The use of δ^2H and $\delta^{18}O$ isotopic analyses combined with chemometrics as a traceability tool for the geographical origin of bell peppers

E. de Rijke [a,*], J.C. Schoorl [b], C. Cerli [b], H.B. Vonhof [c], S.J.A. Verdegaal [c], G. Vivó-Truyols [d], M. Lopatka [e,f], R. Dekter [g], D. Bakker [g], M.J. Sjerps [e,f], M. Ebskamp [g], C.G. de Koster [a]

[a] Mass Spectrometry of Biomacromolecules, Swammerdam Institute of Life Sciences, University of Amsterdam, Sciencepark 904, 1090 GE Amsterdam, The Netherlands
[b] Earth Surface Science, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Sciencepark 904, 1090 GE Amsterdam, The Netherlands
[c] Faculty of Earth and Life Sciences, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
[d] van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Sciencepark 904, 1090 GE Amsterdam, The Netherlands
[e] Korteweg-de Vries Institute, University of Amsterdam, Sciencepark 904, 1090 GE Amsterdam, The Netherlands
[f] Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands
[g] Naktuinbouw Laboratories, Sotaweg 25, P.O Box 40, 2370 AA Roelofarendsveen, The Netherlands

ARTICLE INFO

ABSTRACT

Two approaches were investigated to discriminate between bell peppers of different geographic origins. Firstly, $\delta^{18}O$ fruit water and corresponding source water were analyzed and correlated to the regional GNIP (Global Network of Isotopes in Precipitation) values. The water and GNIP data showed good correlation with the pepper data, with constant isotope fractionation of about −4. Secondly, compound-specific stable hydrogen isotope data was used for classification. Using n-alkane fingerprinting data, both linear discriminant analysis (LDA) and a likelihood-based classification, using the kernel-density smoothed data, were developed to discriminate between peppers from different origins. Both methods were evaluated using the δ^2H values and n-alkanes relative composition as variables. Misclassification rates were calculated using a Monte-Carlo 5-fold cross-validation procedure. Comparable overall classification performance was achieved, however, the two methods showed sensitivity to different samples. The combined values of δ^2H IRMS, and complimentary information regarding the relative abundance of four main alkanes in bell pepper fruit water, has proven effective for geographic origin discrimination. Evaluation of the rarity of observing particular ranges for these characteristics could be used to make quantitative assertions regarding geographic origin of bell peppers and, therefore, have a role in verifying compliance with labeling of geographical origin.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The Netherlands is the second largest exporter of agricultural products in the world (after the US) (LEI, 2013). Tomatoes, cucumbers and bell peppers (*Capsicum annuum*) are among the most important Dutch agricultural products. Within the European Union (EU), the Netherlands and Spain are the largest bell pepper producers, however, there is increasing competition from other countries such as Israel and Egypt. Increase in global trade and free markets has created an increased need for better regulation in the distribution chain of agricultural products. EU law on the provision of food information to consumers EEC No. 1169/2011 (2011) came into effect in December 2014. This law includes EU regulation 1337/2013 regarding the mandatory country of origin (COO) or place of provenance labeling for unprocessed meat from pigs, sheep, goats and poultry. This paves the way for new regulations regarding COO labeling of other food products. This regulatory shift has spurred the development of new techniques for the discrimination between foods of different origins, sources and farming systems. Stable isotope analysis of specific elements, including: hydrogen, carbon, nitrogen, oxygen, and sulfur, has been applied in food authentication for more than 20 years (Kelly, Heaton, & Hoogewerff, 2005; Rossmann, 2001). These methods are based on

stable isotope ratio measurements of a bulk product or a specific component such as an ingredient or target molecule. The measurements provide information on the botanical and geographical origin.

Fractionation of hydrogen and oxygen isotopes takes place during the evaporation and condensation processes of the water cycle: ocean water passing from liquid to vapor in the atmosphere undergoes strong isotope fractionation resulting in depletion of the heavier isotopes (^{2}H or ^{18}O) in vapor and clouds. The extent of the fractionation is dependent on the temperature, and thus is influenced by the local climate. While water is the only source of hydrogen for photosynthesis, oxygen is taken in by plants from several sources including atmospheric oxygen and carbon dioxide, and soil water. Consequently, the ^{2}H- and ^{18}O-contents of water from fruits should reflect the isotopic composition of ground water and, therefore, their place of origin. Plant physiology plays an important role in isotope fractionation processes within the plant, for example the opening or closing of stomata in response to water availability in the environment (atmospheric water and soil moisture). Evapotranspiration taking place during the maturation period of a plant also plays a factor in isotopic fractionation (Kahmen, Schefuß, & Sachse, 2013); this process causes the enrichment of heavy isotopes of oxygen and hydrogen in the water of fruits. In previous work, the stable isotopes of wine water (δ^{18}O, δ^2H) and ethanol (δ^{13}C) have been used to authenticate the location of wine production (Christoph, Rossmann, Schlicht, & Voerkelius (2009, chap. 11); Calderone & Guillou, 2008; Raco, Dotsika, Poutoukis, Battaglini, & Chantzi 2015).

In food authenticity studies, statistical analysis plays an important role in model generation and subsequent use of models for classification of unknown-origin samples (Oulhote, Le Bot, Deguen, & Glorennec 2011). These models may be either univariate (e.g. using only ^{2}H for classification) or multivariate (e.g. using ^{2}H and ^{18}O in combination with relative concentrations of several chemical compounds for classification). The variables are normally complementary (i.e. not fully correlated). Therefore, at least theoretically, the use of multivariate models should be advantageous over univariate ones. Multivariate methods normally require the extraction of latent information to reduce the dimensionality of the data and provide a rapid overview of data prior to multivariate analysis. This approach has been used in authenticity studies aiming to discriminate between organic and conventional plants (Laursen, Schjoerring, Kelly, & Husted 2014). In some techniques, the data reduction process and the classification models are intrinsically related, and completed in one step, e.g. linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA). Unsupervised dimension reduction methods such as principal component analysis (PCA) are not specifically tuned to classification as an objective, but rather describe the underlying data as a reduced dimensional projection, defined as a linear combination of the original variables. On the other hand, supervised classification methods such as LDA and PLS-DA perform dimension reduction in such a way as to maximize separation between known groups of samples. In this way, supervised methods rely on a priori assertions regarding sample groupings (Berrueta, Alonso-Salces, & Héberger 2007). Common to both unsupervised and supervised classification methods is the importance of careful model validation to avoid over-fitting and thus over-optimistic classification results (Kjeldahl & Bro, 2010).

In this study, isotope ratio mass spectrometry (IRMS) was used to develop a traceability tool for the geographical origin of bell peppers in both compound-specific (^{2}H/^{1}H) and bulk (^{18}O/^{16}O) analysis mode. The ^{18}O/^{16}O isotope ratios of the bell pepper fruit water and corresponding source water was determined using a continuous flow IRMS technique. Because the bell pepper fruit water samples contained high concentrations of dissolved organic

substances, we use an isotope equilibration technique by which the ^{18}O/^{16}O isotope ratio of the fruit water is determined. This was achieved by analyzing the ^{18}O/^{16}O isotope ratio (δ^{18}O value) of CO_2 gas after isotope equilibration with the fruit water samples. This equilibration technique (Epstein & Mayeda, 1953) is a widely accepted method for high-precision δ^{18}O analysis of water samples. It works particularly well for waters that are difficult to analyze with other techniques due to high salt or dissolved organic content. Furthermore, the ^{2}H/^{1}H isotope ratio of n-alkanes in the surface wax layer of the peppers was determined using gas chromatography coupled to IRMS (GC-C-IRMS). The lipid composition of different varieties of peppers was recently investigated by Parsons et al. (2013). Isotope ratios vary per lipid and give a good reflection of the isotopic composition of the meteorological spring water in different plant species; moreover, the hydrogen atoms are not likely to exchange due to transpiration during transport and sample handling (McInerney, Helliker, & Freeman 2011; Sachse, Gleixner, Wilkes, & Kahmen 2010).

On the basis of n-alkane fingerprinting data, two chemometric classification models were developed and compared based on their ability to discriminate between peppers from different countries of origin. Firstly, an LDA model was developed, which is a model that has been used in many food authenticity studies (e.g. Kelly et al. 2005; Laursen et al. 2014). Secondly, a naïve Bayesian classifier using kernel-density estimation of the variables was evaluated using the same data set. Since the problem of food authenticity is also of forensic interest, probabilistic assertions regarding classification outcomes are particularly useful. A common strategy for "soft" classification is the use of a likelihood ratio (LR) approach. This has been performed in both (LDA and kernel-density-based) classifiers. Similar methods using LR were recently developed for wine authentication (Martyna, Zadora, Stanimirova, & Ramos 2014). The classification performances of both models were externally validated using cross-validation.

2. Material and methods

2.1. Samples, standards and software

Dutch and Spanish peppers were collected from two greenhouses in the Netherlands (Naktuinbouw, Roelofarendsveen; Helderman, Middenmeer) and two greenhouses in Spain (Enza seeds, Almeria; IMIDA, Murcia). Italian (Sicily) and Israeli peppers were collected from a local Italian Bio market, and a Dutch supermarket (Albert Heijn, Amsterdam). The Dutch supermarket-bought peppers specified Israel as the country of origin. The number of samples (n) indicated below is per sampling date. Peppers ($n = 20$) and water samples ($n = 8$) from the Roelofarendsveen greenhouse were collected each month from July 2013 to September 2014; peppers ($n = 20$) and water samples ($n = 4$) from the Middenmeer greenhouse were collected monthly from March to September 2014. The Spanish peppers ($n = 20$) and water samples ($n = 2$) were collected in May and August 2014 in both Murcia and Almeria; Israeli and Italian peppers ($n = 5$ and $n = 6$, respectively) were bought in May; no corresponding water was available for these samples as they were obtained from supermarkets. All other water samples were collected from the main water supply in the greenhouses on the day the peppers were harvested. In the greenhouse in Middenmeer, water (and pepper) samples were collected from four different sections to test the intra-greenhouse variability. Water was sampled in this greenhouse both at the supply and at the drainage point of each section. As a control, the rainwater collection basin was also sampled as well as water from the underground storage and window condensate. The underground storage contained rainwater that is collected and stored in antici-