11 shows the currently measured spectrum and the measurement parameters. The Mini 11 is operated in full scan MS mode at 0.5 Hz. The window at the lower right displays the classification result for this spectrum. File name of the spectrum, identified milk type, confidence score and distance to the group center of the milk species are listed. Mahalanobis distance was used as a classifier here.

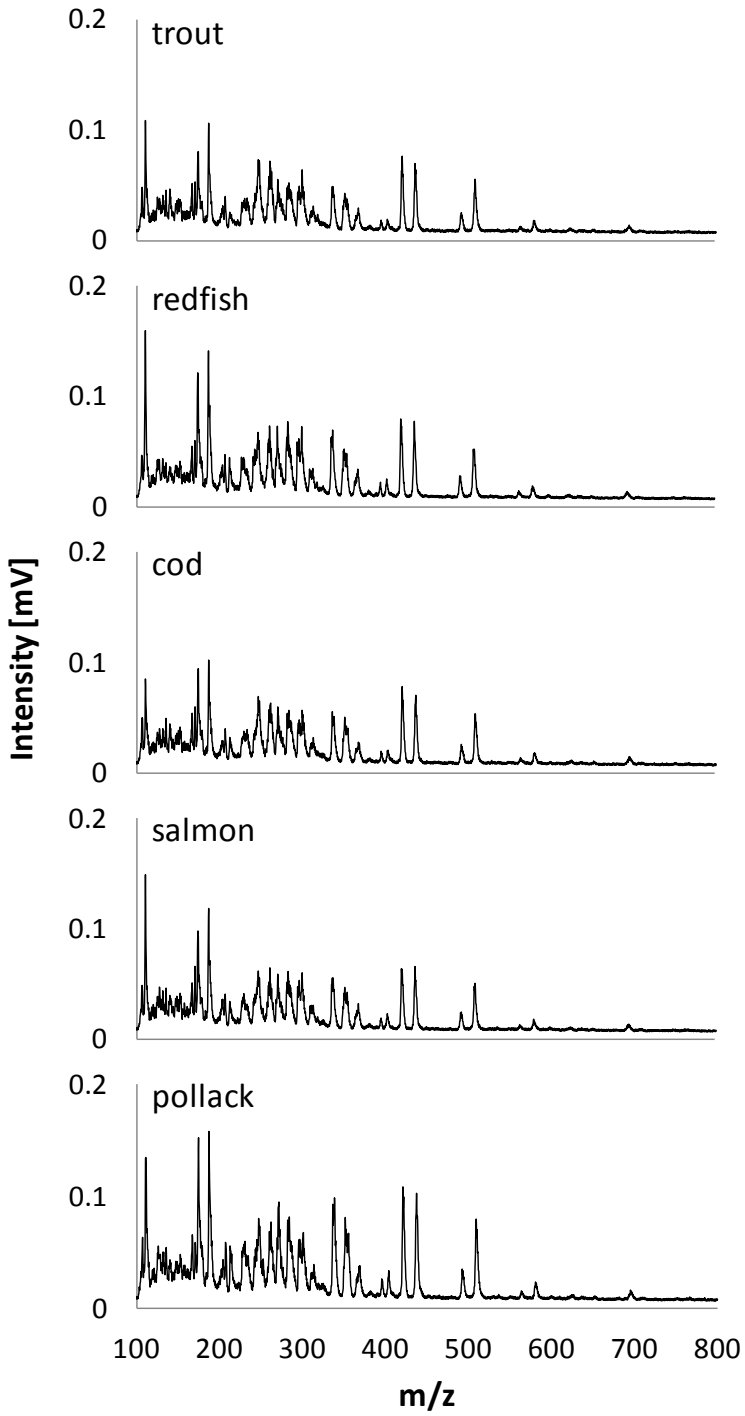

Figure S-12: DESI mass spectra of trout, redfish, cod, salmon and pollack. Average of 30 spectra.

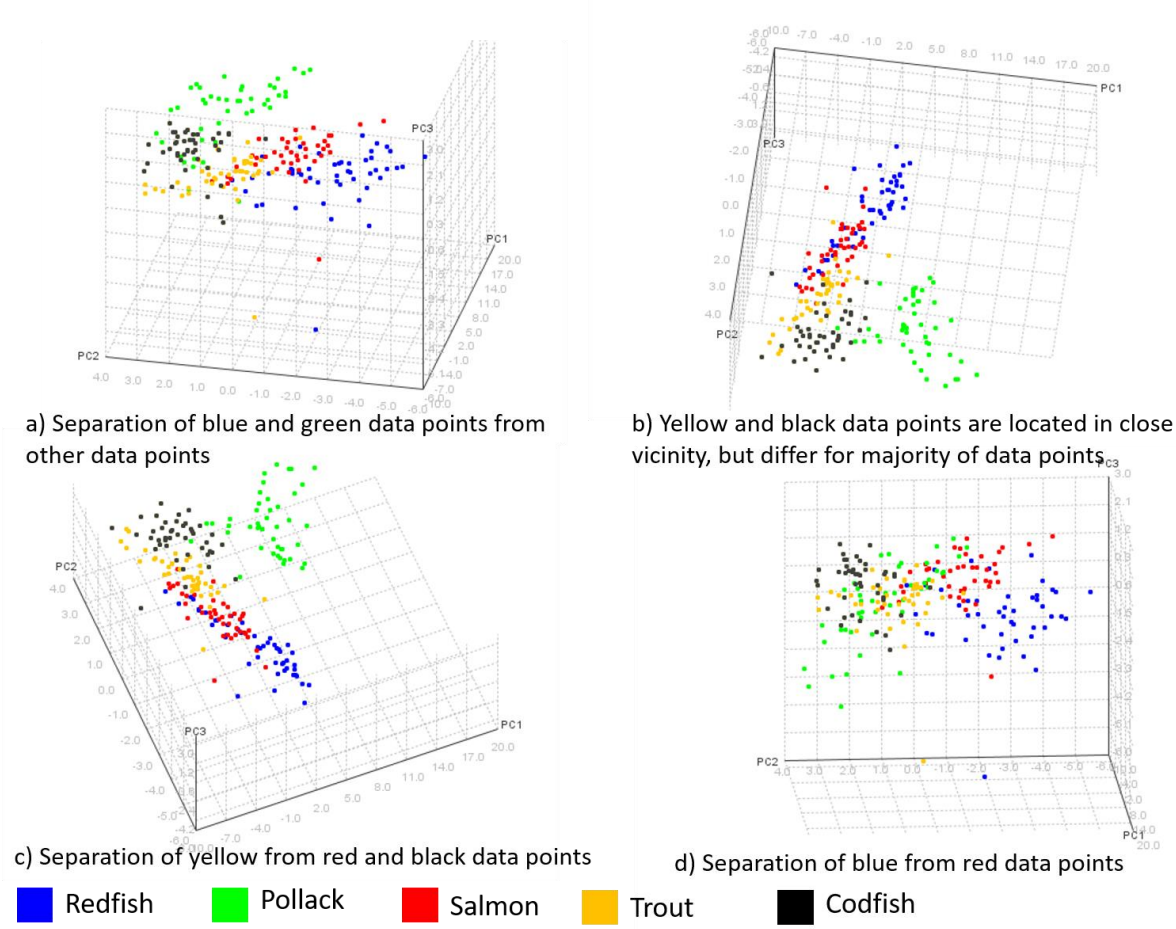

Figure S-13: PCA score plots for the separation of different types of fish. The same data is shown in Fig. 2a in the main manuscript from a different point of view. Please refer to the main manuscript for experimental details. Part a) shows the separation of redfish and pollack data points from the other species, b) depicts the separation of trout and codfish data that seem overlapping from other points of view. c) shows that the data points of codfish can be separated from pollack and trout and d) gives a detailed view on the separation of redfish and salmon data points.

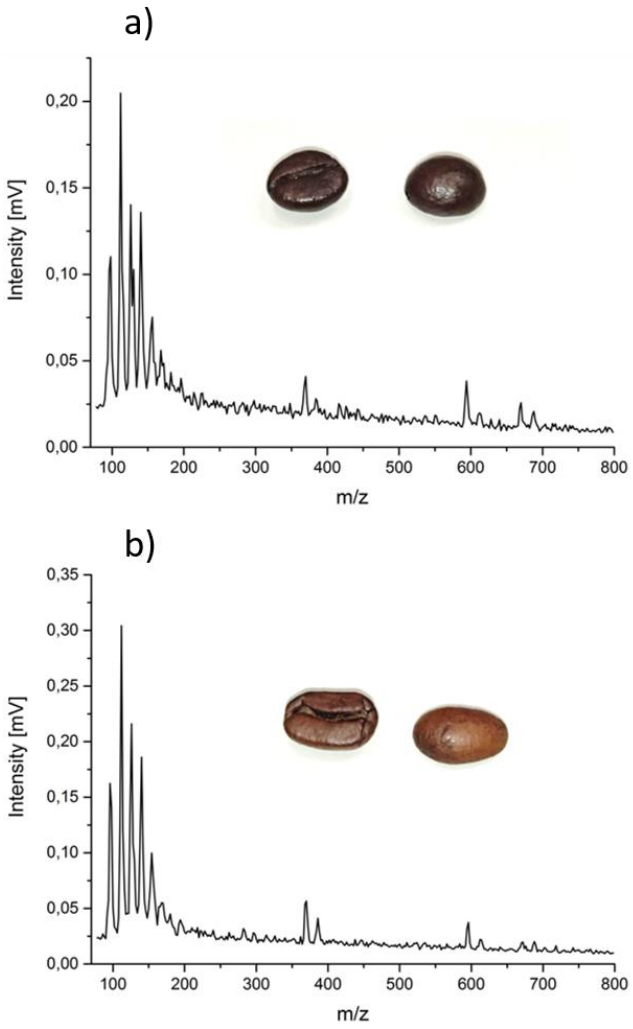

Figure S-14: LTP mass spectra of a) Robusta and b) Arabica coffee beans. Sum of 30 spectra.

	Euclidean distance	Mahalanobis distance	Linear Discriminant analysis
Cross-validation 1: 0.5 u, log-scaled, NIPALS, 60 pc	99.9%	100%	100%
Cross-validation 2: 2.0 u, log-scaled, NIPALS, 60 pc	99.9%	100%	100%
Cross-validation 3: 0.5 u, non-log, NIPALS, 60 pc	85.5%	100%	100%
Cross-validation 4: 2.0 u, non-log, NIPALS, 7 pc	85.2%	100%	100%
Cross-validation 5: 0.5 u, log-scaled, QR, 90%	99.9%	100%	100%
Cross-validation 6: 2.0 u, log-scaled, QR, 90%	99.9%	100%	100%
Cross-validation 7: 0.5 u, non-log, QR, 90%	85.5%	100%	100%
Cross-validation 8: 2.0 u, non-log, QR, 90%	85.2%	100%	100%

Table S-1: The results of all eight performed cross-validations for the milk data set.

	Euclidean distance	Mahalanobis distance	Linear Discriminant analysis
Cross-validation 1: 0.5 u, log-scaled, NIPALS, 60 pc	65.86	100	99.8
Cross-validation 2: 2.0 u, log-scaled, NIPALS, 60 pc	65.26	100	99.2
Cross-validation 3: 0.5 u, non-log, NIPALS, 60 pc	61.65	95.78	96.39
Cross-validation 4: 2.0 u, non-log, NIPALS, 7 pc	61.45	97.57	90.76
Cross-validation 5: 0.5 u, log-scaled, QR, 90%	65.86	99.8	96.18
Cross-validation 6: 2.0 u, log-scaled, QR, 90%	65.26	99.6	97.79
Cross-validation 7: 0.5 u, non-log, QR, 90%	61.85	88.96	79.12
Cross-validation 8: 2.0 u, non-log, QR, 90%	59.84	59.44	56.43

Table S-2: Results of all eight performed cross-validations for the coffee data set.

	Euclidean distance	Mahalanobis distance	Linear Discriminant analysis
Cross-validation 1: 0.5 u, log-scaled, NIPALS, 60 pc	88.2%	28.1%	99.6%
Cross-validation 2: 2.0 u, log-scaled, NIPALS, 60 pc	85.1%	82.0%	99.6%
Cross-validation 3: 0.5 u, non-log, NIPALS, 60 pc	85.1%	61.1%	98.6%
Cross-validation 4: 2.0 u, non-log, NIPALS, 7 pc	85.1%	82.0%	99.6%
Cross-validation 5: 0.5 u, log-scaled, QR, 90%	88.2%	62.0%	99.6%
Cross-validation 6: 2.0 u, log-scaled, QR, 90%	85.1%	55.2%	99.6%
Cross-validation 7: 0.5 u, non-log, QR, 90%	85.1%	58.4%	99.1%
Cross-validation 8: 2.0 u, non-log, QR, 90%	82.8%	100%	98.2%

Table S-3: Results of all eight performed cross-validations for the fish data set.