Authentifizierung pflanzlicher Lebensmittel mittels *Direct Analysis in Real Time*-Massenspektrometrie

DISSERTATION

Zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Chemie

der Universität Hamburg

HAMBURG SCHOOL OF FOOD SCIENCE

Institut für Lebensmittelchemie

vorgelegt von

Felix Philipp Schmauder

Hamburg, 2025

1. Gutachter der Dissertation Herr Prof. Dr. Markus Fischer

2. Gutachter der Dissertation Herr Prof. Dr. Hartmut Schlüter

Datum der Disputation: 10.04.2026

Datum der Druckfreigabe: 10.04.2026

I. Publikationen

Peer-Reviewte Publikationen

F. Schmauder, D. Schütz, M. Creydt, J. Riedl, M. Fischer, Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS). Food Control 163, 110489, 2024.

F. Schmauder, M. Creydt, M. Fischer, Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (*Corylus avellana* L.). Food Chemistry 467, 142265, 2025.

J. Brockelt, F. Schmauder, K. Brettschneider, M. Creydt, S. Seifert, M. Fischer, Competing technologies: determining the geographical origin of strawberries (*Fragaria× ananassa*) using laboratory based near-infrared spectroscopy compared to a simple portable device. Molecular Omics 21, 7–18, 2025.

F. Schmauder, J. Brockelt, K. Brettschneider, S. Seifert, M. Creydt, M. Fischer, DART-MS Approaches for Metabolomics-Based Authentication: Challenges in Determining the Geographical Origin of Strawberries (*Fragaria× ananassa*). Talanta Open 12, 100496, 2025.

J. Brockelt, F. Schmauder, M. Creydt, M. Fischer, Lipid profiling: proving the geographical origin of strawberries (*Fragaria× ananassa*) using a non-targeted LC-IM-MS approach. Molecular Omics 21, 353-364, 2025.

Übersichtsartikel

F. Schmauder, M. Creydt, M. Fischer, Food Authentication goes Green: Geographische Herkunftsbestimmung von Körnermais (*Zea mays* L.) mittels Direct Analysis in Real Time-Massenspektrometrie, Deutsche Lebensmittel Rundschau, 09/2024.

F. Schmauder, M. Creydt, M. Fischer, Grüne Lebensmittelanalytik mittels DART-MS, LABO, 11/2024.

M. Creydt, J. Brockelt, F. Schmauder, T. Kehrer, M. Fischer (2025), Herkunftskontrolle bei Erdbeeren: Wie zuverlässig ist „Made in Germany“?, Der Lebensmittelbrief, 05/2025.

Vorträge

F. Schmauder, M. Creydt, M. Fischer, Food authentication goes green: Herkunftsauthentifizierung mittels Direct Analysis in Real Time-Massenspektrometrie. Deutsche Lebensmittelchemietage, 16.-18.09.2024, Freising.

Posterbeiträge

F. Schmauder, M. Creydt, S. Panto, S. Vinnenberg, L. Dubois, M. Fischer, Food Authentication Herkunftsbestimmung von weißem Spargel mittels GC×GC-TOF-MS und chemometrischer Auswerteverfahren. Deutscher Lebensmittelchemikertag, 19.-22.09.2022, Hamburg.

F. Schmauder, D. Schütz, M. Creydt, J. Riedl, M. Fischer, Food Authentication Herkunftsbestimmung von Körnermais mittels DART TOF-MS und chemometrischer Auswerteverfahren. Deutsche Lebensmittelchemietage, 21.-23.08.2023, Bonn.

F. Schmauder, D. Schütz, M. Creydt, J. Riedl, M. Fischer, DART-MS zur geographischen Herkunftsbestimmung von Körnermais (*Zea mays* L.). Arbeitstagung des Regionalverbandes Nord, 22.-23.04.2024, Bremerhaven.

F. Schmauder, M. Creydt, M. Fischer, Geographische Herkunftsbestimmung von Haselnüssen (*Corylus avellana* L.) mittels Direct Analysis in Real Time-Massenspektrometrie (DART-MS) und chemometrischer Auswertung. Deutsche Lebensmittelchemietage, 16.-18.09.2024, Freising.

K. Brettschneider, J. Brockelt, F. Schmauder, M. Creydt, M. Fischer, S. Seifert, Anwendung mobiler und stationärer NIR-Geräte zur Authentifizierung von Erdbeeren: Ein Vergleich. Arbeitstagung des Regionalverbandes Nord, 10.-11.03.2024, Kiel.

II. Inhaltsverzeichnis

I. Publikationen V

II. Inhaltsverzeichnis VII

III. Abkürzungsverzeichnis XI

1 Zusammenfassung 1

2 Abstract 3

3 Einleitung 5

4 Theoretische Grundlagen 7

4.1 Erdbeeren 7

4.1.1 Anbau 7

4.2 Körnermais 8

4.3 Haselnüsse 9

4.4 Lebensmittelschutz 9

4.4.1 Lebensmittelsicherheit 9

4.4.2 Lebensmittelfälschung 10

4.5 Lebensmittelauthentifizierung 11

4.5.1 Die *Omics*-Kaskade 11

4.5.2 Analytische Ansätze zur Herkunftsbestimmung 13

4.5.2.1 Metabolomics-basierte Ansätze zur Herkunftsbestimmung 14

4.5.2.1.1 Massenspektrometrische Methoden 15

4.5.2.1.2 Elektrospray-Ionisation 15

4.5.2.1.3 Time-of-flight-Massenanalysatoren 16

4.5.2.2 Analytische Schnellmethoden zur Herkunftsbestimmung 17

4.5.2.2.1 Direct Analysis in Real Time-Ionenquellen 17

4.5.2.2.2 DART-qToF-MS 18

4.5.2.2.3 Signalidentifizierung mittels DART-qToF-MS 19

4.5.2.2.4 Fourier-Transformations-Nahinfrarotspektroskopische Methoden 19

4.6 Multivariate Statistische Verfahren 20

4.6.1 Datenprozessierung von DART-MS Daten 21

4.6.2 Datenprozessierung von FT-NIR Daten 22
4.6.3 Hauptkomponentenanalyse 23
4.6.4 Random Forest 24
4.6.5 Lineare Diskriminanzanalyse 24
4.6.6 Validierungsstrategien 25
4.6.7 Variablenwichtigkeit und Variablenselektion 26
5 Zielsetzung der Arbeit 29
6 Kumulativer Teil der Dissertation 31
6.1 Herkunftsauthentifizierung von Erdbeeren (*Fragaria* × *ananassa*) mittels DART-MS. 31
6.2 Herkunftsauthentifizierung von Körnermais (*Zea mays* L.) mittels DART-MS. 41
6.3 Herkunftsauthentifizierung von Haselnüssen (*Corylus avellana* L.) mittels DART-MS. 53
6.4 Herkunftsauthentifizierung von Erdbeeren (*Fragaria* × *ananassa*) mittels FT-NIR Spektroskopie. 65
7 Zusammenfassende Diskussion 79
7.1 Herkunftsauthentifizierung von Erdbeeren 79
7.1.1 Methodenoptimierung 79
7.1.2 Authentifizierung von Erdbeeren 81
7.2 Herkunftsauthentifizierung Körnermais 84
7.3 Herkunftsauthentifizierung von Haselnüssen 88
7.4 DART-MS zur Herkunftsunterscheidung 92
7.5 Vergleich der Authentifizierung mittels DART-MS und FT-NIR 96
7.6 Limitationen und Ausblick 100
8 Literaturverzeichnis 103
9 Anhang 121
9.1 Supporting Information zu: DART-MS approaches for metabolomics-based authentication: Challenges in determining the geographical origin of strawberries (*Fragaria* × *ananassa*) 121

9.2 Supporting Information: Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS) 138

9.3 Supporting Information zu: Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (*Corylus avellana* L.) 158

9.4 Supporting Information zu: Competing technologies: determining the geographical origin of strawberries (*Fragaria* × *ananassa*) using laboratory based near-infrared spectroscopy compared to a simple portable device 175

9.5 Verwendete Gefahrstoffe nach GHS 188

9.6 Zeitschriftenartikel: Food Authentication goes Green: Geographische Herkunftsbestimmung von Körnermais (*Zea mays* L.) mittels Direct Analysis in Real Time-Massenspektrometrie 190

9.7 Zeitschriftenartikel: Grüne Lebensmittelanalytik mittels DART-MS 197

10 Danksagung 201

11 Eidesstattliche Erklärung 202

III. Abkürzungsverzeichnis

ANOVA	Varianzanalyse (engl. *analysis of variance*)
CCS	Kollisionsquerschnitt (engl. c*ollision cross secton*)
CV	Kreuzvalidierung (engl. c*ross-validation)*
DART	Direkte Ionisationsquelle (engl. *direct analysis in real time*)
DART-qToF	Flugzeitmassenspektrometer gekoppelt mit einem DART (engl. d*irect analysis in real time quadrupole time-of-flight mass spectrometer*)
ESI	Elektrospray-Ionisation (engl. *electrospray ionization*)
EHEC	Enterohämorrhagische Escherichia coli
FT-NIR	Fourier-Transformations-Nahinfrarotspektroskopie (engl. *Fourier-transform near-infrared spectroscopy*)
g.g.A.	geschützte geographische Angabe
g.U.	geschützte Ursprungsbezeichnung
GC	Gaschromatographie
GC-ToF	Gaschromatographie gekoppelt mit Flugzeitmassenspektrometer (*engl. gas chromatographie time-of-flight mass spectrometer*)
GHS	Global harmonisiertes System (engl. globally harmonized system)
HPLC	Hochleistungsflüssigkeitschromatographie (engl. *high performance liquid chromatographie*)
IMS	Ionen-Mobilitäts-Spektrometer (engl. *ion mobility spectrometer*)
LC	Flüssigkeitschromatographie (engl. *liquid chromatography*)
LC-MS	Flüssigkeitschromatographie-Massenspektrometer (engl. *liquid chromatography mass spectrometer*)
LDA	Lineare Diskriminanzanalyse (engl. *linear discriminant analysis)*
m/z	Masse-zu-Ladung-Verhältnis
MS	Massenspektrometer
MS/MS	Tandem-Massenspektrometrie
MSC	Multiplikative Streulichtkorrektur (engl. *multiplicative signal correction*)
NIR	Nahinfrarotspektroskopie (engl. *near-infrarated spectroscopy*)
PC	Hauptkomponente (engl. *principal componant*)
PCA	Hauptkomponentenanalyse (engl. p*rincipal componant analysis)*
Q	Quadrupol
QQQ	Triple-Quadrupolsystem
qToF	Quadrupol Flugzeitmassenspektrometer (engl. *quadrupol time-of-flight mass spectrometer*)

RF	Random Forest (engl. *Random Forest*)
SPE	Festphasenextraktion (engl. *solid phase extraction*)
SVM	Multivariates Klassifikationsverfahren (engl. *support vector machine*)
ToF	Flugzeitmassenspektrometer (engl. *time-of-flight mass spectrometer*)
UHPLC	Ultra-Hochleistungsflüssigkeitschromatographie (engl. *ultra* high *performance liquid chromatography*)
UV	Ultraviolett

Maßeinheiten und Größen

$	Dollar
%	Prozent
€	Euro
°C	Grad Celsius
µ	mikro
cm	Zentimeter
eV	Elektronenvolt
g	Gramm
h	Stunde
Hz	Herz
kg	Kilogramm
L	Liter
m	Meter
m	milli
m	Masse
m/z	Masse-zu-Ladungsverhältnis
min	Minute
mol	Mol
n	nano
p	piko
rpm	Umdrehungen pro Minute
s	Sekunde
z	Ladung

1 Zusammenfassung

Personen, Unternehmen und die Gesellschaft benötigen zuverlässige, schnelle und kosteneffiziente Methoden zur Bestimmung der Herkunft von Lebensmitteln. Diese Kontrollmethoden können betrügerische Praktiken wie Lebensmittelverfälschung verhindern, welche einen jährlichen Schaden von 10–15 Milliarden Dollar verursachen (Johnson 2014). Daher beschäftigt sich diese Dissertation mit der Entwicklung und Etablierung neuer analytischer Verfahren zur Authentifizierung von Lebensmitteln, wobei der Fokus auf schnellen, umweltfreundlichen und kostengünstigen Verfahren liegt.

Das zentrale Verfahren dieser Arbeit ist die *Direct Analysis in Real Time*-Massenspektrometrie (DART-MS), die es ermöglicht, Proben ohne chromatographische Trennung, direkt zu ionisieren. Diese Technik zeigt das Potential für eine disruptive Leistungssteigerung der chemischen Analytik, da sie die Analysenzeit von Lebensmitteln erheblich verkürzt und den Einsatz von umweltschädlichen Chemikalien minimiert.

Diese Dissertation betrachtet insbesondere die Anwendung der DART-MS-Kopplung für die *Metabolomics*-basierte Authentifizierung von Lebensmitteln. Die Disziplin der *Metabolomics*-Forschung betrachtet kleine Stoffwechselprodukte (*small molecules*), wobei die Konzentrationsunterschiede ausgewählter Moleküle für die Authentifizierung von Lebensmitteln herangezogen werden können. Dieser Ansatz wird im Rahmen dieser Dissertation zur Herkunftsbestimmung von drei verschiedenen Lebensmitteln angewendet: (i) Erdbeeren als wasserreiches Lebensmittel, (ii) Körnermais als kohlenhydratreiches Lebensmittel und (iii) Haselnüsse als fettreiches Lebensmittel. Diese Auswahl an verschiedenen Lebensmitteln bietet die Möglichkeit einer umfassenden Bewertung der Leistungsfähigkeit von DART-MS Systemen.

Die Arbeit hat zum Ziel, einen Arbeitsablauf zur *Metabolomics*-basierten Lebensmittelauthentifizierung mittels DART-MS zu entwickeln und anhand der drei genannten Datensätze zu testen. Das Vorgehen umfasst die Methodenoptimierung und Analyse (Erdbeeren, Körnermais, Haselnüsse) sowie die Entwicklung eines *Open-Source*-Arbeitsablaufes zur Datenanalyse.

Die Ergebnisse demonstrieren, dass die Herkunftsbestimmung mittels DART-MS-basierter Analytik in vielen Fällen erfolgreich realisiert werden konnte. Für die Erdbeerproben ergibt sich eine Richtigkeit der Zuordnungen der Herkunft von 63,8%, während Körnermais mit 84,4% und Haselnüsse mit 93,2% richtig klassifiziert wurden. Die Unterschiede in den Richtigkeiten zeigen, dass die Herkunft von kohlenhydratreichen und fettreichen Lebensmitteln mittels DART-MS bestimmt werden kann. Die Richtigkeit der Klassifizierung der Herkunft von Erdbeeren hebt die Herausforderungen bei der Analyse von wasserreichen Lebensmitteln

hervor. Aus allen Datensätzen konnten, für die Herkunftsbestimmung wichtige Marker, identifiziert werden. Dies wurde durch die Aufklärung der Fragmentierungsmuster verschiedener Stoffklassen mittels Standards und DART-MS/MS Experimenten gewährleistet. Über alle Datensätze hinweg werden Di- und Triacylglyceride als herkunftsrelevante Verbindungen identifiziert und zusätzlich können Tocopherole als relevante Marker für die Herkunft von Haselnüssen bestimmt werden.

Neben der DART-MS-basierten Analytik wird der entwickelte Arbeitsablauf mit einer alternativen Schnellmethode, der Fourier-Transformations-Nahinfrarotspektroskopie (FT-NIR), am Beispiel der Erdbeeren, verglichen. Hierfür werden diese Proben zusätzlich mittels FT-NIR analysiert, klassifiziert und die Ergebnisse der DART-MS-basierten Analytik gegenübergestellt. Dieses Experiment zeigt Vorteile der FT-NIR Analytik bei der Probenvorbereitung und der einfachen Handhabung des Analysengerätes, während der DART-MS-basierte Ansatz große Vorteile bei der gezielten Markeridentifizierung und der Sensitivität der Methode bietet.

Abschließend werden die Limitationen der DART-MS Analytik diskutiert und Bedingungen für eine Übertragung in die Routineanalytik identifiziert. Diese Arbeit leistet einen Beitrag zur Etablierung von DART-Ionenquellen auf dem Feld der Lebensmittelauthentifizierung auf dem Weg zu schnellerer, kostengünstigerer und umweltfreundlicherer Analytik.

2 Abstract

It is evident that reliable, efficient and cost-effective methods for determining the provenance of food are essential for individuals, companies and society. Such methods are crucial in order to avoid fraudulent practices such as food adulteration, which have been shown to cause an annual loss of 10–15 billion dollars (Johnson 2014). The present work is dealing with the development and establishment of new analytical methods for the authentication of food, with a particular focus on rapid, cost-efficient and ecological methods.

The main method of this work is Direct Analysis in Real Time-Mass Spectrometry (DART-MS), which can ionize samples directly without the necessity of chromatographic separation. This technique demonstrates the potential for an enhancement in the performance of chemical analysis, as it significantly reduces the analysis time of foodstuffs and minimizes the use of chemicals.

This work focuses on the application of DART-MS methods for metabolomics-based food authentication. The discipline of metabolomics research examines small molecules that can be used to address different topics in food authentication. This approach is applied to determine the origin of three different foods: (i) strawberries as a water-rich food, (ii) grain maize as a carbohydrate-rich food and (iii) hazelnuts as a fat-rich food. This selection offers the opportunity to comprehensively analyze the performance of DART-MS systems.

The aim of the study is to develop a workflow for food authentication using DART-MS and to assess its efficacy through comprehensive testing using diverse matrices. This includes method optimization, analysis (strawberries, grain maize, hazelnuts) and development of an open-source workflow for data analysis.

The results demonstrate the potential of DART-MS-based analysis in determining the origin of a wide range of foods. The strawberry samples exhibit a classification accuracy of 63,8%, while grain maize demonstrates 84,4% and hazelnuts attain 93,2% accuracy. The observed variations in accuracy highlight the potential of DART-MS for the differentiation of the origin of carbohydrate-rich and fat-rich foods. The analysis of strawberries, which exhibits a classification accuracy of 63,8%, demonstrates the challenges associated with the analysis of water-rich foods. Utilizing DART-MS/MS experiments and fragmentation patterns previously elucidated by measuring standards, significant markers are identified across all datasets. The identification of di- and triacylglycerides as origin-relevant compounds, along with the determination of tocopherols as markers for the origin of hazelnuts, proves the potential of DART-MS-based analysis for the characterization of complex food samples.

In addition to DART-MS-based analysis, the developed workflow is compared with another rapid method: Fourier transform near-infrared spectroscopy (FT-NIR). For this purpose, the strawberry samples are analyzed and classified using FT-NIR and are compared with the results from the DART-MS-based workflow. The comparison yields the conclusion that FT-NIR analysis exhibits advantages in terms of sample preparation and handling, while the DART-MS approach offers significant benefits in terms of targeted marker identification and the sensitivity of the method.

Finally, the limitations of DART-MS analysis are discussed and the requirements for transfer to routine analysis are addressed. This work contributes to the establishment of DART ion sources in the field of food authentication on the way to faster, cost-efficient and more environmentally friendly analytics.

3 Einleitung

Umweltfreundlichkeit, Zeitersparnis und Kostenreduzierung sind wichtige Kriterien für die Entwicklung und Etablierung neuer analytischer Verfahren. Daher kommt der Etablierung von neuen geeigneten Arbeitsabläufen und Analysentechniken eine große Bedeutung zu. Dies ist insbesondere wichtig, da es in den letzten Jahren eine stetige Weiterentwicklung der etablierten Verfahren gab, aber keine durchgreifenden Neuerungen stattfanden. So müssen für die Zukunft neue Wege beschritten werden, um den ökologischen Fußabdruck der Analytik zu reduzieren und auch schnellere Antworten zu generieren. Eine bisher wenig etablierte Analysetechnik ist die Verwendung von DART-MS-Kopplungen. Diese Kopplung ermöglicht es Proben in flüssiger, fester oder gasförmiger Form unter Atmosphärenbedingungen zu ionisieren und anschließend massenspektrometrisch zu vermessen. Die auf DART-MS-basierende Analytik zeigt große Vorteile, da ohne chromatographische Trennung, ein hoher Probendurchsatz zu geringen Kosten und ganz ohne den Einsatz von Chemikalien möglich ist (Schmauder et al. 2024). Somit bietet DART-MS-basierte Analytik die Chance, eine große Veränderung der chemischen Analytik herbeizuführen. Jedoch wurden bisher die Möglichkeiten dieser Ionisationstechnik noch nicht mit umfangreichen Datensätzen auf ihre Leistungsfähigkeit getestet.

DART-Ionenquellen erfassen primär kleine Moleküle und eignen sich hervorragend für den Bereich der *Metabolomics*-basierten Authentifizierung, da dieses Verfahren auf der Erfassung niedermolekularer Metabolite beruht. Eine solche *Metabolomics*-Untersuchung basiert auf der Bestimmung kleiner Moleküle unter 1500 Dalton (Drexler et al. 2011). Durch die Erfassung möglichst vieler Moleküle des Pflanzenmetaboloms ermöglicht das Verfahren, die Anpassungsreaktion der Pflanze auf verschiedene Umwelteinflüsse zu untersuchen, welche sich in unterschiedlichen Metaboliten abbilden. Diese Umwelteinflüsse können beispielsweise durch unterschiedliche Standorte hervorgerufen werden, die sich in Bezug auf Wetter, Niederschlag und Klimabedingungen unterscheiden. Die Metabolomanalyse ist eine Möglichkeit zur Bestimmung der Herkunft und Authentifizierung verschiedener Lebensmittel. Somit stellt diese ein ideales Einsatzgebiet für DART-MS-Kopplungen dar.

Die Sicherung der Authentizität von Lebensmitteln ist ein wachsender Einsatzbereich für MS-basierte Methoden, da die Lebensmittelfälschung in den letzten Jahren vermehrt zu großen wirtschaftlichen Schäden geführt hat. Zur Lebensmittelfälschung, auch *Food Fraud* genannt, zählen alle Veränderungen an Lebensmitteln, durch welche ein wirtschaftlicher Gewinn erzielt wird, wozu beispielsweise die Verfälschung der Herkunft, Sorte oder des Anbaus zählt. Durch Lebensmittelbetrug wird ein jährlicher Schaden von 10–15 Milliarden Dollar verursacht (Johnson 2014). Diese hohe Schadenssumme unterstreicht die Notwendigkeit, neue Strategien zur Verhinderung von Lebensmittelbetrug zu etablieren. Bisher wird die

Authentizität von Lebensmitteln meist über Lieferdokumente sichergestellt, welche aber selbst anfällig für Fälschungen sind und so ist es wichtig die Authentizität über das vorliegende Lebensmittel feststellen zu können (Manning 2016). Beispielsweise kann die Herkunft durch Veränderung der Länderkennzeichnung einfach verfälscht werden und es können höhere Preise für eine mit höherer Qualität assoziierte Herkunft verlangt werden. Häufig erzielen heimische Produkte oder Produkte eines speziellen oder bekannten Anbaugebietes einen höheren Erlös und sind somit anfällig für Fälschung (Creydt et al. 2018b). Des Weiteren können durch die Verschleierung der Herkunft Sanktionen gegen Produkte aus bestimmten Ländern umgangen werden. Beispielsweise sind Lebensmittel bestimmter Herkünfte häufiger mit Kontaminanten, wie Aflatoxinen belastet, weshalb sie auch öfter kontrolliert werden (Corain 2023; Khodaei et al. 2021). Um eine transparente Herkunftskennzeichnung zu ermöglichen, sind für einige Produkte, wie beispielsweise Obst, die Herkunftsangabe verpflichtend und die Kennzeichnung wurde ab dem Jahr 2025 auf geschnittenes Obst ausgeweitet, um den Verzehr von lokalen Produkten zu fördern (Verordnung (EG) 1308/2013; Delegierte Verordnung (EU) 2023/2429).

Ein Lebensmittel mit besonderer Bedeutung der Herkunft ist in Deutschland die Erdbeere, welche zu den meistverzehrten Obstsorten in Deutschland zählt (Bundesanstalt für Landwirtschaft und Ernährung 2024). Durch die Relevanz für den lokalen Verbraucher werden Erdbeeren in dieser Arbeit als Beispiel für ein wasserreiches Lebensmittel ausgewählt und näher betrachtet.

Neben der Länderkennzeichnung können Herkunftsbezeichnungen zusätzlich durch geschützte geographische Angaben (g.g.A.) der EU ergänzt werden, was beispielsweise bei Haselnüssen aus verschiedenen Regionen in Italien der Fall ist (Verordnung (EG) Nr. 510/2006). Da Haselnüsse besonderen Schutz genießen, werden diese als Beispiel eines fettreichen Lebensmittels ausgewählt. Zur Komplettierung wird Körnermais als kohlenhydratreiches Lebensmittel herangezogen, da er weltweit die Nutzpflanze mit der größten Erntemenge darstellt und global angebaut sowie gehandelt wird (Erenstein et al. 2022).

Aufgrund der verschiedenen Matrices bieten diesen drei Lebensmittel eine hervorragende Basis, um einen Leistungsnachweis für DART-MS-basierte Analytik durchzuführen und diese als neuartigen analytischen Ansatz für den Herkunftsnachweis von Lebensmitteln zu etablieren. Somit soll diese Arbeit einen Beitrag dazu leisten die Möglichkeiten von DART-MS-basierter Analytik zu testen und die Grundlagen für eine Übertragung in die Routineanalytik zu schaffen. Durch ein grüneres, kostengünstigeres und effizienteres Verfahren soll die Lebensmittelauthentifizierung verbessert werden, um zukünftigen Lebensmittelbetrug zu verhindern.

4 Theoretische Grundlagen

4.1 Erdbeeren

Erdbeeren sind mit einem jährlichen Verzehr von 3,9 kg pro Kopf das beliebteste Saisonobst der Deutschen. In höheren Mengen verzehrt werden nur die ganzjährig verfügbaren Äpfel, Bananen und Tafeltrauben (Bundesanstalt für Landwirtschaft und Ernährung 2024). Neben 0,53–1,17% Proteinen, 0,33–0,48% Fett und 2,14–4,73% Kohlenhydraten weisen Erdbeeren einen großen Anteil an Wasser von 90,6–92,2% auf (Adak et al. 2017). Aufgrund des anhaltend hohen Konsums von Erdbeeren in Deutschland wurden im Jahr 2023 etwa 118.600 t frischer Erdbeeren importiert (Statistisches Bundesamt 2024). Hinzu kommen 130.000 t in Deutschland angebaute Erdbeeren, was eine starke Konkurrenz zwischen heimischen und importierten Früchten demonstriert (Statistisches Bundesamt 2025b). Dabei wird insbesondere in der Früh- und Spätsaison Ware nach Deutschland importiert, welche dann, aufgrund der besseren Kostenstruktur, zu günstigeren Preisen angeboten werden kann (Bundeszentrum für Ernährung 2023). Die wichtigsten Produktionsländer in Europa, mit Produktionsmengen von 79.000–329.000 t sind Spanien, Polen, Deutschland, Italien, Griechenland und die Niederlande (Eurostat 2025). Hinzu kommt Ägypten als wichtiger Exporteur für Erdbeeren in den Wintermonaten, was die Exportmenge im Jahr 2021 von circa 20.000 t Erdbeeren belegt (Amr Kadah 2021).

Erdbeeren unterliegen generell starken Preisschwankungen, welche einerseits durch die stark schwankenden Erntemengen und andererseits durch die unterschiedlichen Herkünfte begründet sind. Deutsche Erdbeeren wurden im Jahr 2024 für 5,32–8,11€ pro Kilogramm angeboten, während importierte Ware, beispielsweise aus Spanien für 2,04–3,00€ pro Kilo zu kaufen war (Spargel & Erdbeerprofi 2024; Blueberries - Magazine and Consulting 2024). Dieser deutliche Preisunterschied birgt ein großes Potential, durch Falschdeklaration der Herkunft einen zusätzlichen Gewinn zu erzielen. Dies ist einfach möglich, da die Herkunft der Ware meist durch Lieferdokumente von mehreren Zwischenhändlern garantiert wird und nur sehr begrenzt analytische Untersuchungen des Produktes stattfinden.

Erdbeeren, die nach Deutschland importiert werden, stammen zum Großteil aus Spanien, gefolgt von Griechenland, den Benelux-Ländern, Italien und Ägypten. Ein weiteres wichtiges Produktionsland stellt Polen dar, welches hauptsächlich für den Tiefkühlmarkt und den Eigenverzehr Erdbeeren produziert (Statistisches Bundesamt 2024; Bundeszentrum für Ernährung 2023).

4.1.1 Anbau

Erdbeeren zählen aufgrund des komplexen Anbauverfahrens, ähnlich wie Spargel, zu den Sonderkulturen. Grundsätzlich werden drei Anbaumethoden unterschieden: (i) Der Anbau im

Freiland findet auf dem Feld in Reihen statt und bei dieser Anbauform haben der Wind sowie das Wetter und Klima den größten Einfluss auf die Pflanze. (ii) Etwas geschützter ist der Anbau in Folientunneln, da hier Niederschlag keinen direkten Einfluss auf die Pflanze hat und die Temperaturen durch den Treibhauseffekt erhöht werden. (iii) Auch ein vom Boden des Feldes getrennter Anbau in Stellagen in Bodensubstrat kann im Gewächshaus stattfinden. Hierbei werden die Pflanzen über das Wasser mit allen wichtigen Nährstoffen versorgt und wachsen beispielsweise in Torf oder Kokosfasern (Akhatou et al. 2016). Durch die wärmeren Temperaturen in Folientunneln oder Gewächshäusern lässt sich die Ernteperiode deutlich verlängern (Hernández-Martínez et al. 2023). So werden mittlerweile in Deutschland jährlich 38.000 t Erdbeeren unter Schutzabdeckungen und in Gewächshäusern geerntet, während der Freilandanbau mit 93.700 t in etwa doppelt so viel zur Erntemenge beiträgt (Statistisches Bundesamt 2025b).

Gemeinsam haben die Anbaumethoden, dass meist mit Setzlingen gearbeitet wird, die vorher in Torf angezogen wurden und dann an den letztlichen Anbauort verbracht werden. Folglich werden die Setzlinge häufig über weite Strecken transportiert. Allein Ägypten exportierte 2021 etwa 4 Millionen Setzlinge, was die Globalisierung des Erdbeermarktes verdeutlicht (Amr Kadah 2021).

Im Erdbeeranbau werden viele unterschiedliche Varietäten der Gartenerdbeere (*Fragaria* × *ananassa*) eingesetzt. Die Auswahl erfolgt nach Erntezeitpunkt, Lagerstabilität, Anbaumethode, Kundenvorlieben und auch nach regionalen Gesichtspunkten. So sind verschiedene Sorten unterschiedlich anfällig für Hitze- und Trockenstress und manche Sorten sind für Transport sowie Lagerung besser geeignet (Nezhadahmadi et al. 2015; Cervantes et al. 2020). Aus den verschiedenen Anbaumethoden und den vielen verfügbaren Varietäten ergibt sich für die Erdbeere ein komplexer Anbau mit unterschiedlichsten Variationsmöglichkeiten.

4.2 Körnermais

Körnermais (*Zea mays* L.) ist mit einer Erntemenge von circa einer Milliarde Tonnen pro Jahr, noch vor Reis und Weizen, die am häufigsten angebaute Pflanze der Welt. Diese dient als wichtiger Baustein der Ernährungssicherheit entweder zum direkten Verzehr oder als Futtermais für die Produktion von tierischen Produkten. Durch den hohen Kohlenhydratgehalt von circa 72% eignet sich Körnermais besonders als Futtermittel sowie zur Stärkegewinnung und zählt folglich zu den kohlenhydratreichen Lebensmitteln. Weiterhin besteht Körnermais aus 9% Proteinen und zu 5% Fett. (Rouf Shah et al. 2016).

Die Hauptanbauländer von Körnermais sind die USA, China, Brasilien, EU-Länder wie Frankreich und Deutschland sowie die Ukraine. Diese Länder machen den Großteil der weltweiten Ernte aus und versorgen durch den Export von Körnermais weitere Länder, was

durch die gute Transportfähigkeit von Körnermais begünstigt wird (United States Department of Agriculture 2025).

4.3 Haselnüsse

Haselnüsse sind hochpreisige Nüsse, welche sowohl direkt verzehrt werden als auch einen wichtigen Rohstoff für eine große Bandbreite an Produkten der Süßwarenindustrie darstellen. Die Kerne der Haselnuss enthalten 50–73% Fett, 10–22% Kohlenhydrate und 10–24% Proteine, womit Haselnüsse zu den fettreichen Lebensmitteln zählen (Wani et al. 2020).

Die europäische Haselnuss (*Corylus avellana* L.) wird hauptsächlich in den Ländern Türkei, Italien, Aserbaidschan, Georgien und Frankreich angebaut. Mit 650.000 t ist die Türkei weltweit, vor Italien mit 103.000 t, das Land mit der größten Erntemenge. Aufgrund der großen Erntemengen der Türkei ist die Erntemenge dieses Landes maßgeblich an der Bildung des Weltmarktpreises beteiligt (Gönenc et al. 2006). So waren 2014 mit 710.000 t gegenüber 2022 mit 1.195.000 t die weltweiten Erntemengen besonders niedrig, was zu einem Preisanstieg von circa 50% für Importländer, wie beispielsweise Deutschland, führte (Statistisches Bundesamt 2025a; Food and Agriculture Organization of the United Nations 2024).

Ergänzend zu den jährlichen Preisfluktuationen erzielen Haselnüsse aus verschiedenen Ländern unterschiedliche Preise. So sind die italienischen Haselnüsse, „Nocciola Romana“, „Nocciola di Giffoni“ und „Nocciola del Piemonte“, als geschützte geographische Angabe (g.g.A.) durch die Europäische Union (EU) protektiert (Verordnung (EG) Nr. 510/2006). Dies führt zu höheren Preisen italienischer Haselnüsse, was es lukrativ macht, günstigere Ware mit teurer Ware zu verfälschen (Sammarco et al. 2023b).

4.4 Lebensmittelschutz

Angesichts der wachsenden globalen Bevölkerung und der verstärkten Herausforderungen durch Umweltveränderungen gewinnt der Schutz der globalen Ernährungssicherheit zunehmend an Bedeutung. Hierbei können verschiedene Teilbereiche betrachtet werden, welche im Folgenden erläutert werden.

4.4.1 Lebensmittelsicherheit

Im Bereich der Lebensmittelsicherheit (*food safety*) werden alle unbeabsichtigten Kontaminationen sowie von der Erzeugung und Verarbeitung des Lebensmittels hervorgerufenen Rückständen zusammengefasst. Bei Kontaminationen werden beim Anbau, der Verarbeitung, der Lagerung oder des Transports von Lebensmitteln unabsichtlich für den Konsumenten schädliche Kontaminationen in das Lebensmittel induziert. Diese können chemischer, biologischer, anthropogener Natur sein oder aus der Umwelt ins Lebensmittel eingetragen werden (Fung et al. 2018). Ein Beispiel für eine Kontamination war der im Jahr 2011 geschehene Eintrag von enterohämorrhagische *Escherichia coli* (EHEC) in Sprossen.

Durch diese Kontamination während des Anbaus der Sprossen starben durch Infektion mit EHEC 53 Menschen in Europa und Afrika (Yeni et al. 2016). Diese immer wiederkehrenden Ereignisse, demonstrieren die Relevanz, neue Strategien zu entwickeln, die zu einer Verbesserung der Lebensmittelsicherheit beitragen. Dies kann zum einen durch die Vermeidung der Kontamination während der Herstellung geschehen, wofür die Kontaminanten bekannt sein müssen und regelmäßige Analysen während der Produktion durchgeführt werden. Zum anderen können Risikoanalysen und daraus entwickelte Präventionsstrategien einen Beitrag dazu leisten, die Lebensmittelversorgung vor unabsichtlichen Kontaminationen zu schützen (Fung et al. 2018).

Im Gegensatz dazu werden Rückstände, während der Produktion des Lebensmittels absichtlich eingesetzt und verbleiben während der Weiterverarbeitung in diesem. Ein Beispiel hierfür sind der Einsatz von Pflanzenschutzmitteln beim Erdbeeranbau (Looser et al. 2006). Diese werden eingesetzt, um die Früchte gegen Schädlinge oder Pilzkrankheiten zu schützen, gehen aber dabei auch ins Lebensmittel über. Zur Minimierung dieser Rückständen werden Beispielsweise ein möglichst geringer Einsatz von Pflanzenschutzmitteln oder die Wahl anderer Schutzmaßnahmen empfohlen (Kobi et al. 2018). Zudem ist ein Monitoring, regelmäßige Analysen und dadurch die Einhaltung der festgelegten Grenzwerte ein Mittel, um mögliche Risiken zu minimieren (Verordnung (EG) Nr. 396/2005; Safi et al. 2002).

4.4.2 Lebensmittelfälschung

Gegensätzlich zur Lebensmittelsicherheit handelt es sich bei Lebensmittelbetrug (*food fraud*) um die absichtliche Fälschung von Lebensmitteln. Fälschung oder Veränderung können dabei das vorsätzliche Ersetzen, Hinzufügen, Manipulieren oder Falschdeklarieren von Lebensmitteln oder Verpackungen sein, um einen wirtschaftlichen Gewinn zu erzielen (Spink und Moyer 2011). Somit ist das Ziel des Lebensmittelbetrugs üblicherweise die Realisierung eines Gewinns durch die absichtliche Verfälschung von Parametern der Lebensmittel. Dieser Gewinn kann beispielsweise durch die Streckung eines teuren Lebensmittels mit günstigeren Zutaten oder durch die Umdeklaration wertbestimmender Parameter erreicht werden. Beispiele für Streckungen wären das Hinzufügen von Olivenblättern oder Majoran zu Oregano, zu hohe Gehalte an Kakaoschale in Kakao oder das Verdünnen von Honig mittels günstigerem Zuckersirup (Creydt et al. 2023; Drees et al. 2023; Hansen et al. 2024). Das Umdeklarieren kann beispielsweise bei biologisch angebauten Produkten, Produkten mit spezifischer Angabe der Art bzw. Varietät, wie Thunfisch oder Trüffel oder durch die Fälschung der Herkunftsangabe, geschehen (Capuano et al. 2013; Michelini et al. 2007; Losso et al. 2023; Katerinopoulou et al. 2020). Ein aktuelles Beispiel hierfür ist die in den Jahren 2023–2024 von Europol und INTERPOL durchgeführte Operation OPSON XIII (EUROPOL 2024). Hierbei wurden in den EU-Staaten und verschiedenen Partnerländern eine Schwerpunktkontrolle im Bereich der Beerenfrüchte durchgeführt. Dabei wurden insbesondere Heidelbeeren

untersucht, da Waldheidelbeeren (*Vaccinium myrtillus* L.) häufig mit anderen Heidelbeerarten und der durch Sortenkreuzung entstandenen Kulturheidelbeere, vermischt werden. Im Ergebnis bestätigte sich die Vermutung: In der Hälfte der untersuchten Proben von Waldheidelbeeren, waren keine Waldheidelbeeren enthalten (BVL 2024). Bei diesen Streckungen oder verfälschten Angaben liegt der Schaden beim Verbraucher oder Abnehmer in der Wirtschaft, da eine höhere Qualität des Lebensmittels vorgetäuscht wird und weiterhin der Preis für die höhere Qualität gezahlt wird. Insbesondere die Verfälschung der Herkunft von Lebensmitteln kann eine große Herausforderung darstellen, da bei globalen Lieferketten die Authentizität der Lebensmittel meist über Lieferdokumente festgestellt wird. Jene Lieferdokumente sind jedoch einfach zu fälschen und so muss eine Überprüfung der Herkunft im besten Fall anhand der bereitgestellten Ware sichergestellt werden können (Manning 2016). Aufgrund der großen Herausforderungen im Bereich der Herkunftsauthentifizierung stellt dies eine optimale Fragestellung für die Erprobung neuer analytischer Techniken dar, weswegen sich diese Arbeit auf die Authentifizierung der Herkunft von Lebensmitteln fokussiert.

4.5 Lebensmittelauthentifizierung

Die Lebensmittelauthentifizierung befasst sich mit der Sicherstellung der Echtheit von Lebensmitteln. Zu diesem Zweck ist die Entwicklung innovativer Methoden erforderlich, um eine Authentizitätsprüfung idealerweise am Produkt selbst durchführen zu können. In den zu untersuchenden Lebensmitteln kommen verschiedene Inhaltsstoffgruppen wie beispielsweise DNA, RNA, Proteine, Elemente und Stoffwechselprodukte, sogenannte Metabolite, vor. Je nach Fragestellung der Authentifizierung können verschiedene Disziplinen bei der Erfassung dieser Gruppen zur Anwendung kommen.

4.5.1 Die *Omics*-Kaskade

Die *Omics*-Kaskade beschreibt den Zusammenhang zwischen den verschiedenen Bestandteilen des Informationsflusses im molekularen Aufbau einer Zelle. Zu den *Omics*-Technologien gehören vorrangig *Genomics*, *Transcriptomics*, *Proteomics* und *Metabolomics*. Hierbei erfolgt der Informationsfluss, wie in Abbildung 01 dargestellt, beginnend mit dem Genom einer Zelle.

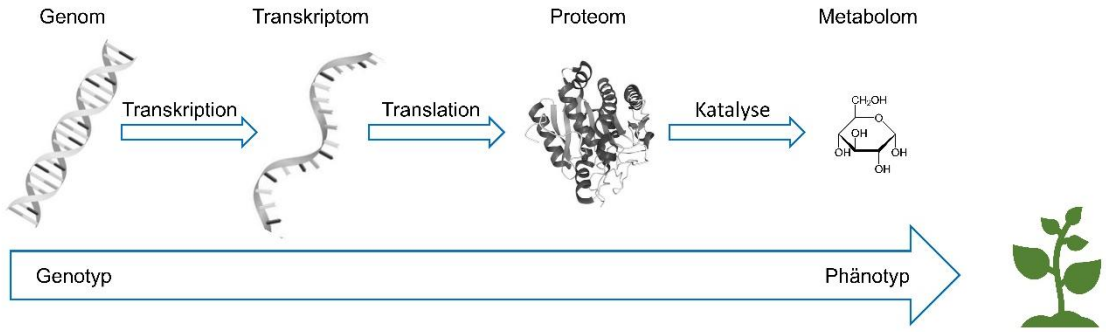


Abbildung 01: Schematische Darstellung der *Omics*-Kaskade zur Visualisierung des Informationsflusses in einem biologischen System vom Genom, über das Transkriptom und Proteom zum Metabolom.

Die *Genomics*-Analyse beschäftigt sich mit der Aufklärung der Struktur und der Zusammensetzung der DNA einer Zelle. Die Genexpression führt zur Transkription der DNA in mRNA und die Gesamtheit des Transkripts sowie die Untersuchung dessen wird in der Disziplin der *Transcriptomics* zusammengefasst. *Proteomics* erfasst die aus der Translation der mRNA resultierenden Proteine, ihre Zusammensetzung, ihren Aufbau und ihre Häufigkeit sowie der aus den Proteinen aufgebauten Enzymen. Die letzte Stufe des zellulären Regulationsprozesses sind die Metabolite, wobei es sich um niedermolekulare Stoffwechselprodukte (< 1500 Da) handelt (Pichersky und Lewinsohn 2011). Die Metabolite werden durch Enzyme gebildet, oder abgebaut und reflektieren damit die Reaktion des biologischen Systems auf veränderte Umwelteinflüsse. Des Weiteren können Metabolite und Proteine bei Pflanzen durch den Boden aufgenommen werden und so eine weitere Interaktion zwischen Pflanze und Umgebung aufzeigen (Reeve et al. 2008; Klampfl 2019). Analog kann es bei tierischen Lebensmittel zu eine Aufnahme durch die Nahrung kommen, welche dann selbst oder nach einer Metabolisierung Teil des Metaboloms bzw. Proteoms werden (Gilbert et al. 2008). Die Reaktion des biologischen Systems wird durch das Transkriptom und das Proteom an die Metaboliten weitergegeben und beeinflussen deren Zusammensetzung. Dabei stellt die DNA des Genoms die weitestgehend statische Komponente des Systems dar, da sie hochkonserviert ist, während entlang der Omics-Kaskade die Variabilität zunimmt. Schlussendlich sind Metabolite die schnelle Reflexion für veränderte Umweltbedingungen. Folglich eignen sich Metabolomanalysen, um Rückschlüsse auf Einflüsse während des Zellwachstums zu ziehen (Karahalil 2016).

Vom biologischen Informationsfluss der Zelle separiert ist die *Isotopolomics*, welche die Gesamtheit der Elemente betrachtet. Durch unterschiedliche Umwelteinflüsse, wie beispielsweise verschiedene Böden, variiert auch die Element- und Isotopenkonzentration in den Pflanzen. Zudem verteilen Pflanzen Elemente selektiv und steuern die Konzentration in verschiedenen Pflanzenbestandteilen (DalCorso et al. 2014). Somit können auch mittels der

unterschiedlichen Elementkonzentrationen, Rückschlüsse auf den Boden und damit den Anbauort gezogen werden (Shen et al. 2013).

4.5.2 Analytische Ansätze zur Herkunftsbestimmung

Die bereits etablierte und in der Routineanalytik angewendete Methode zur Herkunftsbestimmung ist die Analyse der leichten stabilen Isotope $^{2}H/^{1}H$, $^{13}C/^{12}C$, $^{15}N/^{14}N$, $^{18}O/^{16}O$ und $^{34}S/^{32}S$ (Kelly et al. 2005). Die Verteilung der Stabilisotopen $^{2}H/^{1}H$ und $^{18}O/^{16}O$ wird durch das unterschiedliche Gewicht der Isotopen und das Wetter beeinflusst (Bontempo et al. 2009). Über den Ozeanen verdunstet Wasser und regnet Stück für Stück beim Treffen auf Landmasse ab. Durch das unterschiedliche Gewicht der Isotopen regnen die schweren Isotope in Küstennähe ab, während die leichteren Isotope erst in Regionen ohne Küstennähe niederschlagen und sich so eine Verteilung in den Isotopenverhältnissen ergibt. Die Unterschiede in den Stabilisotpen $^{13}C/^{12}C$, $^{15}N/^{14}N$ hängen von der landwirtschaftlichen Praxis und bei Stickstoff mit dem Einsatz von synthetischen Düngern zusammen. Für die Stabilisotopen $^{34}S/^{32}S$ sind Effekte aus der vulkanischen Aktivität, dem Wetter und der Geologie der Böden verantwortlich (Camin et al. 2017). Im Ergebnis zeigen sich so verschiedene Stabilisotopenverteilungen, welche für die Herkunftsbestimmung genutzt werden können. Die gute Vergleichbarkeit der Ergebnisse, die vergleichsweise einfachen Messungen und die überschaubare Anzahl an Elementen (Wasserstoff, Sauerstoff, Stickstoff und Schwefel) haben dazu geführt, dass sich diese Methode in der Routineanalytik etabliert hat (Camin et al. 2017). Jedoch ist der Stabilisotopengradient auch eine große Schwäche, da alle Gebiete mit derselben Entfernung zum Meer in etwa die gleiche Isotopenverteilung für die Isotope $^{2}H/^{1}H$ und $^{18}O/^{16}O$ aufweisen. Daher kommt es zu Einschränkungen in der Detektion von Ländern, welche ähnliche geographische Lagen in Bezug zum Meer aufweisen (Creydt et al. 2018b).

Des Weiteren können neben den Unterschieden in den Stabilisotopenverteilungen einer Pflanze Differenzen in den einzelnen Elementen zur Herkunftsbestimmung genutzt werden. Dieses Verfahren beruht darauf möglichst viele Elemente in den Proben zu erfassen und aufgrund der unterschiedlichen Elementkonzentrationen Rückschlüsse auf die geographische Herkunft der Proben zu ziehen. Je nach Standort nehmen Pflanzen diese, hauptsächlich durch den Boden, aufgrund von Niederschlägen, vulkanischer Aktivität sowie industrieller Luftschadstoffe, verschiedene Elemente auf (Kravchenko et al. 2025; Caboi et al. 2021). So können dann aufgrund einzelner oder mehrerer Elemente Rückschlüsse auf die geographische Herkunft der Proben gezogen werden (Bertoldi et al. 2016). Ein großer Vorteil dieser Methode ist der geringe Einfluss der Varietäten und die direkte Übertragbarkeit in die Routineanalytik. Nachteilig kann die Unterscheidung von Regionen mit ähnlichem Elementprofil der Böden sein, wodurch es zu Fehlklassifikationen kommen kann (Richter et al. 2019).

In Ergänzung dazu stellen Metabolite Reaktionen der Zelle auf endogene und exogene Einflüsse dar und so können diese genutzt werden, um Rückschlüsse auf Standortfaktoren wie Klima, Wetter, den Boden, aber auch die Pflanzenvarietät zu ziehen. Die unterschiedlichen Standortfaktoren bewirken differenzierte Zellantworten, so bilden beispielsweise Pflanzen unter Stress eher weniger gesättigte Fettsäuren aus (Upchurch 2008). Ebenfalls können Korrelationen zwischen Schwermetallen und ungeradzahligen Fettsäuren gezogen werden, was den Einfluss von Bodenbestandteilen auf das Metabolom demonstriert (Lösel et al. 2024). So hat sich die Metabolomanalyse bereits im Bereich der Herkunftsbestimmung weit etabliert und es war möglich die Herkunft unterschiedlichster Lebensmittel zu authentifizieren (Cubero-Leon et al. 2014; Katerinopoulou et al. 2020).

4.5.2.1 Metabolomics-basierte Ansätze zur Herkunftsbestimmung

Das Metabolom einer Zelle kann über eine Vielzahl von analytischen Techniken erfasst werden, welche im Folgenden erläutert werden:

Im Bereich der **MS-Methoden** werden sowohl die Gaschromatographie gekoppelte Massenspektrometrie (GC-MS) als auch die Flüssigkeitschromatographie gekoppelte Massenspektrometrie (LC-MS) eingesetzt. Mittels GC-MS kann sowohl das flüchtige als auch nach einer Derivatisierung das nicht-flüchtige Metabolom untersucht werden (Sammarco et al. 2023a; Moros et al. 2017). Dagegen wird mittels LC-MS das nicht-flüchtige Metabolom untersucht, wobei eine Derivatisierung nicht erforderlich ist (Roca et al. 2021). Für die Erfassung eines möglichst großen Teils des Metaboloms bieten sich *non-targeted* Methoden an. Bei diesen nicht zielgerichteten Vorgehen werden meist Detektoren verwendet, welche gleichzeitig möglichst Metaboliten erfassen können (Wu et al. 2021). Durch die Erfassung eines möglichst großen Teils des Metaboloms können – bei einer ausreichend hohen Anzahl an Proben – Rückschlüsse zwischen bestimmten Metaboliten und der geographischen Herkunft einer Region gezogen werden. Vorteilhaft sind bei MS-basierten Methoden die hohe Anzahl an erfassten Metaboliten, die niedrige Nachweisgrenze und die Selektivität der Methode (Li et al. 2021). Ein großer Nachteil ist die lange Messzeit, welche i.W. durch die Chromatographie bestimmt wird und 15–60 Minuten in Anspruch nehmen kann. Weiter ist bei Übertragung einer Methode in die Routine, aus Kostengründen die Entwicklung von *targeted* Methoden erforderlich, was gemeinsam mit der hohen Messzeit, eine Hürde für die Anwendung von Metabolomanalysen in der Routineanalytik darstellt (Creydt und Fischer 2020; McMillen et al. 2023).

Eine Alternative zur MS-basierten Analytik ist die **Magnetresonanzspektroskopie (NMR)**. Bei dieser Methode kann durch magnetische Anregung, meist von Wasserstoff- oder Kohlenstoffatomen und der Messung der Zeit, bis ein Atom wieder im Ursprungszustand ist, Rückschlüsse auf die chemische Umgebung der Atome gezogen werden. Somit können für

Atome unterschiedliche Metabolite verschiedene chemische Verschiebungen zugeordnet werden. Die Messzeit für eine einfache Aufnahme eines ^{1}H-NMR Spektrums beträgt in etwa 15 Minuten, wobei komplexere Fragestellungen in längeren Messzeiten von mehreren Stunden resultieren können (Pandey et al. 2025). Somit haben NMR-Geräte in vielen Anwendungsfällen einen höheren Durchsatz als LC-MS-basierte Methoden (Hayashi et al. 2021). Dieser Zeitvorteil geht mit der gleichzeitigen Erfassung vieler Metabolite bei der NMR-Spektroskopie einher, was zu einer Überlagerung der Analyten in komplexen Mischungen führt und zu einer geringeren Sensitivität dieser Methode beiträgt. Zudem bietet sich die NMR-Spektroskopie eher für die Erfassung polarer Metabolite an und die Vermessung von unpolaren Metaboliten stellt eine zusätzliche Herausforderung dar. Lipide als Bestandteil des unpolaren Metaboloms weisen lange aliphatische Seitenketten auf und alle Atome dieser Seitenketten haben ähnliche chemische Verschiebungen im NMR-Spektrum, was eine Unterscheidung herausfordernd macht (Williams et al. 2025). Ein wichtiger Vorteil von NMR-Messungen ist die gute Robustheit der Methode und die Möglichkeit, geräteübergreifend Daten zu vergleichen, was den Aufbau von Datenbanken und den Routineeinsatz möglich macht (Powers et al. 2024).

4.5.2.1.1 Massenspektrometrische Methoden

Massenspektrometer (MS) sind weit verbreitete Analysengeräte, die in der Forschung und in der Wirtschaft für ein breites Analysenspektrum eingesetzt werden. In den Abschnitten 4.5.2.1.2, 4.5.2.1.3, 4.5.2.2.1 und 4.5.2.2.2 werden der Aufbau und die Funktionsweise erklärt.

4.5.2.1.2 Elektrospray-Ionisation

Zur Durchführung LC-MS-basierter Analytik werden für die Detektion der Metabolite nach der chromatographischen Trennung häufig MS-Geräte mit einer Elektrospray-Ionenquelle (engl. *electrospray ionization*, ESI) eingesetzt (Cubero-Leon et al. 2014; Roca et al. 2021). Diese wurde in den 1980er Jahren von Masamichi Yamashita und John Fenn entwickelt, wofür Fenn 2002 mit dem Nobelpreis ausgezeichnet wurde (Mann 2011; Yamashita und Fenn 1984). Das Prinzip der ESI beruht auf einem Spray aus Analytenlösung, welches durch das Austreten aus einer Nadel mit einem Durchmesser von wenigen Mikrometern, entsteht. Um ein stabiles Spray zu erzeugen, wird die Probe mit Hilfe eines inerten Gases, meistens Stickstoff, beim Austritt aus der Nadel zerstäubt. Dabei wird an die Nadel und einem Gegenpol, je nach gewünschtem Ionisationsmodus, eine positive bzw. negative Spannung von 2000–6000 V angelegt, wodurch eine Spannungsdifferenz aufgebaut wird. Es formiert sich am Ende der Nadel im Spannungsfeld der sogenannte Taylor-Konus, in welchem der Anteil am Lösemittel abnimmt und der Anteil an geladenen Molekülen in Richtung der Gegenelektrode zunimmt. Sobald mit sinkender Tröpfchengröße das Rayleigh-Limit überschritten wird, kommt es zur Coulomb-Explosion und die Tropfen werden zerstört. Die hierbei ablaufende Kaskade setzt die Explosionen in weiteren Tropfen fort, bis die Moleküle annähernd vollständig ionisiert in

der Gasphase vorliegen. Der genaue Mechanismus der Ionenbildung und Ionisierung ist hierbei nur zum Teil aufgeklärt (van Berkel und Kertesz 2007; Creydt 2019). Im Anschluss an die Ionisierung werden nicht-ionisierte Bestandteile durch weitere Gasströme abgetrennt und die entstandenen Ionen durch eine Transferkapillare in das MS eingezogen (Gross 2019; Awad et al. 2015).

Generell wird von der ESI als weiche Ionisationsmethode gesprochen, da sie sich gut zur Ionisation größerer intakter Moleküle eignet (Wang et al. 2018; Prabhu et al. 2023). Jedoch kommt es während des Ionisationsprozesses in geringerem Umfang zu Fragmentierungen von Molekülen, weswegen nicht alle Metabolite in intakter Form erfasst werden. Welcher Anteil an Molekülen während eines mittels ESI-Ionenquelle durchgeführten *Metabolomics*-Experiments nicht intakt Ionisiert wird, ist nicht vollständig bekannt (Da Silva et al. 2015; Giera et al. 2024). Während der Coelution von chromatographisch nicht getrennten Analyten oder bei Verfahren ohne chromatographische Trennung kann es, während der gleichzeitigen Ionisation mehrerer Moleküle, zu gegenseitigen Beeinflussung der Ionisation durch Ionensuppression kommen (Ghosh et al. 2012). Hierdurch kann die Zahl an erfassten Molekülen und die Nachweisgrenze vermindert werden, was sich beides negativ auf die Erfassung möglicher Markermoleküle auswirkt (Ghosson et al. 2021). Eine Minimierung der Ionensuppression kann durch eine selektive Extraktion der Analyten, eine Probenaufreinigung oder eine Optimierung der chromatographischen Trennung bzw. der Ionisationsparameter erfolgen (Furey et al. 2013).

4.5.2.1.3 Time-of-flight-Massenanalysatoren

Flugzeitmassenspektrometer (ToF, engl. *time-of-flight*) stellen eine Möglichkeit dar, mittels ESI erzeugte Ionen, zu detektieren. Das ToF-MS wurden 1946 von W. E. Stephens und M. M. Wolff erfunden (Wolff und Stephens 1953).

Beim ToF-MS werden die Ionen gepulst und über eine Flugstrecke anhand ihrer unterschiedlichen Masse-zu-Ladung-Verhältnisses (*m/z*) im feldfreien Raum aufgetrennt. Dies ist möglich, da die Ionen bei gleicher Ladung mit größeren Massen eine längere Flugzeit aufweisen. Die anschließende Detektion erfolgt mittels der auf dem Prinzip der Sekundärelektronenvervielfacher (SEV) beruhenden Mikrokanalplatten-Detektoren (engl. *microchannel plate detector*, MCP-Detektor) (Smith 1951; Ge 2024).

In der einfachsten Ausführung ist die Flugstrecke eine Gerade. Für das Erreichen einer längeren Flugstrecke und damit höheren Auflösungen werden Reflektrons eingesetzt. Ein Reflektron weist eine feldfreie Driftregion auf, welche durch Anlegen einer Spannung die Ionen abbremst und dann reflektiert (Mamyrin 1994). Dabei verweilen größere Ionen länger im Reflektron, da sie, aufgrund ihrer größeren kinetischen Energie, weiter eindringen. Der Auswurf aus dem Reflektion erfolgt im umgekehrten Einfallswinkel, sodass die Ionen im ToF eine parabelähnliche Flugstrecke zurücklegen (Gross 2019). Dies hat den Vorteil, dass die

Flugbahnen der ein- und austretenden Ionen sich nicht kreuzen. Durch die längere Flugstrecke und die längere Verweildauer der Ionen im Reflektron kann die Auflösung verbessert werden (Cornish und Cotter 1997; Guilhaus 1995; Ioanoviciu 1995).

Die Detektion der nach *m/z*-Verhältnissen getrennten Ionen erfolgt meist mittels MCP-Detektoren. Diese Detektoren bestehen aus einer Platte mit vielen Kanälen, welche mit einem Halbleitermaterial beschichtet sind. Beim Auftreffen eines Ions auf eine Kanalwand werden Sekundärelektronen aus dem Halbleitermaterial herausgeschlagen. Die Sekundärelektronen werden durch die umgekehrt angelegte Spannung in Richtung der gegenüberliegenden Kanalwand beschleunigt. Beim Auftreffen der Sekundärelektronen auf die gegenüberliegende Kanalwand werden wiederum mehr Sekundärelektronen aus der Elektrode emittiert. Diese Kaskade setzt sich über die weiteren Elektroden fort und die Spannung vervielfacht sich bei jeder weiteren Kollision mit der Kanalwand (Stanton et al. 1956; Ge 2024). Detektiert werden die aus der Mikrokanalplatte austretenden Sekundärelektronen in einem Vorverstärker, der die vorhandene Spannung nochmals erhöht und dann wird die Spannung in das eigentliche Signal umgewandelt (Gross 2019; Ge 2024)

4.5.2.2 Analytische Schnellmethoden zur Herkunftsbestimmung

Einige der in Abschnitt 4.5.2.1 beschriebenen Methoden zur Metabolomanalyse weisen den Nachteil einer zeitaufwändigen Vorbereitung und Messung auf. Daher rücken Schnellmethoden immer mehr in den Fokus der Forschung, welche in den folgenden Abschnitten erläutert werden.

4.5.2.2.1 Direct Analysis in Real Time-Ionenquellen

Direct Analysis in Real Time (DART) ist eine Ionisationstechnik bei der Ionen unter Atmosphärendruck und direkt in der Umgebungsatmosphäre entstehen (Cody et al. 2005)

Eine DART-Quelle besteht grundsätzlich aus drei unterschiedlichen Kammern. Zur Erzeugung der Ionen wird ein Trägergas, meist Helium oder Stickstoff, in die erste Kammer geleitet, welcher mit einer Anode und einer Kathode bestückt ist. Wenn nun eine hohe Spannung an die Kathode und die Anode angelegt wird, entsteht ein Plasma. Dieses enthält Ionen, Elektronen und angeregte Atome des Trägergases. Durch den Gasfluss gelangt das angeregte Trägergas in die nächste Kammer, welche mit einer zweiten Elektrode bestückt ist. Diese dient dazu, die entstandenen Ionen aus dem Trägergas zu filtern. Die dritte Kammer ist eine Heizkammer, die den Heliumstrom bei Bedarf erhitzen kann. Üblich sind Temperaturen zwischen 50 und 550°C (Cody et al. 2005; Saang'onyo und Smith 2012). Abschließend strömt das Gas vor dem Austritt aus dem DART durch ein Gitter bzw. dritte Elektrode, welche dazu dient, Ionen der nicht gewünschten Polarität zu entfernen. Dieses Gitter ist in einer Keramikspitze verbaut, welche es vor Berührungen mit dem Probenmaterial schützt und den Gasstrom in Richtung der Probe fokussiert (Cody et al. 2005).

Nach Austreten des angeregten Trägergases trifft dieses auf die Probe und löst eine Gasphasenionisation aus. Üblicherweise wird direkt hinter der Probe der Einlass für den mit dem DART gekoppelten Massenspektrometer montiert. Die Ionisation erfolgt im positiven Ionisationsmodus nicht direkt von angeregtem Helium auf die Probe, sondern über ein Trägermolekül. Über die Penning-Ionisation wird Stickstoff aus der Luft ionisiert. Dieser ionisiert Wasser, welches sich zu Wasserclustern formiert (Andrade et al. 2008; Kishimoto et al. 1998). Die Wassercluster sind die Moleküle, welche ihre Energie auf die Probe übertragen und diese ionisieren (Song et al. 2011). Bei DART werden meist $[M+H]^+$- oder $[M-H]^-$-Ionen von kleineren Molekülen beobachtet. In Anwesenheit von Ammonium können auch $[M+NH_4]^+$-Ammoniumaddukte entstehen, welche insbesondere bei Triacylglyceriden beobachtet werden können (Vaclavik et al. 2009). Analog zur positiven Ionisation wird im negativen Modus die Energie vom Helium auf Stickstoff übertragen, der via Penning Ionisation ionisiert wird (Cody et al., 2005). Der ionisierte Stickstoff überträgt ein Elektron auf Luftsauerstoff, welcher durch Reaktion mit dem Molekülion dieses ionisiert. Das Elektron wird an das Molekülion abgegeben, welches durch Abgabe eines Protons ein $[M-H]^-$-Ion formiert. In der Ionisation unter Atmosphärendruck wird durch die hohe Kollisionsrate mit der Luft die Energie von Ionen leicht wieder abgegeben, was zu einer geringen Fragmentierung von Molekülen bei DART führt (Gross 2014, 2019)

4.5.2.2.2 DART-qToF-MS

Die Kopplung aus einer DART-Ionenquelle und einem qToF stellt eine neue und leistungsfähige Kombination für massenspektrometrische Anwendungen dar (Gross 2014; Avula et al. 2015) Eine schematische Darstellung der Kopplung findet sich in Abbildung 02.

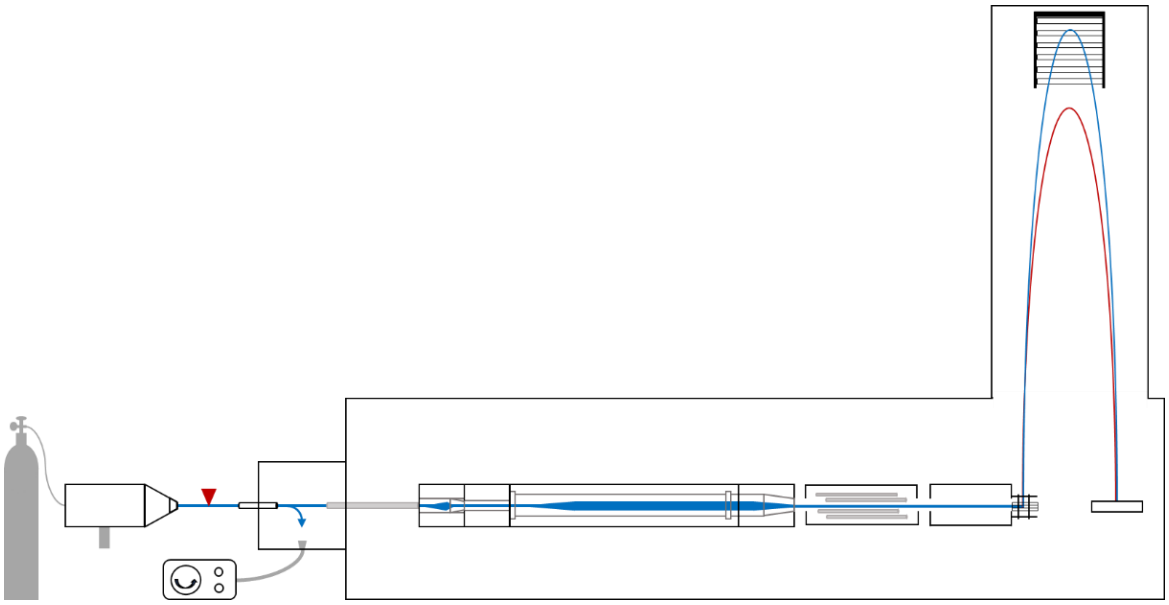


Abbildung 02: Schematische Darstellung einer DART-Ionenquelle gekoppelt mit einem ToF-Massenspektrometer.

Durch die Ionisierung mittels einer DART-Quelle kann eine sehr kurze Ionisationszeit pro Probe von wenigen Sekunden erzielt werden. Dies führt in Kombination mit einer Haltezeit zwischen den Proben im Sekundenbereich, zu einer sehr kurzen Analysendauer (Gross 2014; Novotná et al. 2012; Gross 2019). Für die geringe Analysendauer ist ein Massenspektrometer mit einer hohen Scanrate erforderlich. Die Kopplung mit einem hochauflösenden qToF bietet den Vorteil einer hohen Scanrate und einer hohen Massengenauigkeit. Die hohe Massengenauigkeit ist zudem notwendig, da die Analytik bei Direktionisationsmethoden ohne vorgeschaltete chromatographische Trennung erfolgt und dadurch insgesamt mehr Analyten gleichzeitig erfasst und unterschieden werden müssen. Zudem ist es allgemein für einen *non-targeted* Ansatz Voraussetzung einen MS mit hoher Massenauflösung zu verwenden, da nur so die unbekannten Analyten unterschieden und zugeordnet werden können (Fisher et al. 2022).

4.5.2.2.3 Signalidentifizierung mittels DART-qToF-MS

Die mittels DART-qToF-MS erfassten *m/z*-Verhältnisse können mittels Datenbanksuche den Metaboliten mit der entsprechenden Masse zugeordnet werden. Hierfür wird zunächst mittels der hochauflösenden Masse ein Vorschlag für die Summenformel errechnet oder direkt die hochauflösende Masse in die Datenbank eingegeben. Diese schlägt dann Metabolite vor, die zur errechneten Summenformel oder Masse keine oder eine geringe Abweichung aufweisen (Sud et al. 2007; Wishart et al. 2022).

Jedoch ist eine Zuordnung mittels Datenbanksuche fehlerbehaftet, weil für eine Masse viele isomere Moleküle in Betracht kommen können. Daher erfolgt die Zuordnung der Metabolite über eine Tandem-Massenspektrometrie (MS/MS) durch Fragmentierung der Ionen und Identifizierung über die erhaltenen MS/MS-Spektren. So wird beispielsweise bei Diacylglyceriden eine der beiden Esterbindungen gespalten und als Produkte lassen sich eine Fettsäure und das verbleibende Monoacylglycerid im Fragmentspektrum beobachten (Byrdwell und Neff 2002). Idealerweise wird das Fragmentierungsverhalten mit einem Standard verglichen, um sicherzustellen, dass es sich bei dem beobachteten Fragmentspektrum um den richtigen Metaboliten und die richtige Masse handelt. Dies ist insbesondere mit den in dieser Arbeit verwendeten DART-qToF-Plattformen ratsam, da es derzeit nur eine limitierte Anzahl an Studien zur Identifizierung von Metaboliten mittels DART-MS/MS Spektren gibt (Wang und Liu 2024).

4.5.2.2.4 Fourier-Transformations-Nahinfrarotspektroskopische Methoden

Eine bereits etablierte Methode zur schnellen Lebensmittelauthentifizierung stellt die Fourier-Transformations-Nahinfrarotspektroskopie (FT-NIR) dar (Wang et al. 2017; Arndt et al. 2021). Bei dieser Methode wird das Lebensmittel mit elektromagnetischer Strahlung unterschiedlicher Wellenlängen im Nahinfrarotbereich (4000–12500 cm^{-1}) bestrahlt. Hierbei kommt es zur

Anregung der Moleküle und die Strahlung wird in Abhängigkeit von der chemischen Struktur der Probe gestreut, absorbiert bzw. emittiert. Diese Streuung, Absorption bzw. Emission kann dann durch einen Detektor erfasst werden und die Frequenzinformationen werden durch Foutier-Transformation in ein NIR-Spektrum umgewandelt. Unterschiedliche funktionelle Gruppen können durch Valenz- und Deformationsschwingungen in verschiedenen Signalen des NIR-Spektrums resultieren. Durch die Messung dieser NIR-Strahlung kann von einer Probe ein chemischer Fingerabdruck des Metaboloms aufgenommen werden. Dies ist möglich, da alle Makrokomponenten eines Lebensmittels, wie Kohlenhydrate, Fette und Proteine, über ihre spezifischen funktionellen Gruppen die Reflexion, Absorption und Emission beeinflussen (Weidlein et al. 1982). Bei festen Proben kommt meist die Messung der diffusen Reflexion zur Anwendung. Hierbei wird die Strahlung erfasst, welche von der Probe ungerichtet zurückgeworfen wird. Dagegen wird bei flüssigen Proben häufig die Transmissionsmessung angewendet, da durch den hohen Flüssigkeitsanteil der Probe nur eine geringe Reflexion der Strahlung gegeben ist (Wang et al. 2021). Von Vorteil ist bei der NIR-Spektroskopie die schnelle, kostengünstige und umweltfreundliche Messung, welche jedoch von einer hohen Nachweisgrenze begleitet wird. Deswegen eignet sich die NIR-Spektroskopie gut als Schnellmethode, beispielsweise für die Wareneingangskontrolle und die Analyse großer Unterschiede, ist jedoch bei der Aufklärung komplexer Probleme aufgrund der geringen Auflösung und der fehlenden Selektivität limitiert (Arndt et al. 2020b; Nagy et al. 2022).

Eine Neuerung auf dem Gebiet der NIR-Spektroskopie stellt der Einsatz von mobilen NIR-Spektrometern dar. Normalerweise werden NIR-Geräte in Laboren eingesetzt und sind an die Klimatisierung und Stromversorgung gebunden. Dagegen ist es möglich mittels miniaturisierter NIR-Spektrometer mobile Messungen unabhängig von Laborinfrastruktur durchzuführen. Dies stellt einen entscheidenden Vorteil gegenüber komplizierteren analytischen Verfahren dar, da es möglich ist mobile NIR-Spektroskopie im Feld einzusetzen. Dagegen erfordern das geringere Gewicht und die reduzierte Größe der Geräte Kompromisse bei der Auflösung und Leistungsfähigkeit der Geräte, weswegen mobile NIR-Spektrometer bisher nur für simplere Fragestellungen eingesetzt wurden (Rukundo et al. 2021; Müller-Maatsch und van Ruth 2021).

4.6 Multivariate Statistische Verfahren

Die Anwendung von massenspektrometrischen Verfahren auf komplexe Matrices führt zu einer großen Anzahl an Daten. Daher erlaubt eine Anwendung herkömmlicher univariater Statistik für hochdimensionale Datensätze mit mehreren hundert bis über tausend Variablen nur eine begrenzte Betrachtung der Zusammenhänge über alle Variablen hinweg. Aufgrund dessen werden multivariate statistische Verfahren angewendet, um eine umfassende Auswertung der Zusammenhänge zwischen den verschiedenen Variablen zu ermöglichen.

Insbesondere die Möglichkeit der Untersuchung unterschiedlicher Probengruppen und die Analyse der Zusammenhänge und Unterschiede dieser Gruppen machen multivariate Verfahren ideal für die Auswertung von MS-Datensätzen (Worley und Powers 2013).

4.6.1 Datenprozessierung von DART-MS Daten

Die Datenprozessierung stellt die Verbindung zwischen Rohdaten und multivariater Auswertung dar. Zunächst werden die Rohdaten, welche herstellerabhängige Dateiformate aufweisen, in universelle Dateiformate umgewandelt, damit diese auch in nicht-kommerzielle Software eingelesen werden kann.

Im darauffolgenden Schritt wird eine Massendetektion durchgeführt. Dabei werden die Peaks zentriert, und aufgrund einer Mindesthöhe alle Peaks, die sich nicht eindeutig vom Rauschen abgrenzen, exkludiert. Als Ergebnis wird sich eine Liste erhalten, die für jeden Scan einer Messung die detektierten *m/z*-Verhältnisse und die korrespondierenden Signalintensitäten aufweist. Anschließend werden die resultierenden extrahierten Ionenchromatogramme (EIC) aller detektierten *m/z*-Verhältnisse berechnet und diese in einer *feature*-Liste dokumentiert (Li et al. 2018). Bei einem *feature* handelt es sich um ein Signal eines bestimmten *m/z*-Verhältnisses (Labory et al. 2024).

Da sich in den Daten Isotope derselben Verbindungen befinden können, werden mittels eines Isotopenfilters die ^{13}C-Isotope zusammengefasst, sodass nur der Hauptpeak in der *feature*-Liste verbleibt. Anschließend werden die Daten der unterschiedlichen Proben, beim sogenannten *Alignment*, zusammengeführt. Durch die Verwendung von hochauflösender MS können die *m/z*-Verhältnisse in den unterschiedlichen Proben kleine Abweichungen aufweisen. So werden alle Proben eines Datensatzes beim Massen-*Alignment* zusammengeführt und die gleichen *m/z*-Verhältnisse aller Proben gruppiert. Dabei ist die Wahl der erlaubten Abweichung von besonderer Wichtigkeit, da zu kleine Abweichungen die Schwankungen von qToF-MS-Geräten nicht ausgleichen, während große Abweichungen zur Gruppierung von unterschiedlichen *m/z* Verhältnissen führen können. Bei Chromatographie-basierten Methoden wird zeitgleich noch ein Peak-*Alignment* durchgeführt. Dadurch werden Peaks, die durch Verschiebungen in der Retentionszeit nicht zur selben Zeit retardieren, auf eine gemeinsame Retentionszeit zusammenführt. Veränderungen der Retentionszeiten während einer LC-MS-Messung werden, beispielsweise durch die Abnutzung der LC-Säule verursacht, und können durch das Peak-*Alignment* behoben werden (Podwojski et al. 2009).

Die resultierende Datentabelle kann nun durch die Anwendung von Filtern reduziert werden. So ist es möglich, nur *features* in einem Bestimmten *m/z*-Bereich zu untersuchen oder eine Mindestzahl an Proben vorzugeben, in denen das *feature* vorhanden sein muss. Beispielsweise können nur *features* ausgewählt werden, die in mindestens 20% der Proben vorhanden sind, womit sich die Anzahl der *features* reduziert. Eine Reduzierung der *features*

kann zu besseren Ergebnissen in der multivariaten Klassifizierung führen und bei Unterschiedsbetrachtungen muss ein *feature* auch in einer gewissen Anzahl an Proben detektiert worden sein, um Unterschiede feststellen zu können (Ballabio et al. 2008; Do et al. 2018).

Im finalen Schritt werden fehlende Signalintensitäten (*missing values*) in der Tabelle aufgefüllt, denn wenn ein *feature* in der Tabelle nicht aufgeführt ist, ist es entweder nicht in der Probe vorhanden oder wurde bei der Datenprozessierung nicht richtig erfasst. Für das nachträgliche Auffüllen von Signalintensitäten möglicherweise nicht erfasster *features* werden Funktionen wie das *gap-filling* oder *recursive feature finding* angewendet. Dabei werden durch einen Algorithmus die nach dem *Alignment* fehlenden Werte in der Tabelle mit den originalen Massenspektren abgeglichen und sollten in den Massenspektren weitere *features* detektiert worden sein, werden diese in die Tabelle übernommen. Dieses Vorgehen senkt die Anzahl an fehlenden Werten in der Tabelle. Weiterhin fehlende Werte werden als Null betrachtet (Pluskal et al. 2010). Eine Alternative dazu ist das Ersetzen der Werte durch den kleinsten gemessenen Wert. Jedoch werden hierbei möglicherweise höhere Intensitäten eines *features* angenommen, welche nicht gemessen wurden und es kann zu einem Bias in den Daten kommen, weswegen in dieser Arbeit die fehlenden Werte mit „0“ ersetzt wurden (Hrydziuszko und Viant 2012; Do et al. 2018).

Vor der Durchführung einer multivariaten statistischen Analyse werden die Daten häufig normalisiert und zentriert. Eine Normalisierung der Daten hilft dabei, systematische Variationen zu reduzieren und die Leistungsfähigkeit darauffolgender statistischer Analysen zu verbessern. Bei dieser werden die Daten beispielsweise über die Summe der Daten oder den Median normalisiert, sodass die Daten anschließend normalverteilt sind. Eine Zentrierung der Daten hat zum Ziel, die Veränderungen zwischen den Datenpunkten, die durch Unterschiede in den gemessenen Detektorantworten vorkommen, zu reduzieren. So reduziert die Skalierung große Unterschiede auf einen relativen Wert, der nicht der absoluten Häufigkeit entspricht. Eine gebräuchliche Skalierungsmethode ist die Medianskalierung, bei der die Varianz der Metabolite um den auf null gesetzten Mittelwert skaliert wird. So werden alle Metabolite in der darauffolgenden Analyse unabhängig von ihrem realen Wert betrachtet und der Fokus kann auf die Variation zwischen den Proben gelegt werden (Di Guida et al. 2016; Boccard et al. 2010).

4.6.2 Datenprozessierung von FT-NIR Daten

Bei FT-NIR-basierten Studien wird häufig eine Datenvorbehandlung angewendet, um die durch die NIR-Messungen hervorgerufenen Schwankungen auszugleichen und die Vergleichbarkeit innerhalb der Daten zu gewährleisten. Verantwortlich für messspezifische Abweichungen sind Streulichteffekte der Proben und wellenlängenabhängige

Streulichteffekte. Zur Minimierung der Streulichteffekte werden Normalisierungsmethoden wie beispielsweise die multiplikative Streulichtkorrektur (engl. *multiplicative scatter correction,* MSC) angewendet (Arndt et al. 2020a). Das Verfahren normalisiert die Probenspektren über ein Referenzspektrum, das aus dem Mittelwert aller gemessenen Spektren gebildet wird und kann so probenspezifische Unterschiede minimieren (Isaksson und Næs 1988). Darüber hinaus sind die einzelnen Messpunkte eines spektroskopischen Verfahrens hochkorreliert, weswegen sich zur Minimierung der Variablenanzahl ein sogenanntes *Binning* eignet. Dabei werden mehrere benachbarte Signale in *Bins*, durch Mittelwertbildung, zusammengefasst und als ein Messwert betrachtet. Hierdurch kann das Signal-Rausch-Verhältnis verbessert und die Variablenanzahl deutlich reduziert werden. Durch die Reduzierung der Variablen kann die Effektivität in der anschließenden Auswertung mittels multivariater Methoden verbessert werden (Zhong et al. 2023). Bei der Auswertung der NIR-Daten wurde zudem der Mittelwert aus einer fünffachen Bestimmung von technischen Replikaten einer Probe gebildet, um Varianzen durch mehrfache Messungen auszugleichen.

4.6.3 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (engl. *principal component analysis*, PCA) ist ein ungerichtetes Trennverfahren zur Dimensionsreduktion, welches häufig bei *Metabolomics*-Studien angewendet wird, um Ausreißer zu identifizieren und erste Trends abzuleiten (Worley und Powers 2013). Mittels dieses Verfahrens können komplexe Daten auf wenige Hauptkomponenten reduziert werden, welche sich dann für eine zwei- oder dreidimensionale Darstellung eignen und eine einfache visuelle Analyse ermöglichen. Dafür werden Hauptkomponenten errechnet, wobei die erste Hauptkomponente als Linearkombination die größte Varianz der Proben erklärt. Die zweite Hauptkomponente steht orthogonal zur ersten und erklärt die zweigrößte Varianz usw. (Pearson 1901). Meist werden PCAs mit standardisierten und normalisierten Daten (siehe 4.6.1) durchgeführt, um zu vermeiden, dass einige sehr hohe Werte einen zu großen Einfluss auf das Ergebnis der PCA haben und alle Werte in das Ergebnis gleichermaßen eingehen. Zur Visualisierung und Auswertung einer PCA werden meist der *Score Plot* und der *Loading Plot* herangezogen. Der *Score Plot* trägt zwei der Hauptkomponenten in der x- und y-Richtung gegeneinander auf und die Proben werden als Datenpunkte in dieser zweidimensionalen Grafik dargestellt. Für die Analyse der auf die Hauptkomponente vereinigten Datenpunkte wird der *Loading Plot* verwendet. Dieser ist eine zweidimensionale Darstellung, welche die *Loadings* auf den Hauptkomponenten visualisiert und so Rückschlüsse auf die Einflüsse der Metabolite auf die einzelnen Proben zulässt. Meist werden in mehreren *Scores*- und *Loading Plots* die ersten drei bis fünf Hauptkomponenten untersucht, da diese größte Varianz eines Datensatzes wiedergeben (Ringnér 2008).

4.6.4 Random Forest

Der *Random Forest*-Algorithmus zählt zu den gerichteten Verfahren des maschinellen Lernens und beruht auf der Berechnung einer Vielzahl von Entscheidungsbäumen. Für den Aufbau der verschiedenen Entscheidungsbäume wird ein Teil der Proben, üblicherweise zwei Drittel, randomisiert ausgewählt und mit diesen ein Entscheidungsbaum trainiert (Breiman 2001). Dabei wird der Entscheidungsbaum durch rekursives binäres Partitionieren erzeugt, in dem die Datenmenge durch eine Prädiktorvariable in zwei disjunkte Teilmengen unterteilt wird, welche eine Aufteilung in möglichst reine Gruppen gewährleistet. Die Prediktorvariable wird, durch Ausprobieren aller möglichen Variablen oder einer zufällig ausgewählten Teilmenge der Variablen, bestimmt. Nach der ersten Teilung wird das Vorgehen so lange wiederholt, bis die Proben nicht mehr teilbar sind oder ein Stoppkriterium, beispielsweise eine minimale Probenanzahl, erreicht wurde. Die Klassifikation einer unbekannten Probe findet statt, indem die Probe durch den Entscheidungsbaum geschickt wird und anhand der Zuordnung der Klasse eines Endpunktes klassifiziert wird (Li et al. 1984).

Für das Training eines *Random Forest*s wird eine zufällige Auswahl der Trainingsdaten verwendet, um mit diesen einen Entscheidungsbaum aufzubauen. Dieser Prozess wird 500–10.000-mal wiederholt, um eine große Anzahl an unterschiedlichen Entscheidungsbäumen zu trainieren. Anschließend werden die im jeweiligen Entscheidungsbaum nicht genutzten Daten als Testdaten für den Entscheidungsbaum genutzt und ihre, über alle Entscheidungsbäume gemittelten Richtigkeiten, zur Bewertung des *Random Forest* herangezogen. Die Mittelung der Ergebnisse aller einzelnen Testdatensätze über alle Entscheidungsbäume hinweg ergibt den *out-of-bag-error* (OOB-error). Dieser lässt eine gute Vorhersage der Klassifikationsrichtigkeit des gesamten *Random Forest* zu und stellt einen wichtigen Vorteil dar, da kein externer Testdatensatz zur Evaluierung des Algorithmus benötigt wird. Die Klassifikation von neuen Proben findet statt, indem die neuen Proben alle trainierten Entscheidungsbäume des *Random Forest* durchlaufen und die Probe der Klasse zugeordnet wird, für welche die Mehrheit der Entscheidungsbäume votiert haben (Breiman 2001).

4.6.5 Lineare Diskriminanzanalyse

Die Lineare Diskriminanzanalyse (LDA) ist ein überwachtes Klassifikationsverfahren. Das Verfahren beruht auf der Annahme, dass normalverteilte Daten vorliegen. Beim Training des Modells werden Linearkombinationen der Variablen betrachtet, um eine Dimensionsreduktion durchzuführen. Dabei wird die Varianz zwischen den Gruppen maximiert, während die Varianz innerhalb der Gruppen minimiert wird. Die Klassen der Trainingsdaten werden im multidimensionalen Raum als Hyperellipsoid angeordnet und eine Trennebene zwischen den Klassen definiert. Diese Ebene definiert die Räume der Klassen und so werden neue Proben entsprechend ihrer Lage bezüglich der Ebene klassifiziert (Bortz und Schuster 2011). Da die

LDA ein lineares Klassifikationsverfahren ist, eignet sie sich nur für lineare Zusammenhänge. Klassifikationen mit nicht-linearen Zusammenhängen können mittels LDA nicht klassifiziert werden (Zhao et al. 2024).

4.6.6 Validierungsstrategien

Überwachte Klassifikationsverfahren werden auf einem vorliegenden Datensatz trainiert, wobei es zur Überanpassung (engl. *Overfitting*) an diesen Datensatz kommen kann. Diese Überanpassung kann mit einer schlechteren Klassifikationsleistung neuer Proben einhergehen und sollte vermieden werden, um generalisierbare Aussagen ableiten zu können. Um eine Überanpassung auszuschließen, wird die Richtigkeit der überinterpretierten Verfahren mittels eines externen Testdatensatzes zur Validierung des Verfahrens überprüft. Im einfachsten Ansatz wird dafür der Datensatz in Test- und Trainingsdaten geteilt. Nach der erfolgten Modellberechnung mittels der Trainingsdaten werden die Proben der Testdaten im Modell klassifiziert und die Ergebnisse des Modells mit der wahren Klasse der Proben verglichen. Der prozentuale Anteil, der im Testdatensatz richtig Klassifizierten proben wird als Richtigkeit (engl. *accuracy*) bezeichnet und ist ein wichtiges Maß für die Güte des erstellten Modells.

Die Auswahl des Testdatensatzes erfolgt üblicherweise per Zufallsziehung, wobei es vorkommen kann, dass Proben aus manchen Klassen über- oder unterrepräsentiert sind. Zur Vermeidung einer zufälligen Zuordnung kann die Ziehung des Testdatensatzes als geschichtete Stichprobe erfolgen, um die Klassenverteilung in den Daten auch im Testdatensatz abzubilden. Die Zufallsziehung des Testdatensatzes weist den Nachteil auf, dass auch zufällig gut oder schlecht klassifizierende Proben in den Testdatensatz aufgenommen werden, und so kann die Richtigkeit des Modells unter- oder überschätzt werden. Dieses Problem kann mittels einer mehrfach durchgeführten Validierung, wie beispielsweise der Kreuzvalidierung (engl. *cross-validation*) gelöst werden, welche schematisch in Abbildung 03 dargestellt ist.

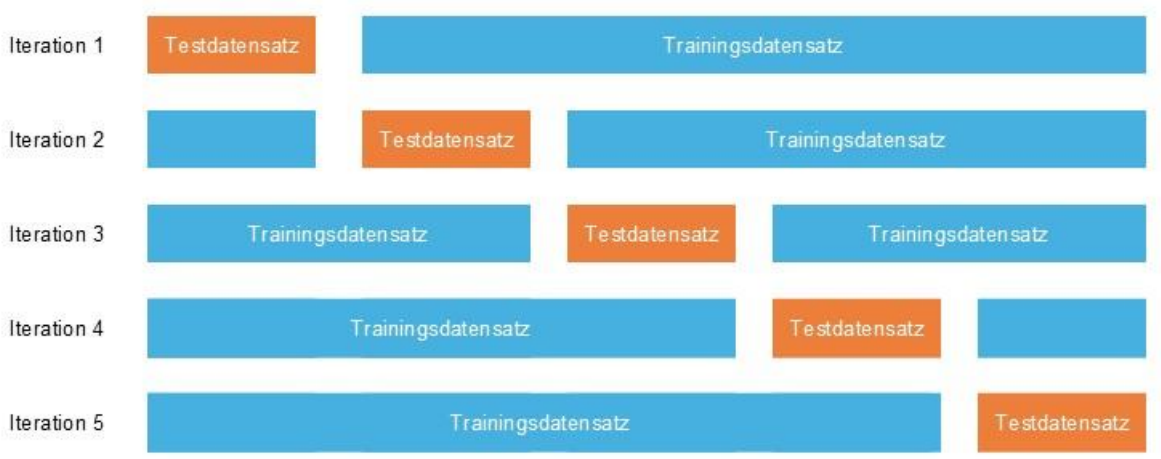


Abbildung 03: Darstellung einer fünffachen Kreuzvalidierung. Die Iterationen eins bis fünf zeigen die unterschiedliche Einteilung des Datensatzes in Test- und Trainingsdaten.

Bei einer fünffachen Kreuzvalidierung wird der Datensatz in fünf gleich große Teile geteilt und ein erstes Modell wird mit vier der fünf Teile trainiert. Der fünfte Teil wird dann zur Evaluation als Testdatensatz verwendet. Das Vorgehen wird fünfmal wiederholt, bis jeder der fünf Teilmengen einmal, als Testdatensatz verwendet wurde. Im Ergebnis werden in fünf Iterationen fünf Modelle trainiert und die Gesamtrichtigkeit als Mittelwert aus den Richtigkeiten der fünf Testdatensätze gebildet. Dieses Vorgehen minimiert die Wahrscheinlichkeit für zufällig gute oder schlechte Testdatensätze deutlich (Rubingh et al. 2006). Darüber hinaus kann die Kreuzvalidierung mit einer anders gewählten fünffachen Unterteilung der Daten weitere hundert Male durchgeführt werden, um den Zufallsfehler gänzlich zu minimieren. Dieses Verfahren der sehr häufigen Wiederholung einer Validierung wird als Monte-Carlo-Verfahren bezeichnet wird (Manly 2018).

4.6.7 Variablenwichtigkeit und Variablenselektion

Für die Ermittlung der für eine Klassifizierung verantwortlichen Metabolite wird häufig eine Rangfolge der Variablen erstellt, um die wichtigsten Variablen eingehender zu untersuchen. Hierfür kann entweder die Variablenwichtigkeit eines Algorithmus des maschinellen Lernens herangezogen werden, welche neben der Klassifikation eine Wichtigkeit der Variablen für die Klassifizierung errechnet. Alternativ dazu kann eine Variablenselektion angewendet werden, in der ein Algorithmus gezielt die wichtigsten Variablen mittels maschinellen Lernens selektiert.

Die Variablenwichtigkeit kann beispielsweise mittels *Random Forest* bestimmt werden. Hierbei wird die Abnahme der Richtigkeiten bei der Klassifizierung einzelner Variablen durch Permutation ermittelt und als *mean decrease in accuracy* (MDA) in Prozent angegeben. Somit ergibt sich für wichtige Variablen eine hohe Abnahme der Klassifikationsrichtigkeiten und es kann hieraus eine Rangfolge der wichtigsten Variablen erstellt werden (Breiman 2001). Die Varianzanalyse (engl. *analysis of variance*, ANOVA) ist eine weitere Möglichkeit, die

Variablenwichtigkeit zu bestimmen. Zur Berechnung einer ANOVA wird die Varianz der unterschiedlichen *features* mittels eines F-Tests untersucht und über ein Signifikanzkriterium entschieden, welche Variablen signifikant unterschiedlich zwischen den Klassen sind. Mit den aus dem F-Test resultierenden entsprechenden p-Werten kann eine Rangfolge der wichtigsten Variablen erstellt werden (Kirpich et al. 2018).

Eine auf *Random Forest*-basierende Variablenselektion kann mittels des Boruta Algorithmus durchgeführt werden. Boruta vergleicht die Variablen mit sogenannten Schattenvariablen, d. h. zufällig permutierten Versionen ihrer selbst, um durch einen statistischen Test diejenigen auszuwählen, die wichtiger sind (Kursa und Rudnicki 2010). Dabei bietet Boruta den Vorteil kurzer Berechnungszeiten und eine geringe Fehlerrate, welche durch etliche Anwendungsbeispiele im *Metabolomics*-Bereich unterstrichen wurden (Speiser et al. 2019; Lösel et al. 2023; Labory et al. 2024; Degenhardt et al. 2019).

5 Zielsetzung der Arbeit

Ziel der Arbeit war es einen Arbeitsablauf für die Lebensmittelauthentifizierung mittels DART-MS Analytik zu entwickeln und diesen an unterschiedlichen Matrices zu erproben. Hierfür sollte die Metabolomanalyse zur Herkunftsbestimmung von Lebensmitteln herangezogen werden, da sich kleine Metabolite gut für die Analyse mittels DART-Ionenquellen eignen. Zur Authentifizierung der Herkunft von Lebensmitteln sollten drei repräsentative Probensätze vermessen werden: (i) Erdbeeren als wasserreiches Lebensmittel, (ii) Körnermais als Beispiel für ein kohlenhydratreiches Lebensmittel und (iii) Haselnüsse als fettreiches Lebensmittel.

Zunächst sollten die Parameter der DART-Ionenquelle für eine *non-targeted* Analyse optimiert werden und ein *Open-Source Workflow* für die Datenauswertung entwickelt werden. Im Anschluss an die Datenauswertung sollte die Möglichkeit der Identifizierung von herkunftsrelevanten Metaboliten mittels DART-MS/MS Messungen erprobt werden, um diese zu identifizieren und einen Vergleich der Fragmentierungsmuster mit der etablierten ESI-MS Analytik zu ermöglichen.

Für die Untersuchung von Erdbeeren, sollten Proben aus den wichtigsten Erzeugerländern akquiriert und ein Protokoll zur Probenaufarbeitung etabliert werden. Ziel der Methodenentwicklung sollte eine Optimierung der Extraktion, der Ionisationsmodi und des Probenaufgabesystems für eine *non-targeted* Analyse anhand eines kleineren Datensatzes sein. Die gewonnen Erkenntnissen sollten auf einen größeren Datensatz übertragen werden, um die Herkunftsbestimmung mittels DART-MS-basierter Metabolomanalyse zu testen. Zudem sollten die wichtigsten Metabolite für die Klassifizierung ermittelt und diese mittels MS/MS Analytik fragmentiert und identifiziert werden, um eine vergleichende Betrachtung der Marker durchzuführen. Des Weiteren sollte ein Vergleich der DART-MS Analytik mit der NIR-Spektroskopie als bereits etablierte Methode der Schnellanalytik durchgeführt werden. Hierfür sollte ein Teil der Erdbeerproben ergänzend mittels NIR-Spektroskopie vermessen und klassifiziert werden.

Der final entwickelte Arbeitsablauf für die Herkunftsbestimmung mittels DART-MS-basierter Metabolomanalyse sollte mittels der weiteren Matrices, Körnermais und Haselnüssen, getestet werden. Analog zur Herkunftsbestimmung der Erdbeere sollte mittels MS/MS Analytik die Aufklärung relevanter Fragmentierungsmuster und die Identifizierung wichtiger Marker zur Herkunftsbestimmung erfolgen. Abschließend sollte ein Vergleich der drei Datensätze durchgeführt werden, um die Möglichkeiten und Limitationen der DART-MS Analytik bei der Authentifizierung verschiedener Lebensmittel zu beurteilen.

6 Kumulativer Teil der Dissertation

6.1 Herkunftsauthentifizierung von Erdbeeren (*Fragaria* × *ananassa*) mittels DART-MS.

Für die Etablierung eines Arbeitsablaufes zur Metabolomanalyse mittels DART-MS wurde die Erdbeere als wasserreiches Lebensmittel gewählt, da diese europaweit angebaut und gehandelt wird. Insgesamt wurden 207 Erdbeerproben aus 6 Ländern akquiriert und zunächst eine ausführliche Methodenoptimierung durchgeführt. Die Methodenoptimierung umfasste die Extraktion, den Ionisierungsprozess und das Probenaufgabesystem. Es zeigte sich, dass eine unpolare Extraktion mit positiver Ionisation und dem Dip-it-Probenaufgabensystem am besten geeignet für die Klassifikation von Erdbeeren nach ihrem Herkunftsland war.

Anschließend wurden alle 207 Proben mittels der optimierten Methode und der DART-MS Plattform vermessen. Die multivariate Datenanalyse zeigte eine begrenzte Auftrennung der Proben nach ihren Herkunftsländern in einem Sechs-Klassenmodell, da die Klassifikation eine reduzierte Richtigkeit von 63,8% ergab. Dagegen war eine Klassifikation der deutschen Proben gegenüber den ausländischen Proben mit einer Richtigkeit von 90,3% möglich.

Des Weiteren konnten 20 Lipide identifiziert werden, deren Signalintensitäten in den Proben eine Abhängigkeit von der geographischen Herkunft aufwiesen. Jedoch war die Aussagekraft der identifizierten Lipide limitiert, da es Überschneidungen insbesondere zwischen benachbarten Ländern gab. Die Überschneidungen in der Klassifikation konnten auf die große Variabilität des Erdbeeranbaus und auf die geographischen Ähnlichkeiten zwischen benachbarten Ländern gleicher Klimazonen zurückgeführt werden.

Hinweis zum Copyright:

This article is available under the Creative Commons CC-BY-NC-ND license and permits non-commercial use of the work as published, without adaptation or alteration provided the work is fully attributed.

URL: https://www.sciencedirect.com/science/article/pii/S2666831925000980

DOI: https://doi.org/10.1016/j.talo.2025.100496

Talanta Open 12 (2025) 100496

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open

DART-MS approaches for metabolomics-based authentication: Challenges in determining the geographical origin of strawberries (*Fragaria* × *ananassa*)

Felix Schmauder, Johannes Brockelt, Kim Brettschneider, Stephan Seifert, Marina Creydt, Markus Fischer *

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, Hamburg 20146, Germany

ARTICLE INFO

Keywords:
Authenticity analysis
Chemometrics
DART-MS
Mass spectrometry
Metabolomics
Strawberry

ABSTRACT

Metabolomics-based approaches are a widely used method for determining the origin of food. The aim of this study was to investigate the suitability of metabolomics analysis using direct analysis in real time mass spectrometry (DART-MS) to ascertain the origin of strawberries. In addition, optimization possibilities and limitations were identified by testing different extraction methods, two different sample delivery systems (Quick-Strip and Dip-it approaches), and running some of the measurements in both positive and negative ionization mode proved to be the most suitable. A total of 207 strawberry samples from six countries were measured and classified using random forest with this optimized approach. The classification accuracy was 63.8%, which shows certain limitations of the approach presented. However, another aim of this study was to protect German strawberries from adulteration with foreign strawberries, and this was achieved with a promising accuracy of 90.3%.

1. Introduction

In recent years, the discipline of metabolomics has been increasingly developed and many different approaches have been tested on a wide range of datasets [1]. Metabolomics experiments are based on the analysis of metabolites, which are small molecules up to a mass of 1,500 Da [2]. Plants contain more than 25,000 metabolites, only some of which can be detected using analytical techniques such as nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) [3,4]. The higher the resolution and sensitivity of an analytical technique, the more metabolites can be distinguished and also detected. For instance, with an liquid chromatography mass spectrometry (LC-MS) approach, it is possible to detect up to 3000 features in an asparagus sample [5]. The untargeted analysis of metabolites can be used, for example, to authenticate origin, determine variety or detect admixtures in foods. However, this approach has its limitations when classes that are too similar need to be distinguished [6]. Such limitations are thoroughly investigated in this publication to contribute to a more targeted use of metabolomics strategies in the future.

Strawberries represent an ideal example for metabolomic studies, as they are grown in a wide range of geographical regions and are a widely traded commodity [7]. More than 9 million tons of strawberries are harvested worldwide every year [8]. Germany, for example, imported 128,000 tons of fresh strawberries from Spain, Greece, Poland, the Netherlands, and Egypt in 2021 [9]. There are significant price differences between importing and exporting countries, making it lucrative to falsify the origin of strawberries and thus achieve higher financial profit [10,11].

To prevent such adulterations in the future, it is necessary to explore the advantages and limitations of new strategies based on metabolomics. Direct analysis in real time mass spectrometry (DART-MS) is an ionization technique that allows the investigation of food samples without chromatographic separation using mass spectrometric analyzers. The resulting advantages are a very short analysis time of about one minute and very low solvent consumption, making the method more environmentally friendly than LC-MS. Furthermore, DART-MS-based approaches are more cost-effective than LC-MS approaches due to the lower consumption of chemicals [12]. These advantages of DART technology and its potency for ionizing many small molecules make it an optimal choice for food authentication using a non-targeted metabolomics approach [13].

To date, DART-MS has been used in only a few studies for food

* Corresponding author.
E-mail address: markus.fischer@uni-hamburg.de (M. Fischer).

https://doi.org/10.1016/j.talo.2025.100496
Received 16 February 2025; Received in revised form 30 May 2025; Accepted 1 June 2025
Available online 5 June 2025
2666-8319/© 2025 The Authors. Published by Elsevier B.V. CCBYLICENSE This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

authenticity testing. The foods studied were chicken, beef, tomatoes, grain maize, peppers and hazelnuts [14–17]. In most cases, the extraction conditions were selected based on the food of interest or considering the results of previous studies. Similarly, the choice of positive or negative ionization was not determined experimentally in most studies [18–20]. Selecting extraction and ionization conditions based on theoretical considerations or literature research is possible, but better results can be achieved with a preliminary method evaluation. For this reason, a detailed method optimization was carried out in this study and the advantages and limitations of the various settings were taken into account. Furthermore, most of the published method optimization studies were performed by evaluating individual mass spectra or feature intensities. This procedure is well suited for approximate optimization, but does not allow the entire dataset to be evaluated as an analytical fingerprint [14, 21]. Therefore, the complete fingerprints were used to evaluate the method optimization based on different classification approaches.

Due to the high number of features that can be detected using non-targeted approaches, data analysis is carried out using multivariate methods. In this context, a distinction is generally made between two different approaches: supervised and unsupervised methods. The unsupervised method "Principal Component Analysis (PCA)" creates latent variables, called principal components, via linear combinations of the original variables. These principal components represent the main variances in the dataset and allow identification of samples with similar composition [22]. Supervised approaches focus on specific differences of predefined groups by incorporating class information, e.g. the geographical origin of the samples. A supervised method that offers some advantages for the classification of analytical data is random forest (RF). RF is a non-parametric, ensemble learning algorithm that is based on a large number of binary decision trees [23]. Since each of these decision trees is trained on a different subsample, called a bootstrap sample, an independent out-of-bag (OOB) error can be obtained by using the respective samples that were not used to train the individual tree. Without parameter optimization - which, unlike with support vector machines, for example, often does not result in major differences in performance anyway - the use of RF enables independent validation without the need for additional test dataset [24]. This is particularly advantageous when only a few samples have been analyzed overall or only a few samples from individual groups are available. RF can also be applied to analyze the importance of individual features based on importance measures such as the decrease of impurity or accuracy, which are used by selection approaches to separate important from unimportant features. Boruta compares the features with so-called shadow variables, randomly permuted versions of themselves, to select those that are more important through a statistical test [25].

For the first time, a comprehensive method development for a DART-MS approach was carried out, along with an in-depth investigation of its suitability for metabolomics fingerprinting studies. To develop an optimal workflow, two datasets containing 62 and 200 strawberry samples from different origins were measured in six different parameter combinations to evaluate the optimal conditions for i) the extraction procedure by using polar and nonpolar solvents, ii) the ionization mode and iii) the sample delivery system (Quick-Strip- and Dip-it approach, Fig. 1). In the Dip-it sample delivery system, a glass rod is dipped into the sample extract and automatically passed through the DART ion source, where the adherent molecules are desorbed and ionized. In the Quick-Strip system, the sample extract is dispensed onto a strip made of mesh. After a short drying period, the mesh is automatically moved into the DART ion source and the dried sample is desorbed and ionized from the mesh. The 200 samples used for this optimization step, many were German, making this dataset ideal for classifying German strawberries, but limiting the classification of individual countries. Therefore, the final method was applied to a dataset consisting of 207 strawberries from the main strawberry-producing countries. With at least 25 samples per country, it was possible to use a balanced dataset without unequally weighted classes. An unsupervised PCA was carried out to analyze the main variances in the data and RF was applied to classify the data according to their geographical origin. Furthermore, the Boruta algorithm was used to determine the most important features for classification and features were identified using MS/MS spectra. In addition, important signals were identified that provide a more detailed understanding of the non-polar metabolome of strawberries.

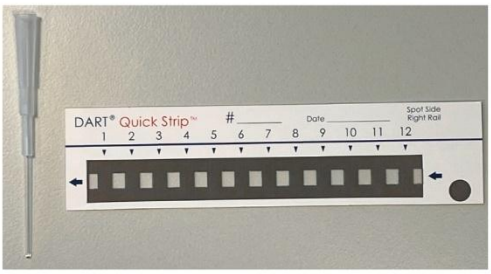


Fig. 1. Picture of the Dip-it sampler with glass rod (left) and the Quick-Strip wire card (right) with 12 pipetting places for sample extracts.

2. Materials and methods

2.1. Chemicals

Chloroform (HPLC grade) and methanol (LC-MS grade) were purchased from Carl Roth GmbH (Karlsruhe, Germany). Ultrapure water was obtained using a Millipore water purification system (Merck Direct-Q 3 UV-R, Merck KGaA, Darmstadt, Germany). Helium 5.0 was purchased from Westfalen Gas (Münster, Germany). Agilent Tuning mix was purchased from Agilent Technologies (Agilent Technologies, Santa Clara, CA, USA).

2.2. Strawberry samples

Samples from Germany, Poland and the Netherlands were acquired directly from the farms. Samples from Egypt, Greece and Spain were provided by importers and traders, and their authenticity was guaranteed by delivery documents. For each sample, 500 g of strawberries were used. A list of metadata, such as countries, regions, varieties and vintages of the samples, can be found in Tables S1-S3 of the electronic supplementary information.

Two datasets were created for method optimization: Dataset 1 included 62 samples from countries Germany (20), Egypt (12), Greece (15), and Spain (15) and was used to determine the most appropriate extraction method and ionization mode. Dataset 2 consisted of 200 samples and was used to compare the two different sample preparation systems, Quick-Strip and Dip-it. This was done by performing a classification with the aim of separating all German samples from all non-German samples. In addition, the number of features detected was taken into account. 105 of the 200 samples originated from Germany, and the other class consisted of samples from Egypt (12), Greece (21), Italy (3), Morocco (8), Spain (42) and the Netherlands (9).

The final model (dataset 3) was based on samples from the most market-relevant growing countries, for which a representative number of at least 25 samples was available. In order to avoid too much class imbalance, a maximum of 50 samples were used for each country. A total of 207 samples from 6 countries were used for this model, consisting of samples from Germany (51), the Netherlands (32), Poland (28), Greece (27), Spain (36) and Egypt (33).

2.3. Sample preparation and extraction

After the strawberries arrived at the research center, the stalks were removed and the fruits were quartered. Then, the quartered strawberries were flash-frozen in liquid nitrogen and stored at -20°C until further processing. Homogenization was performed with 300 g of strawberries and dry ice by means of a Grindomix GM 300 knife mill (Retsch, Haan, Germany). The pre-grinding was carried out with a backward-running knife to protect the blades (1 min, 1000 rpm). The main grinding was done at 4000 rpm for 2 min. Subsequently, the homogenized strawberries were transferred to pre-cooled 500 mL round-bottom flasks and freeze-dried for four days at 0.02 mbar using a freeze-dryer (Beta 2-8 LSCplus, Martin Christ, Osterode am Harz, Germany). On the third day, the samples were briefly removed and stirred using a glass rod. After freeze-drying, the samples were ground using a mortar (Haldenwanger, Waldkrainburg, Germany) and stored at -80°C until further analysis. For the analysis of the homogenized and freeze-dried samples, 50 ± 0.5 mg of each sample was weighed into a 2 mL safe-lock tube (Eppendorf, Hamburg, Germany).

Polar metabolites were extracted by pipetting 1.5 mL of methanol into the tubes. Two steel balls (Ø 3 mm, THS, Mühlacker, Baden-Württemberg, Germany) were also added. The sample material was then ground using a ball mill (BeadRuptor24, Omni International, Kennesaw, GA, USA, 3.1 m/s) for three minutes. Subsequently, the obtained suspensions were centrifuged for 20 minutes at 12700 rpm and at a temperature of 4°C (Eppendorf 5430 R, Hamburg, Germany). After centrifugation, 500 μL of the supernatants were transferred to new safe-lock tubes and centrifuged again for 7 minutes to ensure that clear extracts were obtained.

The non-polar compounds were extracted using a procedure developed by Bligh and Dyer and optimized by Creydt et al. [26,27]. After weighing the homogenized and freeze-dried samples in 2 mL safe-lock tubes, two steel balls and 750 μL of an ice-cold chloroform/methanol mixture (1:2; v/v) were added. The metabolites were extracted for one minute using the ball mill under the previously mentioned conditions. After ball milling, 500 μL of ultra-purified water and 250 μL of chloroform were added, and the extracts were again homogenized for two minutes using the ball mill under the conditions mentioned above. Centrifugation was also carried out as described above for 20 minutes, after which 500 μL of the lower chloroform phase was taken and transferred to new extraction tubes. The second centrifugation step was also performed under the conditions mentioned above for 7 minutes. The supernatants of the obtained extract were used directly for analysis.

2.4. Mass spectrometry measurements

For the measurements, a quadrupole time-of-flight (QToF) mass spectrometer (Agilent 6560 IM-QTOF-MS, Agilent Technologies, Santa Clara, CA, USA) equipped with a DART source (Ion Sense, Saugus, MA, USA) was used. The mass range of the qToF was set to m/z 100-1000 and the spectral rate was 10 Hz. Before the DART source was installed, the qToF analyzer was tuned and calibrated using the commonly used electrospray ionization (ESI) source (Dual JetStream, Agilent Technologies, Santa Clara, CA, USA). For this step, the small mass procedure (m/z 50-1700) with the Agilent ESI tuning mix was used. After completion of the tuning procedure, the ESI source was replaced by the DART source.

The DART source was connected to the mass spectrometer via a vapor interface with a ceramic capillary. A 2 mm gap was left between the ceramic capillary and the sample holder, as well as between the DART source and the sample holder. Furthermore, the vapor interface was connected to a vacuum pump (Vacuubrand, Wertheim, Germany) to generate a pre-vacuum in the vapor interface, which sucked the ions generated in ambient air into the ceramic capillary. Depending on the experimental setup, a linear rail with a Quick-Strip or Dip-it sample holder (Ion Sense, Saugus, MA, USA) was used.

The DART source temperature was set to 300°C for measurements in positive ionization mode and 250°C for measurements in negative ionization mode (hereafter, in combination with polar and non-polar extraction, referred to as P-pos, P-neg, NP-pos and NP-neg). The pulse duration for the measurements of the non-polar extracts was set to 5 seconds, while 7 seconds were chosen for the measurements of the polar extracts. These changes in the measurement time were made to ensure for complete ionization of the polar extracts. The heater wait time was 30 seconds, nitrogen was set as standby gas, and helium was used for measurements. The standard setting of 350 V was kept for the grid voltage.

Measurements were carried out using either the Quick-Strip or the Dip-it method. For the Quick-Strip method, 4 μL of each extract was pipetted onto a Quick-Strip card (Ion Sense, Saugus, MA, USA). One sample extract was applied to every second spot, and the other spots were left empty. The free spot before each sample served as an air blank to check for sample carryover. For each sample, a new measurement was started to obtain individual data files.

The Dip-it method was performed using Dip-it glass rods (Ion Sense, Saugus, MA, USA). Application of the sample extracts was done by dipping the glass rods into the extracts for approximately 2 seconds, followed by attaching the glass rod at the second position of the Dip-it linear rail. As with the Quick-Strip method, the first empty position on the linear rail was used as an air blank to check for sample carryover. The described procedure of the Dip-it sample application method was used with the same settings as mentioned above. To ensure a valid method, samples were extracted and analyzed in random order to minimize systematic effects or instrumental deviations. In addition, the total ion chromatograms (TIC) and mass spectra of a quality control (QC) sample were analyzed at regular intervals to check for systematic data deviations and batch effects. The QC sample was prepared by mixing 10μL of all extracts obtained.

Marker identification was performed by MS/MS experiments, for which selected m/z ratios were fragmented. A collision energy of 20 eV was used, and if the fragmentation was not sufficient, the collision energy was increased to 40 eV. Samples from different countries were used until a fragment spectrum with high signal intensities was obtained.

2.5. Data handling and chemometrics

For each extracted sample, a full scan mode measurement of the air blank and sample spot was recorded in positive and negative ionization using the Agilent MassHunter Data Acquisition 10.00 software (Agilent Technologies, Santa Clara, CA, USA). Agilent MassHunter Qualitative Analysis 10.0 (Agilent Technologies, Santa Clara, CA, USA) software was used to display and visualize the spectra. A database comparison using Lipid Maps and the Human Metabolome Database (HMDB) were applied to assign the high-resolution masses to possible metabolites [28,29].

The conversion of Agilent .d files to .mzML files was performed using msConvert software. No additional filters were used for this step. Data processing with MZmine (v 2.53) and chemometric evaluation were carried out separately for the samples of each extraction and ionization method [30]. Mass detection was performed in centroid mode with a noise level of 10. Chromatogram building was performed using the ADAP chromatogram builder with a minimum group size of 3, and group intensity threshold as well as a minimum highest intensity of 150. The m/z tolerance was set to 0.01 or 5.0 ppm. The time segment was set to 0.40-0.80 min to cut out the sample peak and separate it from the air blank in the same data file. Deisotoping was performed with an m/z deviation of 0.01 and a retention time deviation of 0.20 minutes. For the alignment, a larger m/z deviation of m/z 0.08 was chosen to ensure alignment of equal peaks with wider buckets, and the m/z to retention time ratio was set to 10:1. Hereafter, a filter was applied to exclude all features that were not present in at least 20 samples. Finally, gap-filling was applied with a retention time tolerance of 0.20 minutes and an m/z deviation of 0.01 or 5.0 ppm. The obtained data table was exported as .

csv file.

Method optimization was evaluated using PCA and RF in MetaboAnalyst 5.0 software [31]. After preprocessing with sum normalization and median standardization, the first three principal components (PC) were examined, and RF models were built using 5000 trees. An mtry corresponding to the square root of the number of variables (non-polar positive = 33, non-polar negative = 40, polar positive = 22, polar negative = 24, Quick-Strip = 25, Dip-it = 24) and nodesize = 1 were used. To estimate the suitability of the different workflows, the OOB errors of the different RF models were compared.

The final analysis of the samples from six countries (dataset 3) was carried out using R (v 4.4.2). PCA was calculated using the packages factoextra (v 1.0.7) and ggplot2 (v 3.5.1) was used for visualization of the results. The RF analysis was performed with the ranger package (v 0.17.0) and the following parameters: ntree = 5000, mtry = 24 (corresponding to the square root of the number of variables), and min.node.size = 1. In addition, the case.weights parameter was chosen to compensate for class imbalances by adjusting for the size of each class. Boruta was applied using the Pomona package (v 1.0.2) with the same RF parameters and the following additional parameters: pvalue = 0.01, maxruns = 100 [32]. For the visualization as a heatmap the signal intensities of the identified features were aggregated for each country by calculating the mean value across all masses of this country using the aggregate() function. The data were min-max normalized and visualized as a heat map using the pheatmap package (v 1.0.12).

3. Results and discussion

3.1. Optimization of the extraction process and ionization mode

To determine the best approach for identifying the origin of strawberries, metabolites were extracted twice from 62 freeze-dried samples (dataset 1). Once with the polar extraction method and once with the non-polar extraction method. The extracts obtained were measured by DART-MS in positive as well as in negative ionization mode using the Dip-it method.

The results are shown in Table 1. In addition, examples of resulting mass spectra of each extraction and ionization combination are shown in Figure S1 of the electronic supplementary material. When evaluating the RF classification of all four approaches, an accuracy of 72.1% for NP-pos and 69.4% for NP-neg was achieved. P-pos and P-neg yielded lower accuracies of 63.5% and 62.9%, respectively. These results indicated that the non-polar extraction method performed better than the polar extraction method in terms of classification accuracy. This result is also consistent with the number of detected features, which was higher for non-polar extraction with 1118 (NP-pos) and 1583 (NP-neg) features, than for polar extraction with 469 (P-pos) and 554 (P-neg) features.

In summary, non-polar extraction proved superior to the polar approach. With an accuracy of 72.1%, the F-NP-pos method achieved the best result, which is why this extraction approach was used for further method optimization. This conclusion is consistent with the results from comparable studies in which LC-MS was used to classify asparagus or hazelnuts, and this extraction approach was used for further method optimization [27,33,34].

3.2. Optimization of the sample delivery system

The next step was to compare the sample delivery methods that could be carried out using the Quick-Strip or Dip-it method. For these experiments, dataset 2 was used to separate German and non-German samples. The results are shown in Table S4 in the Supplementary Electronic Material. The Quick-Strip approach resulted in 605 features and a classification accuracy of 86.5%, while the Dip-it sample holder resulted in 597 features and a classification accuracy of 87.0%. Since the numbers of features and classification accuracies were almost the same, differing by only 0.5%, it can be concluded that there was no significant difference in classification with the different sample delivery systems. When using the Quick-Strip card, the extracts have to be pipetted, whereas when using the Dip-it method, only the glass rods have to be dipped into the sample extract. Since the pipetting step is eliminated, the Dip-it method is the faster approach and was selected for the further procedures. Furthermore, this result shows that a differentiation

Table 1
Confusion matrices showing the percentage accuracies of the RF classifications for determining the geographical origin of the strawberry samples in dataset 1. A total of 12 samples from Egypt, 20 samples from Germany, 15 samples from Greece and 15 samples from Spain were used. The overall classification accuracy was 63.5% for the polar extracts measured in positive ionization mode and 62.9% for the polar extracts measured in negative ionization mode. The data evaluation of the non-polar extractions resulted in 72.1% in positive and 69.4% in negative ionization mode.

Predicted Class [%]

Positive Ionization

True Class [%]	Egypt	Germany	Greece	Spain
Polar Extraction: Egypt	58.3	41.7	0.0	0.0
Polar Extraction: Germany	5.0	85.0	0.0	10.0
Polar Extraction: Greece	6.7	0.0	60.0	33.3
Polar Extraction: Spain	6.7	0.0	46.7	46.7
Non-polar Extraction: Egypt	58.3	8.3	16.7	16.7
Non-polar Extraction: Germany	0.0	95.0	0.0	5.0
Non-polar Extraction: Greece	7.1	21.4	42.9	28.6
Non-polar Extraction: Spain	0.0	0.0	20.0	80.0

Negative Ionization

True Class [%]	Egypt	Germany	Greece	Spain
Polar Extraction: Egypt	58.3	33.3	0.0	8.3
Polar Extraction: Germany	5.0	85.0	0.0	10.0
Polar Extraction: Greece	6.7	6.7	60.0	26.7
Polar Extraction: Spain	0.0	13.3	46.7	40.0
Non-polar Extraction: Egypt	58.3	8.3	16.7	16.7
Non-polar Extraction: Germany	0.0	100.0	0.0	0.0
Non-polar Extraction: Greece	6.7	13.3	46.7	33.3
Non-polar Extraction: Spain	0.0	6.7	33.3	60.0

between German and non-German samples is possible with promising accuracy when the non-polar extraction procedure, the positive ionization mode and sample application with the Dip-it approach are used.

3.3. Application of the optimized method

The optimized parameters were applied to a real market-relevant sample from dataset 3. For this purpose, 207 strawberry samples from six countries: Germany, Greece, Egypt, Spain, Poland, and the Netherlands were analyzed using the DART-MS approach, resulting in the detection of 610 features.

PCA was performed and the scores and loadings plots of the first three principal components are shown in Fig. 2 and Figure S2, respectively. The scores of the first two principal components (Fig. 2A) do not show any grouping based on the geographical origin of the samples. Moreover, the score plot of PC 2 vs. PC 3 was investigated, which demonstrates an enhanced clustering of the samples. The samples from Greece, Spain, and Egypt are located in the positive region of PC 3, whereas the samples from the Netherlands and Poland are located in the negative region of PC 3. The samples from Germany are mostly clustered in the center of the plot. Furthermore, the loadings of the PCs were investigated, which are presented in Figure S2 of the supplementary material. The loadings of PC 2 and PC 3 demonstrate notable correlations, such as a strong positive correlation from m/z 800-920 with PC 2 and a high negative correlation with PC 3. The features within this m/z range were traced back to triacylglycerols (TG) (see section 3.4) and the region of m/z 550-670 was assigned to diacylglycerols (DG). The score plots, in particular those of PC 2 and PC 3, show that the dataset contains information that is relevant to the origin, but that a clear separation of the classes cannot be achieved with unsupervised methods.

Therefore, RF was used as a supervised method to classify the samples and select features relevant for the separation. The results of the classification are presented in Table 2 and an overall accuracy of 63.8% is achieved. The samples from the six countries were correctly classified with an accuracy ranging between 48.1 and 78.4%, which exceeds the a priori probability of 16.7% for all classes. It can be concluded that the classification rates indicate the possibility of adequate classification, although there were considerable overlaps between some classes.

The two classes with the lowest classification accuracies are the Greek samples with 48.1% and the Spanish samples with 55.6% correct classifications. It is noteworthy that in each case there is a significant number of misclassifications in the other class. Overall, 29.6% of the Greek samples were incorrectly assigned to the Spanish class and, conversely, 30.6% of the Spanish samples were incorrectly assigned to the Greek class. This overlap between Greek and Spanish samples can be attributed to the climatic similarities between the two countries, as Huelva in Spain and Western Greece, where the samples mainly come from, have a Mediterranean climate with comparable precipitation levels and temperatures [35,36]. The model is therefore able to distinguish the Greek and Spanish samples from the samples of the other countries, but it is not possible to distinguish between these two countries.

66.7% of the samples from Egypt were classified correctly, and 15.2% and 12.1% of the Egyptian samples were wrongly identified as Greek and Polish, respectively. The similarity between the Greek and Egyptian samples may be attributed to their shared geographical location on the Mediterranean Sea, which results in comparable climatic conditions [37,38]. Furthermore, 78.4% of the samples from Germany were correctly assigned with misclassifications in the other countries ranging from 2.0% to 5.9%. These results show that the German samples can be sufficiently classified and that there is only a small overlap with other classes.

In addition, 12.1% of the samples from Egypt were misclassified as Polish samples, while 21.4% of the samples from Poland were classified as originating from Egypt. Due to the considerable distance between the countries concerned, similar geographical locations and climates cannot be used as a possible explanation. Presumably, the measurement approach used is not sufficiently high-resolution or sensitive to detect a sufficient number of marker compounds suitable for a more detailed discrimination.

The samples from the Netherlands were classified correctly in 59.4% of cases, while all other countries of origin showed misclassifications ranging from 3.1-12.5%. The classification accuracy of the Dutch samples shows that it is possible in principle to identify the Dutch samples, albeit with less accuracy. However, misclassifications are evident for the other two Central European countries, Germany, and Poland, at a rate of 9.4% each. The similarities in climate and geographical location of these three countries in the Central European Plain most likely led to the observed overlaps [39–41]. Poland, Germany and the Netherlands are located in close proximity to each other, stretching from east to west along the Baltic Sea and North Sea. These countries also have fairly similar climates and geography. These factors likely prevented the strawberry plants from developing country-specific markers, resulting in less classification accuracies when distinguishing these countries.

3.4. Classification of German strawberries

In addition to the classification of the individual countries, dataset 3 was again examined for the possibility of distinguishing German strawberries from non-German strawberries, meaning that the countries Egypt, Greece, the Netherlands, Poland, and Spain were analyzed as one group. The results of the RF classification are provided in Table 3 and show an overall classification accuracy of 90.3%.

A closer look at the accuracies of the individual groups shows a perfect classification of the non-German samples with 99.4%. In contrast, the accuracy of the German samples is much lower with 62.7%.

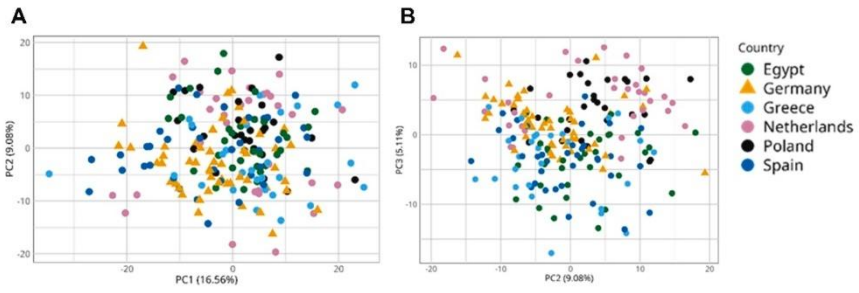


Fig. 2. Results of the PCA showing the scores of the first and second (A), as well as the second and third principal component (B) for the analysis of dataset 3 containing of 207 strawberry samples from Egypt (33), Germany (51), Greece (27), the Netherlands (32), Poland (28) and Spain (36).

Table 2
Percentage accuracies of the random forest classification of dataset 3 using the optimized method on 207 samples from Egypt, Germany, Greece, the Netherlands, Poland and Spain achieving a total accuracy of 63.8%. The number of samples analyzed for each country of origin is shown in the last column (#).

True Class [%] \ Predicted Class [%]	Egypt	Germany	Greece	NL	Poland	Spain	#
Egypt	66.7	3.0	15.2	3.0	12.1	0.0	*33*
Germany	2.0	78.4	3.9	5.9	5.9	3.9	*51*
Greece	3.7	7.4	48.1	0.0	11.1	29.6	*27*
NL	3.1	9.4	12.5	59.4	9.4	6.3	*32*
Poland	21.4	0.0	0.0	14.3	64.3	0.0	*28*
Spain	0.0	5.6	30.6	8.3	0.0	55.6	*36*

Table 3
Percentage accuracies of the random forest classification of dataset 3 using the optimized method on 207 samples to distinguish between German and non-German samples. The number of samples analyzed for each country of origin is shown in the last column (#). The samples were classified using the RF algorithm and the accuracy was 90.3%.

True Class [%] \ Predicted Class [%]	Germany	Other Origin	#
Germany	62.7	37.3	51
Other Origin	0.6	99.4	156

This large discrepancy between the two accuracies is due to the different numbers of samples between the two classes. With 156 to 51 samples, the class of non-German samples is significantly overrepresented, leading to unbalanced classification rates. This fact is confirmed by the results of the method optimization (see chapter 3.2), since the classification accuracies for a dataset consisting of two equally sized classes were 95.2% and 77.9% (see Table S4).

When optimizing the measurement method with dataset 2, an overall accuracy of 87.0% was achieved, which is very close to the 90.3% achieved here. These results demonstrate that it is possible to classify German and non-German samples based on the two different datasets. However, in the case of the unbalanced dataset 3, the accuracy for the classification of German samples is lower. In addition, there are similarities between the strawberry-producing countries, especially between neighboring countries, which have already been discussed in detail in chapter 3.3.

3.5. Selection and identification of relevant features

The *m/z* ratios of the 152 features were selected by the application of Boruta and the respective *m/z* ratios are listed in Table S6 of the electronic appendix. As many of these selected features as possible were then identified using DART-MS/MS. However, the possibilities for identification using DART-MS/MS are restricted by the fact that some lipid classes are ionized to a limited extent using the DART-source [17]. The relative signal intensities of the 5 identified features are shown in Fig. 3.

The identified triglycerides, such as TG 54:8 (Fig. 3A), show only minor differences between the countries, although the intensities of the Greek samples are slightly higher. However, the analysis of intensities reveals discrepancies between the intensities of the markers in certain countries, as illustrated by the Polish and Greek samples. Consequently, the identified TGs contribute only little to the classification result, which is reflected by the small number of relevant signals with *m/z* 800-920, as only 3 out of the 152 selected masses are affected.

In general, the DGs allow some differentiation between the countries, but not for complete separation. For example, DG 36:2, DG 36:3 and DG 36:4 (Fig. 3B, C and D) show higher signal intensities in Greek samples, whereas the signal intensities are lower in German, Polish and Dutch samples. The signal intensities of DG 34:1 (Fig. 3E) is lower in the Dutch and Spanish samples, which makes these samples different from those of the other classes. The higher influence of the DGs on the classification compared to the TGs is also reflected in the increased number of 16 selected masses in the range of *m/z* 550-650. This reduced number of selected masses in the DG and TG range also explains the low number of selected and identified masses.

Furthermore, an attempt was made to identify additional lipid masses in order to assign the mass ranges to the classes of compounds and to gain a more detailed understanding of the chemical fingerprint of strawberries. All identified masses are listed in Table S5 of the Supplementary Information. In order to get a general impression of the signal intensities across the different countries of origin, these were averaged for each feature for the individual countries and the results are shown in Fig. 4. The signals of TG 42.2, DG 36.2 and DG 34:1 showed high intensities in the Mediterranean countries Greece, Egypt and Spain and low signals in the Central European countries Germany, the Netherlands and Poland. Conversely, TG 44:4, DG 34:2, TG 52.3 and TG 52.4 showed

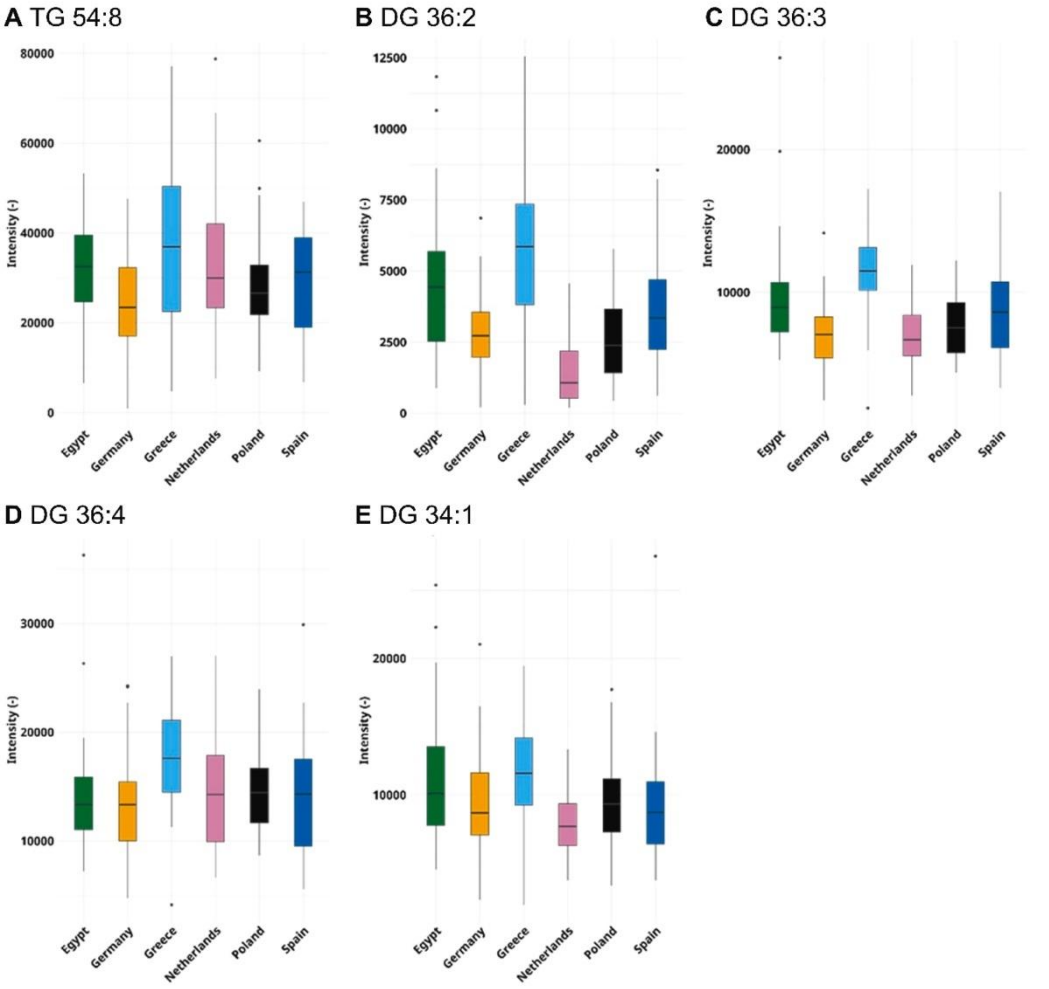


Fig. 3. Boxplots of the signal intensities of important features for the separation of strawberry samples from the six countries of origin: Egypt, Germany, Greece, Netherlands, Poland and Spain.

high signal intensities in the Central European countries and low intensities in the Mediterranean countries. Other identified signals show a more complex picture and do not allow any direct conclusions to be drawn about the climate zone of the countries. However, these different signal intensities can also be observed for different countries, which also contribute to the understanding of the samples.

Overall, the selected features and the heatmap of the identified compounds (Fig. 3 and Fig. 4) show similarities within the Mediterranean and Central European countries. The similarities between Egypt, Greece and Spain, as well as Poland, Germany, and the Netherlands, show that the chosen metabolomics approach has a limitation that makes it challenging to distinguish countries that are geographically close to each other. The approach was chosen under the assumption that the plant metabolome reacts differently to climate differences, as has already been extensively studied [42–45]. However, when growing strawberries, the location and cultivation period are selected in such a way that the climate conditions are as optimal as possible for cultivation, which leads in the observed similarities. In addition, the countries of the Central European Plain do not show significant climatic differences.

Moreover, strawberries show considerable diversity in both varieties and growing conditions. New varieties come onto the market every year, increasing the diversity of the samples and complicating classification. In addition, strawberry plants are pre-cultivated in peat moss and then moved to their intended final location [46]. This pre-cultivation step can also be carried out in different geographical locations, which makes classification according to geographical origin even more challenging. Depending on the location, pre-grown seedlings are used that bear fruit after just a few months. Consequently, the plants at the fruiting location had a short growing period and thus a limited window of opportunity to develop characteristic marker compounds, which explains the

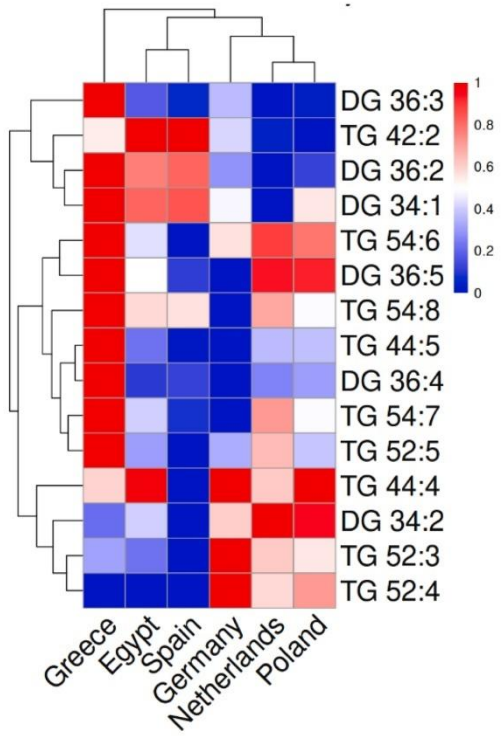


Fig. 4. Results of the heatmap visualization of the signal intensities of the identified signals. The color reflects the peak intensity of the signals.

similarities between strawberries samples of different countries.

Another factor influencing cultivation is the protection of strawberries from adverse weather conditions through the use of foil tunnels. The use of these tunnels allows earlier harvesting of strawberries and the fruits are no longer exposed to precipitation and temperature fluctuations. In this way, the exogenous factors that can lead to differences in the metabolome are further reduced, which in turn makes it more difficult to distinguish between different origins [47]. In summary, the similar growing conditions result in only minor metabolic differences in the sample groups and therefore limit the classification when distinguishing between closely spaced locations. However, despite the existing challenges, the results are quite promising for distinguishing between sample sites that are further apart.

Another limitation that has a negative impact when there are only very slight differences in the metabolites, is the chosen DART ionization approach. The DART-MS approach offers several advantages over commonly used methods, such as fast, environmentally friendly and cost-effective measurements. However, the speed of the method comes at the expense of limited number of detectable features. The short measurement time of only five seconds, combined with the lack of chromatographic separation, which increases ion suppression, results in the detection of a reduced number of features. Furthermore, some lipid classes such as phospholipids are only ionized to a small extent during DART ionization because in-source fragmentation occurs [17]. As a result, some differences may not be well recognized, which limits the method when applied to difficult issues.

Therefore, the very small differences in the signal intensities and the chosen analytical technique contribute to the reduced classification outcome (see Chapter 3.3) and the limited possibility of marker identification. Finally, the example of strawberries origin classification illustrates the advantages of DART-MS based metabolome experiments as well as the limitations of applying this method to a highly complex dataset.

4. Conclusion

The aim of this study was to investigate the benefits and limitations of metabolome analysis for determining the geographical origin of strawberries. Optimization of the DART-MS-based approach revealed that the combination of a non-polar extraction procedure with a positive ionization mode using the Dip-it method is most promising for determining the origin of the samples. Applying the methodology to a real-world market dataset of 207 strawberry samples resulted in a classification accuracy of 63.8%, achieved by RF. A detailed investigation of misclassifications and marker compounds revealed that the classification of strawberries using DART-MS is only possible with a comparatively low degree of accuracy. The identified marker compounds showed strong similarities between neighboring countries. In addition, the detection of various markers was limited by the chosen ionization technology. The metabolomic approach used is therefore more suitable for simple questions, such as distinguishing between German and non-German strawberries, which was possible with a sufficient classification accuracy of 90.3%.

Funding

This study was performed with the project "Detection of the geographical origin of strawberries (*Fragaria* × *ananassa*) using ICP-MS and LC-MS/MS using chemometric evaluation methods". This IGF Project (22909 N) of the FEI is supported within the program for promoting the Industrial Collective Research (IGF) of the Federal Ministry of Economic Affairs and Climate Action (BMWK), based on a resolution of the German Parliament.

CRediT authorship contribution statement

Felix Schmauder: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Johannes Brockelt:** Writing – review & editing, Formal analysis. **Kim Brettschneider:** Writing – review & editing, Visualization, Software, Formal analysis. **Stephan Seifert:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Marina Creydt:** Writing – review & editing, Supervision, Funding acquisition. **Markus Fischer:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The authors thank Christian Ahlers, Lina Cvancar, Robin Dammann, Henri Lösel, Marie-Sophie Müller, Nils Wax and Soeren Wenck for their help with sampling and helpful discussions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.talo.2025.100496.

Data availability

Data will be made available on request.

References

[1] J. Selamat, N.A.A. Rozani, S. Murugesu, Application of the metabolomics approach in food authentication, Molecules. 26 (2021) 7565.
[2] E. Pichersky, E. Lewinsohn, Convergent evolution in plant specialized metabolism, Annu Rev. Plant Biol. 62 (2011) 549–566.
[3] J. Aszyk, H. Byliński, J. Namieśnik, A. Kot-Wasik, Main strategies, analytical trends and challenges in LC-MS and ambient mass spectrometry–based metabolomics, TrAC Trends Analyt. Chem. 108 (2018) 278–295.
[4] R.N. Trethewey, Metabolite profiling as an aid to metabolic engineering in plants, Curr. Opin. Plant Biol. 7 (2004) 196–201.
[5] M. Creydt, M. Fischer, Food profiling: progress in ensuring the authenticity of truffle species and their geographical origin, Curr. Opin. Food Sci. (2024) 101171.
[6] N. Mialon, B. Roig, E. Capodanno, A. Cadiere, Untargeted metabolomic approaches in food authenticity: a review that showcases biomarkers, Food Chem. 398 (2023) 133856.
[7] Simpson, D. The economic importance of strawberry crops. The genomes of rosaceous berries and their wild relatives 2018, 1–7.
[8] Food and Agriculture Organization of the United Nations. Worldwide production quantity of strawberries. Available online: https://www.fao.org/faostat/en/#data/QCL (accessed on 1 November 2023).
[9] Statistisches Bundesamt. Aus- und Einfuhr (Außenhandel): Deutschland: Erdbeerimport Deutschland 2021. Available online: https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de&query=erdbeeren+import#abreadcrumb (accessed on 1 November 2023).
[10] L. Nassar, I.E. Okwuchi, M. Saad, F. Karray, K. Ponnambalam, P. Agrawal, Prediction of strawberry yield and farm price utilizing deep learning, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–7. ISBN 1728169267.
[11] D.H. Suh, Z. Guan, H. Khachatryan, The impact of Mexican competition on the US strawberry industry, Int. Food Agribus. Manag. Rev. 20 (2017) 591–604.
[12] H.J. Kim, Y.T. Seo, S. Park, S.H. Jeong, M.K. Kim, Y.P. Jang, DART–TOF–MS based metabolomics study for the discrimination analysis of geographical origin of Angelica gigas roots collected from Korea and China, Metabolomics. 11 (2015) 64–70.
[13] F. Schmauder, M. Creydt, M. Fischer, Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (Corylus avellana L.), Food Chem. 467 (2025) 142265.
[14] H. Novotná, O. Kmiecik, M. Gałązka, V. Krtková, A. Hurajová, V. Schulzová, E. Hallmann, E. Rembiałkowska, J. Hajšlová, Metabolomic fingerprinting employing DART-TOFMS for authentication of tomatoes and peppers from organic and conventional farming, Food Addit. Contamin.: Part A 29 (2012) 1335–1346.
[15] T. Cajka, H. Danhelova, M. Zachariasova, K. Riddellova, J. Hajslova, Application of direct analysis in real time ionization–mass spectrometry (DART–MS) in chicken meat metabolomics aiming at the retrospective control of feed fraud, Metabolomics. 9 (2013) 545–557.
[16] K. Wang, L. Xu, X. Wang, A. Chen, Z. Xu, Discrimination of beef from different origins based on lipidomics: a comparison study of DART-QTOF and LC-ESI-QTOF, LWT 149 (2021) 111838.
[17] F. Schmauder, D. Schütz, M. Creydt, J. Riedl, M. Fischer, Food authentication goes green: determination of the geographical origin of grain maize (Zea mays L.) using direct analysis in real time mass spectrometry (DART-MS), Food Control 163 (2024) 110489.
[18] L. Vaclavik, K. Riddellova, T. Cajka, J. Hajšlová, GC–TOF-MS and DART–TOF-MS: challenges in the analysis of soft drinks, LCGC Europe 21 (2008) 250–256.
[19] T. Cajka, H. Danhelova, A. Vavrecka, K. Riddellova, V. Kocourek, F. Vacha, J. Hajslova, Evaluation of direct analysis in real time ionization–mass spectrometry (DART–MS) in fish metabolomics aimed to assess the response to dietary supplementation, Talanta 115 (2013) 263–270.
[20] A. Massaro, R. Stella, A. Negro, M. Bragolusi, B. Miano, G. Arcangeli, G. Biancotto, R. Piro, A. Tata, New strategies for the differentiation of fresh and frozen/thawed fish: a rapid and accurate non-targeted method by ambient mass spectrometry and data fusion (part A), Food Control 130 (2021) 108364.
[21] K. Fraser, G.A. Lane, D.E. Otter, S.J. Harrison, S.-Y. Quek, Y. Hemar, S. Rasmussen, Monitoring tea fermentation/manufacturing by direct analysis in real time (DART) mass spectrometry, Food Chem. 141 (2013) 2060–2065.
[22] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37–52.
[23] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[24] P. Probst, A.-L. Boulesteix, B. Bischl, Tunability: importance of hyperparameters of machine learning algorithms, J. Mach. Learn. Res. 20 (2019) 1–32.
[25] M.B. Kursa, W.R. Rudnicki, Feature selection with the Boruta package, J. Stat. Softw. 36 (2010) 1–13.
[26] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Can. J. Biochem. Physiol. 37 (1959) 911–917.
[27] M. Creydt, D. Hudzik, M. Rurik, O. Kohlbacher, M. Fischer, Food authentication: small-molecule profiling as a tool for the geographic discrimination of German white asparagus, J. Agric. Food Chem. 66 (2018) 13328–13339.
[28] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, LIPID MAPS online tools for lipid research, Nucleic. Acids. Res. 35 (2007) W606–W612.
[29] D.S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B.L. Lee, HMDB 5.0: the human metabolome database for 2022, Nucleic. Acids. Res. 50 (2022) D622–D631.
[30] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC. Bioinformatics. 11 (2010) 395, https://doi.org/10.1186/1471-2105-11-395.
[31] Z. Pang, J. Chong, G. Zhou, Lima Morais, D.A. de, Le Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, J. Xia, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic. Acids. Res. 49 (2021) W388–W396.
[32] F. Degenhardt, S. Seifert, S. Szymczak, Evaluation of variable selection methods for random forests and omics data sets, Brief. Bioinform. 20 (2019) 492–503.
[33] M. Creydt, M. Arndt, D. Hudzik, M. Fischer, Plant metabolomics: evaluation of different extraction parameters for nontargeted UPLC-ESI-QTOF-mass spectrometry at the example of white Asparagus officinalis, J. Agric. Food Chem. 66 (2018) 12876–12887.
[34] S. Klockmann, E. Reiner, R. Bachmann, T. Hackl, M. Fischer, Food fingerprinting: metabolomic approaches for geographical origin discrimination of hazelnuts (Corylus avellana) by UPLC-QTOF-MS, J. Agric. Food Chem. 64 (2016) 9253–9262.
[35] Deutscher Wetterdienst. Klimadiagramm von Huelva, Andalusien /Spanien. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_083830_di.pdf (accessed on 4 October 2023).
[36] meteoblue. Beobachtete historische Klima- und Wetterdaten für Patras. Available online: https://www.meteoblue.com/de/wetter/historyclimate/climateobserved/patras_griechenland_255683 (accessed on 4 October 2023).
[37] Deutscher Wetterdienst. Klimadiagramm Alexandria. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_623180_di.pdf (accessed on 29 October 2024).
[38] Deutscher Wetterdienst. Klimadiagramm Athen. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_167160_di.pdf (accessed on 3 December 2024).
[39] Deutscher Wetterdienst. Klimadiagramm Hannover. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_103380_di.pdf (accessed on 3 December 2024).
[40] Deutscher Wetterdienst. Klimadiagramm Warschau. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_123750_di.pdf (accessed on 3 December 2024).
[41] Deutscher Wetterdienst. Klimadiagramm Maastricht. Available online: https://www.dwd.de/DWD/klima/beratung/ak/ak_063800_di.pdf (accessed on 3 December 2024).
[42] M. Creydt, M. Fischer, Omics approaches for food authentication, Electrophoresis 39 (2018) 1569–1581.
[43] E. Cubero-Leon, R. Peñalver, A. Maquet, Review on metabolomics for food authentication, Food Res. Int. 60 (2014) 95–107.
[44] P. Zhong, X. Wei, X. Li, X. Wei, S. Wu, W. Huang, A. Koidis, Z. Xu, H. Lei, Untargeted metabolomics by liquid chromatography-mass spectrometry for food authentication: a review, Compr. Rev. Food Sci. Food Saf. 21 (2022) 2455–2488.
[45] D. Yu, S. Guo, X. Zhang, H. Yan, S. Mao, J. Wang, J. Zhou, J. Yang, Y. Yuan, J. Duan, Combining stable isotope, multielement and untargeted metabolomics with chemometrics to discriminate the geographical origins of ginger (Zingiber officinale Roscoe), Food Chem. 426 (2023) 136577.
[46] P. Legua, F. Hernández, F. Tozzi, R. Martínez-Font, D. Jorquera, C. Rosique Jiménez, E. Giordani, J.J. Martínez-Nicolás, P. Melgarejo, Application of LCA methodology to the production of strawberry on substrates with peat and sediments from ports, Sustainability. 13 (2021) 6323.
[47] K. Wysocki, J. Kopytowski, A. Bieniek, J. Bojarska, The effect of substrates on yield and quality of strawberry fruits cultivated in a heated foil tunnel, Zemdirbyste-Agric. 104 (2017).

6.2 Herkunftsauthentifizierung von Körnermais (*Zea mays* L.) mittels DART-MS.

Der in Abschnitt 6.1 etablierte Arbeitsablauf wurde auf die Herkunftsbestimmung von Körnermais übertragen. Hierfür wurde ein Datensatz aus 200 Körnermaisproben gewählt, der sieben unterschiedliche Länder von drei Kontinenten beinhaltete. Die Analyten wurden mit Hilfe von unpolaren Lösungsmitteln extrahiert und für die Untersuchung des Lipidprofils wurden die Extrakte mittels DART-MS-basierter Metabolomanalyse untersucht.

Die mittels PCA und *Random Forest* durchgeführte multivariate Analyse der 647 erfassten *features* zeigte, dass ein Trend zur Herkunftsunterscheidung in den Daten vorhanden ist und eine darauffolgende Klassifizierung mittels *Random Forest* ergab eine Richtigkeit von 84,4%. Mittels des angewendeten Algorithmus wurde die Wichtigkeit der *features* für die Klassifizierung der Herkunft bestimmt. Zur Aufklärung der relevantesten Markerverbindungen, die eine Korrelation mit den geographischen Herkünften zeigten, wurden zunächst Standards verschiedener Lipidklassen mittels DART-MS/MS vermessen. Anhand der erhaltenen Fragmentierungsmuster wurden die für die Herkunftsbestimmung wichtigen Markermetabolite identifiziert. Auffällig waren vor allem Marker aus den Lipidklassen der Mono-, Di- und Triacylglyceride sowie der freien Fettsäuren. Ein Vergleich der Intensitäten der Metaboliten zeigte, dass Unterschiede zwischen den Klassen bestehen und sich mittels der Aufgeklärten Marker Rückschlüsse auf die Herkunft der Proben ziehen lassen.

Hinweis zum Copyright:

This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

URL: https://www.sciencedirect.com/science/article/pii/S0956713524002068?via%3Dihub

DOI: https://doi.org/10.1016/j.foodcont.2024.110489

Food Control 163 (2024) 110489

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS)

Felix Schmauder [a], David Schütz [a], Marina Creydt [a], Janet Riedl [b], Markus Fischer [a,*]

[a] Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146, Hamburg, Germany
[b] Bundesinstitut für Risikobewertung, BfR, Max-Dohrn-Straße 8-10, 10589, Berlin, Germany

ARTICLE INFO

Keywords:
Authenticity analysis
DART-MS
Food and feed fraud
Lipidomics
Mass spectrometry
Metabolomics
Random forest

ABSTRACT

Time savings, environmental friendliness and cost reduction are important criteria to determine the authenticity of feed and food in routine analysis. One analytical technique that is especially promising in terms of these criteria in metabolomics studies is Direct Analysis in Real Time Mass Spectrometry (DART-MS). In the present study, DART-MS is applied to 200 grain maize samples from seven geographical origins to develop a data processing workflow and an approach for origin analysis of grain maize. For the first time, such a large sample set was investigated using a DART-MS-based non-targeted lipidomics approach. Various open-source programs were used for pre-processing and random forest (RF) combined with repeated cross-validation was used for classification, resulting in an accuracy of 84.4%. Subsequently, the most important markers identified by RF, belonging to the classes of mono-, di- and triacylglycerols, were identified using DART-MS^2. The chosen method significantly minimizes the time and solvent consumption compared to classic liquid chromatography electrospray ionization mass spectrometric (LC-ESI-MS) strategies and thus offers a more cost-effective and environmentally friendly alternative. Furthermore, the results show the high potential for determining the geographical origin of grain maize using DART-MS.

1. Introduction

In recent years, there has been increasing attention to green, fast and cost-effective methods for food and feed authenticity, which requires the development of new approaches, processes and hardware solutions. However, these requirements are offset by the fact that high-resolution methods have often to be used to reliably detect adulterated goods as counterfeits become more sophisticated. In this regard, especially non-targeted metabolomics strategies have gained importance. For metabolomics-based approaches usually two complementary analytical methods are used, nuclear magnetic resonance (NMR) spectroscopy and liquid chromatography mass spectrometry LC-ESI-MS-based technologies. The main advantage of NMR spectroscopic methods is their reproducibility and speed, while LC-ESI-MS-based methods have lower detection limits. Moreover, analysis based on LC-ESI-MS is associated with long analysis times, large data sets and complex data evaluation (Aszyk et al., 2018). Approaches to combine the use of NMR and LC-ESI-MS have already been tested, but are accompanied by an increased workload (Razali et al., 2018).

There is positive experience with testing non-polar extracts and LC-ESI-MS to determine the geographic origin of food and feed using metabolomics experiments (Creydt & Fischer, 2022; Creydt, Hudzik, et al., 2018). A subset of the grain maize dataset used for this study was also analyzed with good classification accuracy of 90.5%, using this approach focused on lipids (Schütz et al., 2021). In addition, parts of the sample set were also analyzed using near-infrared spectroscopy (NIR) and mid-infrared spectroscopy(MIR) (Achten et al., 2019; Schütz et al., 2022). These methods are well suited as rapid testing procedures and provided good classification results, but they do not offer the possibility to identify specific markers relevant for classification.

One method to close this gap between rapid and precise test procedures could be the use of direct analysis in real time mass spectrometry (DART-MS) couplings. DART sources offer the possibility to ionize samples directly and allow short analysis times with many detectable markers. In addition DART-MS instruments enable low detection limits, fast analysis times and the possibility of structural elucidation of

* Corresponding author.
E-mail address: markus.fischer@uni-hamburg.de (M. Fischer).

https://doi.org/10.1016/j.foodcont.2024.110489
Received 1 November 2023; Received in revised form 29 February 2024; Accepted 30 March 2024
Available online 1 April 2024
0956-7135/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

relevant marker compounds (Lian et al., 2017; Xu et al., 2017). Furthermore, ionization by means of a DART source has a lower risk of contamination in contrast to direct infusion (Gross, 2014). DART is a soft ionization technique that can be operated at atmospheric pressure. For this purpose, helium atoms are excited in the DART source by means of electrical voltage, which are then directed to a solid or liquid sample and cause analyte molecules to be ionized (Cody et al., 2005). So far, mainly small molecules have been successfully ionized and used for authentication. Most studies are based on the use of polar sample extracts and are therefore usually only focused on polar and medium-polar molecules (Angelis et al., 2021; Avula et al., 2015; Hrbek et al., 2018). Nevertheless, there are also a few publications that show the successful ionization of larger molecules, especially lipids extracted from e.g. olive oil, beef and chicken meat (Cajka et al., 2013; Vaclavik et al., 2009; Wang et al., 2021). In addition, most plant-based metabolomics studies are based on small sample sets and there is limited experience with sample numbers in the three-digit range (Avula et al., 2015; Kim et al., 2011; Novotná et al., 2012). An exception is a study on the origin analysis of *Angelica gigas* roots with 188 samples (Kim et al., 2015).

Another aspect that is increasingly becoming the focus of research is sustainability. While LC-ESI-MS analyses require large amounts of high-purity solvents, no solvents are necessary for DART-MS measurements. This advantage raises the entire analysis to a new level of sustainability. In addition, DART-MS systems require less maintenance due to the lack of chromatographic separation.

In order to be able to compare the performance of DART-MS measurements with those of LC-ESI-MS results, a DART source was connected to a quadrupole time-of-flight mass spectrometer (qToF), which in particular is suitable for a non-targeted analysis, because of its high mass accuracy and the ability to detect many analytes simultaneously (Núñez et al., 2004). The samples investigated are grain maize samples from Spain, France, Germany, Peru, the Ukraine, Hungary and the USA and the grain maize data set that had already been measured with LC-ESI-MS (Schütz et al., 2021).

Grain maize is one of the most important field crop and has a significant contribution to global security of food supply (United States Department of Agriculture, 2023). As an important food and animal feed, grain maize is particularly susceptible to fraudulent declarations of origin. The cultivation site influences not only the nutritional quality of the feed but also the risk factor for contamination with mycotoxins (Khodaei et al., 2021). In this way, the development of methods serves not only the rapid authentication, but also the safety in the feed and food supply chain. In addition, crop failures can contribute to local and annual price reductions, which plays a role in falsification of grain maize origins. Consumers as well as farmers who purchase grain maize as animal feed are affected by these adulterations. The USA and major European countries, such as France, Germany, Hungary and the Ukraine, contribute a large share of the world's grain maize harvest and were therefore selected for the model. Peru was chosen as an example of a South American country, as this continent is a major grain maize producer (United States Department of Agriculture, 2023).

This study attempts to develop a new DART-MS workflow for the authentication of grain maize with regard to geographical origin based on lipidomics and compare its performance with an LC-ESI-MS approach. For this reason, a total of 200 samples from the seven countries mentioned above, were investigated using a non-polar extraction and a DART-MS platform. The data set had already been measured by LC-ESI-MS, allowing a comparison to be made (Schütz et al., 2021). According to our best knowledge, this is the first time that a large dataset of 200 samples had been measured with a DART-MS platform for the determination of origin. Since this new workflow was a proof-of-concept study, the focus was on the detection of pure countries and mixtures were not considered.

Another aim was to use open-source software tools for data analysis to establish a workflow that is freely accessible to potential end users. RF was used for classification, as it is known to be well suited for high-dimensional data sets containing large numbers of variables and comparatively low numbers of samples (Biau & Scornet, 2016). During development two RF types for general classification of origin and one vs. rest models were explored. Subsequently, relevant markers were determined by RF importance scores and identified in order to characterize the differences of the classes and to get an impression of the fragmentation patterns of MS/MS spectra that can be obtained using a DART source.

For performance comparison of DART-MS with the LC-ESI-MS approach, sustainable aspects, classification results of different RF models, costs and metabolites were compared.

2. Materials and methods

2.1. Chemicals

Ultrapure water was obtained by using a Merck Direct-Q 3 UV-R (Merck, Darmstadt, Germany) Millipore water purification system. Chloroform (HPLC grade), acetonitrile (LC-MS grade), methanol (LC-MS grade) were purchased from Carl Roth GmbH (Karlsruhe, Germany). Helium 5.0 was purchased from Westfalen Gas (Münster, Germany). Phosphatidylethanolamine (PE) and glycerophosphocholine (PC) standards (Avanti Polar Lipids, Alabaster, Alabama, USA) as well as the monoacylglycerol (MG), diacylglycerol (DG) and triacylglycerol (TG) standards (all from Sigma-Aldrich, St. Louis, Missouri, USA) were purchased in LC-MS quality.

2.2. Grain maize samples

The grain maize samples from the already mentioned publication by Schütz et al. (Schütz et al., 2021) were used and expanded by additional samples from Germany. A total of 200 samples were available. Of these, 26 samples came from Germany, 37 from France, 50 from Hungary, 20 from Spain, 25 from Peru, 24 from the United States and 18 samples from Ukraine.

The French samples were from the region Alsace. The samples from the USA were collected in different states like Alabama, Arkansas, Georgia, Louisiana, Mississippi and Texas, to cover the size of the USA. The German samples were from variety trials in the regions of Bavaria, Brandenburg and North Rhine-Westphalia and contain nine different varieties. Also, the Spanish samples came from variety trials, which minimize the influence of different cultivars on the model. Further information on the samples can be found in Table A1 inAppendix A. The samples from Peru, the USA, Spain and Germany are considered authentic as they were received after agricultural production from primary sources through cooperating institutions. In contrast the samples from Ukraine, Hungary and France were provided by state institutions and producers of commercial goods. They were taken at later stages of the supply chain in Germany. Authenticity of these samples relies on known supply chains and, in the case of Ukraine, on information provided by the producers. Shortly after shipment, the samples were subjected to sample preparation to ensure that the metabolome remains unaltered.

2.3. Sample preparation and extraction

The grain maize samples were ground, homogenized and freeze-dried as mentioned by Schütz et al. (Schütz et al., 2021): Grinding of the samples from Peru, the USA, Spain and Ukraine was done with an ultra-centrifugal mill (ZM 200, mesh size 0,5 mm, Retsch GmbH, Haan, Germany), followed by homogenisation with the help of a drum hoop mixer (3 h, 28 rpm, RRM 100. Engelsmann AG, Ludwigshafen, Germany). The samples from Germany, France and Hungary were ground with dry ice in a knife mill (Grindomix GM 300, Retsch, Haan Germany). Dry ice was used during grinding to protect the samples from process heat. All samples were freeze-dried for 24 h using a freeze-drying system

(Beta 1–8 LDplus, Christ, Osterode am Harz, Germany) and stored at −21 °C until the extraction step.

The non-polar compounds were extracted according to a method by Creydt et al., who had previously optimized the extraction process for the detection of lipids from food using mass spectrometry (Creydt, Arndt, et al., 2018). 50 ± 0.5 mg each of the lyophilized samples were weighed into a 2 mL safe-lock tube (Eppendorf, Hamburg, Germany). Then two steel balls (Ø 3 mm, THS, Mühlacker, Baden-Württemberg, Germany) and 750 μL of a pre-cooled chloroform/methanol mixture (1:2; v/v) were added, and the metabolites were extracted with a ball mill (BeadRuptor24, Omni International, Kennesaw, Georgia, USA, 3.1 m/s) for 1 min. Subsequently, 500 μL of ultra-purified water and 250 μL chloroform were added. The suspension was extracted again for 2 min in the ball mill with the conditions mentioned above. The extract was then centrifuged for 20 min at 12700 rpm and a temperature of 4 °C (Eppendorf 5430 R, Hamburg, Germany). Afterwards, 500 μL of the lower non-polar phase was transferred to a new extraction tube. The tube was centrifuged again for 7 min with the conditions mentioned above to remove any remaining precipitation. After centrifugation the liquid supernatant was used directly for analysis and the extract was cooled on ice.

2.4. Mass spectrometry measurements

DART-MS analyzes were performed using a DART ion source equipped with a DIP-it sample holder and an automatic linear rail (Ion Sense, Saugus, Massachusetts, USA). This setup was placed in front of a qToF mass spectrometer (Agilent 6560 IM-QTOF-MS, Agilent Technologies, Santa Clara, California, USA). A vapor interface with a ceramic capillary (Ion Sense, Saugus, Massachusetts, USA) was used to connect the DART source to the qToF. A 1.5 cm wide gap was maintained between the DART source and the capillary to install the linear rail.

Furthermore, a vacuum pump (Vacuubrand MZ 2NT, 7 mbar, 2.2 m^3/h, Wertheim, Germany) was used to generate the pre-vacuum in the vapor interface and draw in ambient air in front of the DART source to transport the generated ions into the capillary. The DART temperature was set at 350 ± 3 °C, the heater wait time was 50 s and the pulse duration was 5 s. Nitrogen was chosen as the standby gas, while helium was used as the measurement gas. The grid voltage was maintained at the default settings of 350 V. Both, the DART ion source and the qToF, were operated in positive mode and an *m/z* range of 100–1000 was set for the qToF. The spectral rate of the qToF instrument was set to 10 Hz. The mass calibration of the qToF device was performed directly before the DART-MS measurements with the ESI source (Dual JetStream, Agilent Technologies, Santa Clara, California, USA) using the Agilent Technologies ESI tuning mix and tuning procedure for small masses (*m/z* 50–1700). After mass calibration (*m/z* 50–1700) was passed, the ESI source was removed, and the DART source was installed.

At the beginning of the measurement, a new air blank sample run was performed for each sample, leaving the first sample position of the DIP-it holder empty. Hereafter, the samples were applied using the DIP-it samplers. The glass rod was dipped into the extract for 2 s and placed in the second slot of the DIP-it sample holder. During the measurement, the same sample extract was measured every 2 h to check repeatability. Further the total ion chromatograms of all samples were overlaid and examined for outliers.

2.5. Marker identification

Since there is limited information available regarding the ionization and fragmentation behavior of non-polar molecules using on DART-MS^2 analysis, standards of different analyte classes were measured and the ionization and fragmentation patterns were revealed. To prepare the standards, about 5 mg of each standard was weighed into a 2 mL safe-lock tube and dissolved in 1 mL methanol. The application into the DART source was carried out in the same way as the samples via DIP-it sampler and with the same procedure. For the MS^2 measurements, the spectral rate of the qToF instrument was increased to 25 Hz, so that a greater intensity of the fragments could be achieved. The collision energy was 20 eV, if no fragmentation could be observed, it was increased to 40 eV. Only one marker was specifically fragmented per measurement (targeted mode). Different samples were measured one after the other until the desired *m/z* was fragmented, and a satisfactory fragment spectrum could be obtained.

2.6. Data handling and chemometrics

A full scan was recorded for each sample using the Agilent MassHunter Data Acquisition 9.00 software (Agilent Technologies, Santa Clara, California, USA). To obtain universally useable mzML files the Agilent .d file format was converted using MsConvert (Chambers et al., 2012). MsConvert was only used for data conversion without applying any additional filters or settings. For feature finding and sample alignment, MZmine version 2.53 (Pluskal et al., 2010) was used. In the first step, feature detection in centroid mode and chromatogram building with the ADAP chromatogram builder was performed. The retention time was limited to 0.50–0.80 min to separate the measurements of the samples from the previously measured blank values. Further, the retention time filter was applied to set the integration limits: The peaks were integrated between the retention time of 0.50–0.80 min and the area values were used. The following parameters were applied: a minimum group size of 3, a group intensity threshold, and a minimum highest intensity of 150. Additionally, an *m/z* tolerance of 0.0100 and 5.0 ppm were employed. Deisotoping and alignment were performed with the same *m/z* deviations as in the chromatogram building and a retention time deviation of 0.20 min. Since DART measurements do not involve a time separation of the masses, the *m/z* ratio was weighted ten times more than the retention time in the alignment. Subsequently, the aligned data were filtered: Only those masses that occurred in at least 150 of the 200 measurements and had an *m/z* of 200–1000 were kept.

As indicated by the measured blanks, most airborne masses were between *m/z* 100–200. Since the interfering ions from the air occurred in an *m/z* range of 100–200 and the relevant masses were expected at higher *m/z*, ions with an *m/z* below 200 were excluded (Schütz et al., 2021). An intensity tolerance of 50%, an *m/z* deviation of 0.0050 or 5.0 ppm and a retention time tolerance of 0.20 min were chosen for gap filling.

For unsupervised analysis of the main variances in the data set, a principal component analysis (PCA) was performed. For the PCA, the data were imported into the MetaboAnalyst 5.0 software and pre-processed using sum normalization and median standardization (Pang et al., 2021). Analysis with MetaboAnalyst continued with a Partial Least-Square Discriminant Analysis (PLS-DA) model, with the pre-processing steps described above. Validation was performed using a 10-fold cross-validation. In addition, RF was used to evaluate the data using a supervised method, using R (4.2.2) and R Studio (2022.12.0), as well as the packages mlr in version 2.19 and randomForest in version 4.7–1.1. The data pre-processing was kept for classification. However, data pre-treatment has limited impact on the analysis using RF, as it is a ranking-based procedure (Chen et al., 2004, p. 24).

The default settings of the randomForest function were used for the mtry and nodesize parameters. For RF analysis, the validation was performed as a 10-fold-cross-validation. To take account of the class imbalance, each RF of the10-fold-cross-validation was performed as balanced RF. Balanced RF uses subsampling of the training data for each tree and 15 samples per class were used for training (Chen et al., 2004, p. 24). Thus, all samples, except for the test data of the cross-validation, were used for training the RF, but subsampling was performed within the RF to balance the classes. The samples were classified using the common RF method: The test data was entered into the trained model and classified into each individual decision tree. Each tree outputs a class and, a majority vote was done by the RF algorithm. A sample was

assigned to the class for which the most trees gave their votes, as default cut-off values were used (1/k with k = 7 (number of classes)). Therefore each sample was assigned to the class with the highest percentage of votes and the votes can be distributed differently between the classes. The results for all samples were obtained by averaging the results of the 10-fold cross-validation.

Two different types of RF models were trained: First, the samples from all seven countries were used as individual classes. This model was performed as described above in a 10-fold-cross-validation with balanced RF and was repeated 100 times. Furthermore, one vs. rest models were obtained for the samples of each country with the same cross-validation and 10 repetitions. For the evaluation of variable importance, the mean decrease in accuracy was used and the 100 most important variables were reported (Table A2). For those variables, the MS^2 spectra were evaluated using Agilent Mass Hunter Qualitative Analysis 10.0 (Agilent Technologies, Santa Clara, California, USA).

3. Results and discussion

In order to develop a fast, inexpensive and environmentally friendly analytical method for determining the geographical origin of grain maize, 200 grain maize samples were measured using a DART-MS platform. An example of a DART-MS measurement is shown in Fig. 1.

According to the preliminary study using LC-ESI-MS, the non-polar extraction conditions and a database comparison using Lipid Maps and the Human Metabolome Database (HMDB), it was assumed that most of the signals originated from lipids (Fahy et al., 2007; Wishart et al., 2022). This assumption could be confirmed. Thus, especially in the range of *m/z* 830–930, the spectrum shows signals with high intensities, which can be assigned to TGs. Also, in the range of *m/z* 530–650 many signals could be detected originating from DGs could be detected. In addition, MGs such as MG (18:2) (*m/z* 337.2831 as $[M + H–H_2O]^+$) and free fatty acids (FA), such as FA (18:2) (*m/z* 263.2365, $[RCO]^-$), could be observed. It was noticeable that the range between 650 and 830 *m/z* in the variable selection showed only a very low mean decrease accuracy and thus a lower importance for the separation of the classes. Therefore, further data evaluation focused on the top ranked features and relevant classes of DGs and TGs. The ranking of the 100 most important variables can be found in Table A2 of the supplementary information. In the blank samples, most of the signals could be detected at *m/z* less than 200. The ionization of small volatile molecules present in laboratory air is typical of atmospheric pressure ionization techniques. In order to remove these signals from the data evaluation, the considered *m/z* range was limited to *m/z* 200–1000 (Arora et al., 2022).

By converting and processing the data in MZmine 2, a table with 647 features was obtained. Due to the gap-filling function by the MZmine 2 software, only a few features are not found in some samples. Based on the resulting features a PCA was performed to analyze the main variances of the data (Fig. 2). Principal components (PC) one to four explain 57.1, 17.8, 11.6, and 3.7% of the variance, respectively. Based on the scores of the first PC, there is a clear separation into two clusters containing the samples from France, Hungary and Germany in one cluster and the samples from Spain, USA, Ukraine and Peru in the other cluster. The separation into two clusters can be explained by the fact that the samples from Germany, France and Hungary came from batches intended for starch production and a slightly different sample preparation was used. However, LC-ESI-MS analysis showed that apart from the concentration of DGs, there was only a minor difference between the feed corn or corn intended for starch production (Schütz et al., 2021). Within these two clusters, a trend toward spatial separation by country could be observed. For example, the German samples showed higher values for the first principal component than the French and Hungarian samples. In general, the clusters of the samples from different origin are overlap spatially in the scores of the first, second and fourth principal component, but the areas of overlap are small and a clear trend towards separation is notable.

Loading plots are shown in Figs. A1 and A2 of the supplementary information. Considering the loading plots of PC 1 vs. PC 2, PC 1 was mainly influenced by masses with an *m/z* ratio between 850 and 920. Thus, the features with an *m/z* of 874.7938, 881.7573, 898.7838 and 900.7977 were the ones with the highest loadings on the PC 1 which could be assigned to the class of TGs. PC 3 showed high loadings on the two features assigned to DGs, with an *m/z* of 601.5163 and 603.5298. Therefore, two PCs demonstrated high loadings on features that were assigned to the classes of DGs and TGs. Since the score plot showed a spatial separation of the countries in the direction of PC 1 and PC 3, it was concluded that DGs and TGs could be relevant for separating the countries. The score plot of PC 1vs. PC 4 indicated a good separation of samples by country and thus potentially important features occurred in the loading plot. Apart from the already known TGs and DGs, there were only a few features with high loadings. These were, for example, features with an *m/z* of 279.2326, 577.5221 and 634.5406 which could not be assigned to a clearly defined mass range. It was therefore possible to detect some initial masses of interest using the loading plots, but it was not possible to assign all masses.

The results of the PLS-DA are shown in the supplementary information in Table A3 and the score plot is given in Fig. A3. The 10-fold cross validation lead to an accuracy of 85.0% and a Q2 value of 0.62 for 8 components. The accuracy indicates a successful classification of the samples, but the rather lower Q2 value shows that the PLS-DA does not represent the dataset that well. Since the PLS-DA has problems with a small number of important variables in large datasets and tends to overfitting, this could be an explanation for the low Q2 value (Gromski et al., 2015). Furthermore, the Q2 values are only applicable to classification models to a limited extent, as they are statistics developed for regression analysis (Golbraikh & Tropsha, 2002; Westerhuis et al., 2008). All in all, the PLS-DA model was rejected due to possible overfitting and the occurrence of a non-linear dataset. Therefore, a

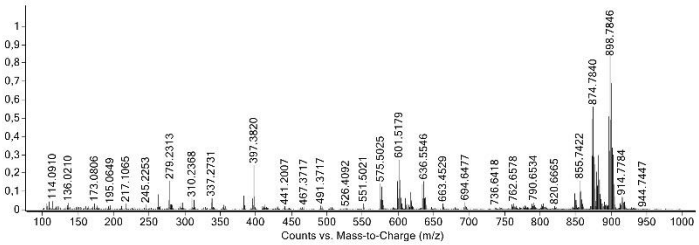


Fig. 1. Exemplary mass spectrum of a German grain maize sample, measured with DIP-it sample holder and a DART-MS instrument at 300 °C. Due to the non-polar derivatization, most of the analytes could be detected in the *m/z* range of the DGs (*m/z* 550–650) and TGs (*m/z* 800–950). Further signals can be recognized in the range of MGs (*m/z* 300–350) and FAs (*m/z* 230–300), whereas only a low signal frequency occurred in the range of phospholipids (*m/z* 650–850).

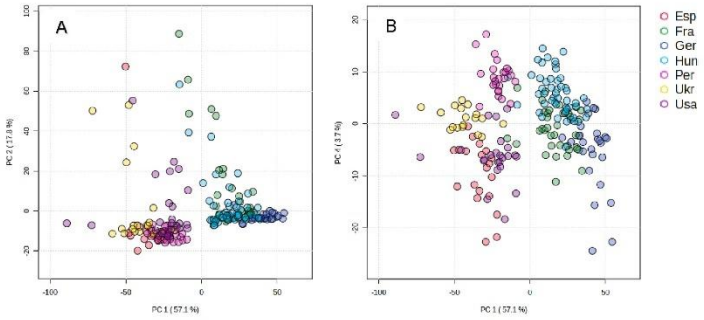


Fig. 2. Results of the PCA: Scores of the first and second (A) as well as the first and fourth principal components (B) for the DART-MS data after sum normalization and mean centering of the samples colored by country of origin are shown. The color code reflects the geographical origin of the countries.

Actual class \ Predicted class	Esp	Fra	Ger	Per	Ukr	Hun	USA	#
Esp	82.1	0.0	0.0	4.1	0.0	0.0	13.9	20
Fra	0.0	82.8	9.6	0.0	0.0	6.4	1.2	37
Ger	0.0	2.9	92.7	0.0	0.0	4.4	0.0	26
Per	3.7	0.0	0.0	95.7	0.0	0.0	0.6	25
Ukr	0.0	3.4	10.5	2.7	73.7	0.0	9.7	18
Hun	0.0	15.4	2.4	2.0	0.0	80.1	0.0	50
USA	6.2	0.0	4.1	4.3	0.6	0.0	84.9	24

Fig. 3. Confusion matrix of the random forest classification for the geographical origin determination of grain maize samples. Mean classification accuracies in percent of 100 times repeated 10-fold cross validations are shown for the countries Spain (Esp), France (Fra), Germany (Ger), Peru (Per), Ukraine (Ukr), Hungary (Hun), and the USA. The overall classification accuracy was 84.4% and the number of the samples for each country (#) is given in absolute numbers in the last column.

non-linear classification model was preferred.

3.1. Random forest classification

To classify the samples by country, an RF data evaluation was performed with all features and samples, achieving an overall accuracy of 84.4%. In detail, the confusion matrix in Fig. 3 shows that the accuracy for the samples of each country range from 73.7 to 95.7%. The samples from Peru and Germany showed the highest accuracies of 95.7 and 92.7%, respectively. For samples from Ukraine, a comparatively low accuracy of 73.7% can be observed, which could be due to the smaller number of samples. The Hungarian samples were classified with an accuracy of 80.1% and the misallocations between France and Hungary are striking. Both countries have a continental and warm climate with very similar annual values for temperature and precipitation, which probably made it more difficult to distinguish between these two classes (Deutscher Wetterdienst, 2023a; 2023d). In addition, there were a few misclassifications between samples from Spain and the USA that show accuracies of 82.1 and 84.9%, respectively. The reason for this could be that USA is a quite large country and, to account for the size, samples were taken from different states with a large distance to each other. This approach offers the advantage of better coverage of the USA in the classification model, but also creates a more diffuse class and similar weather conditions may have also caused the overlap of the samples from Spain and the southern US states (Deutscher Wetterdienst, 2023b; 2023c). Probably, this overlap could be resolved with a larger data set containing more samples from the USA in order to sufficiently represent the large variance of this class in the model.

In order to analyze the correct classification of the samples from the individual classes, one vs. rest models were additionally trained. These models are especially interesting for routine use since a sample is usually labeled with a country of origin and can therefore be checked in a one vs. rest model to confirm the declared country of origin of the sample.

The accuracy and sensitivity of the one vs. rest models range from 87.7 to 96.7% and 74.9% and 92%, respectively (Table 1). The highest accuracy occurs in the Peru vs. rest model with 96,7%, whereas the Spain-, Germany-, Ukraine-, and Hungary vs. rest models show sufficient accuracies with 95.5, 92.8, 94.5, and 90.5%. Only France- and USA vs. rest models yield in lower accuracies of 87.8 and 88.3%. The recognition of the country increases in the one vs. rest models in contrast to the seven-class model for the countries Spain from 81.2 to 89.0% and for Ukraine from 73.7 to 81.1%, which shows that a one vs. rest approach can make the model more accurate.

Subsequently, the features in the seven-class model were ranked based on their importance determined by the mean decrease in accuracy (Table A2). The most important features could be assigned to FAs, MGs, DGs and TGs based on their mass and database research. Additional MS^2 measurements were performed to validate the database search (see section 3.2).

3.2. Identification of relevant markers

To identify the most relevant marker signals, fragmentation patterns of similar reference substances were first analyzed. This was particularly important because there is little literature on the fragmentation of lipophilic molecules with DART-MS^2 and thus no database search could be performed. The lipid standard substances listed in Table 2 were measured and fragmented. The fragmentations in the MS^1 (due to in-

Table 1
Accuracy and sensitivity in percent of the classifications in one vs. rest models for each country using the random forest algorithm with 10 times repeated 10-fold cross validation. The two class models consisted of the respective country of Spain (Esp), France (Fra), Germany (Ger), Peru (Per), Ukraine (Ukr), Hungary (Hun), and the USA and the totality of the other samples as rest.

	Accuracy [%]	Sensitivity per country [%]
Esp	95.6	89.0
Fra	87.7	74.9
Ger	92.8	92.3
Per	96.7	92.0
Ukr	94.5	81.1
Hun	90.5	84.4
USA	88.3	81.3

Table 2
Ionization and fragmentation behavior of the different lipid standards obtained with the DART-qToF platform.

Tentative compound	Proposed formula	Adduct	Modus	Theoretical m/z [Da]	Measured m/z [Da]	Relevant fragment
MG 18:1	$C_{21}H_{40}O_4$	$[M+H]^+$	MS^1	357.3005	357.3005	
		$[M-H–H_2O]^+$	MS^1	339.2894	339.2891	
		$[RCO]^+$	MS^2	265.2532	265.2529	FA 18:1
		$[RCO–H_2O]^-$	MS^2	247.2426	247.2422	FA 18:1
DG 32:0	$C_{35}H_{68}O_5$	$[M+H]^+$	MS^1	569.5145	569.5141	
		$[M-H–H_2O]^+$	MS^1	551.5039	551.5051	
		$[M-NH_4]^+$	MS^1	586.5410	586.5418	
		$[RCO]^+$	MS^2	239.2375	239.2376	FA 16:0
		$[M+H]^+$	MS^2	313.2737	313.2747	MG 16:1
TG 48:0	$C_{51}H_{98}O_6$	$[M-NH_4]^+$	MS^1	824.7707	824.7734	
		$[M-H–H_2O]^+$	MS^1	551.5034	551.5056	
		$[M-H–H_2O]^+$	MS^2	551.5034	551.5058	DG 32:0
TG 54:6	$C_{57}H_{98}O_6$	$[M+H]^+$	MS^1	879.7442	879.7441	
		$[M-NH_4]^+$	MS^1	896.7707	896.7706	
		$[M-H–H_2O]^+$	MS^1	599.5034	599.5035	
		$[M-H–H_2O]^+$	MS^2	599.5034	599.5037	DG 36:4
		$[M]^+$	MS^2	616.5061	616.5015	
		$[RCO]^+$	MS^2	263.2369	263.2365	FA 18:2
PE 36:0	$C_{41}H_{82}NO_8P$	$[M+H]^+$	MS^1	748.5856	748.5859	
		$[M-H–H_2O]^+$	MS^1	607.5660	607.5667	DG 36:0
		$[M-H–H_2O]^+$	MS^2	607.5660	607.5675	DG 36:0
		$[M-H–H_2O]^+$	MS^2	341.3050	341.3063	MG 18:0
PC 36:2	$C_{44}H_{84}NO_8P$	$[M+H]^+$	MS^1	786.6007	786.6015	
		$[M-H–H_2O]^+$	MS^1	603.5347	603.5367	DG 36:2
		$[M-H–H_2O]^+$	MS^2	603.5347	603.5349	DG 36:2
		$[M-H–H_2O]^+$	MS^2	339.2900	339.2888	MG 18:1

Abbreviations: FA, fatty acid; MG, monoacylglycerol; DG, diacylglycerol; TG, triacylglycerol; PE, glycerophosphoethanolamine; PC, glycerophosphocholine.

source fragmentations) and MS^2 spectra as well as all ionization possibilities were taken into account and are shown in the supplementary information's in Figs. A4–A15.

MG (18:1) was chosen as an example of the MGs, from which the fragmentation patterns of other MGs should be inferred. The molecule ionized as an $[M+H]^+$-adduct and as a $[M+H–H_2O]^+$-adduct in MS^1 mode. Furthermore, this compound was already fragmented in the MS^1 mode, resulting, in a lower signal at *m/z* 265.2526 of FA (18:1) due to the elimination of glycerol. The MG (18:1) as $[M+H]^+$-adduct was chosen for fragmentation. As well as in the MS^1 spectra the MS^2 spectrum showed a fragmentation into the corresponding FA under the loss of a fragment of *m/z* 74.0362, which indicates the fragmentation of the ester and the loss of a glycerol molecule. The FA (18:1) occurred as $[RCO]^-$ and $[RCO–H_2O]^-$ fragments with an *m/z* of 265.2529 and 247.2422. The existence of $[RCO]^-$-adducts in the ionization of fatty acids by DART sources has already been described in the literature (Cajka et al., 2013; Vaclavik et al., 2009).

In MS^1 mode, DG (32:0) ionized like the MG as $[M+H]^+$ and $[M+H–H_2O]^+$-adduct, with the ammonium adduct also being detectable. No additional ammonia was used during the analysis, which some groups prefer, by placing an ammonia bottle relatively close to the DART source area (Hajslova et al., 2011). However, apparently the ammonia present in the laboratory air was sufficient for the formation of ammonium adducts (Gross, 2014). Fragments of the DG (32:0) consisted of the corresponding $[M+H]^+$-adduct of MG (16:0) with an *m/z* of 313.2747 and palmitic acid as $[RCO]^+$-adduct with an *m/z* of 239.2376. As already observed for MG (18:1), the MG occurred as $[M+H]^+$-adduct and was further fragmented in the fatty acid as $[RCO]^+$-fragment.

To cover the broad spectrum of TGs, TG (48:0) and TG (54:6) were chosen as examples for saturated and unsaturated TGs. In MS^1 mode the main signal was for both TGs, the ammonium adduct, while in the case of TG (48:0) the $[M+H–H_2O]^+$-ion could also be observed. The $[M+H]^+$-adduct was only present for the unsaturated TG (54:6). Another similarity between the TGs as well as the DGs and MGs was the low level of fragmentation in MS^1 mode. Thus, in the MS^1 spectra of TG (48:0) and TG (54:6), the fragments of the corresponding DGs were present as $[M+H–H_2O]^+$-ions with an *m/z* of 551.5056 and 599.5035. This pattern was repeated in the MS^2 spectra. In both measured TG standards, the fragments were observed as $[M+H–H_2O]^+$-ions of the DGs with almost the same *m/z* as in the MS^1 spectra. Other fragments of TG (54:6) were DG (36:6) as $[M^+]$-ion at *m/z* 616.5051 and with a low signal intensity the FA (18:1) with an *m/z* of 263.2365 as $[RCO]^+$-ion. In contrast the $[RCO]^+$-fragment of FA (16:0) was not present when TG (48:0) was measured. Using ESI sources, TGs are ionized rather than $[M+NH_4]^+$ and $[M+Na]^+$ ions, but fragmentation in the DGs and MGs follows the same pattern. The fragmentation of the TGs by means of DART sources has been described in the literature and has already been used for marker identification (Frick & Weyermann, 2019; Schütz et al., 2021). The fragmentation by DART-MS^2 leads to the formation of further adducts such as $[M+H–H_2O]^+$ and $[M]^+$-ions, besides the $[M+NH_4]^+$-adducts already known from an ionization using ESI. In addition no $[M+Na]^+$-ions were observed by means of DART ionization.

The $[M+H]^+$-ions of phospholipids PE (36:0) and PC (36:2) showed only very low signal intensities in MS^1 mode at an *m/z* of 748.5859 and 786.6015. This is a difference from ionization via ESI, where the $[M+H]^+$-ion is the main ion for phospholipids (Godzien et al., 2015). The low signal intensity of the $[M+H]^+$-ion of PE (36:0) in MS^1 mode resulted from in-source fragmentation of the ester bond between glycerol and the phosphate. This led to the formation of the DG (36:0) with elimination of the phosphoethanolamine group, which is indicated by neutral loss of *m/z* 141 (Zhou et al., 2012). The remaining DG (36:0) ionized as $[M+H–H_2O]^+$-ion and was much more intense than the original PE (36:0). In MS^2 mode, as in MS^1 mode, PEs fragmented into the corresponding DG as $[M+H–H_2O]^+$-ions like in MS^1 mode. In addition, further fragmentation occurred from the DG (36:0) into the MG (18:0) as $[M+H–H_2O]^+$-ion. PC (36:2) followed the same fragmentation patterns like the PE (36:0). In MS^2 mode the fragmentation occurred in the $[M+H–H_2O]^+$-ion of the DG (36:2) with an *m/z* of 603.5349. The neutral loss of *m/z* 184, which is typical for PCs, indicates the fragmentation of the phosphate ester and the elimination of the phosphocholine group (Colsch et al., 2017). It follows that glycerol-phospholipids could not be ionized well with the chosen DART source, but that the fragments were also observed in the MS^1 spectrum. However, it was possible to identify the PE (36:0) and PC (36:2) via the low abundant $[M+H]^+$-ion. The ability to ionize PEs and PCs without fragmentation using an ESI source is an advantage of this ionization

technique. Nevertheless, the fragmented DGs from the PEs and PCs can be observed in DART-spectra and reveal structural information.

Identification of important markers from the grain maize samples was carried out according to the workflow developed with the standards. In general, the MS^2 spectra of the samples were more diffuse than those of the standards and more ions appeared. As an example, the MS^2 spectrum of the signal from m/z 847.7843 is shown in Fig. 4. For identification, the fragments of the FAs were considered, whereby three m/z-values could be assigned to the $[RCO]^+$ fragments of the FA (16:0, m/z 239.3454), (FA 18:1, m/z 265.2519) and FA (18:2, m/z 263.2361). The spectrum also shows the DGs as $[M + H–H_2O]^+$-ions, which are the most intense fragments of the spectrum and can also be used to identify non-ionized FAs.

In this way, an attempt was made to identify as many relevant markers as possible. The m/z ratios of relevant markers were fragmented in descending order of importance. Thus, the structures of seven signals were determined and 3 TG as well as four DGs could be identified. They are listed together with their importance values in Table 3 and the corresponding spectra can be found in Figs. A16–A22. One disadvantage that reduced the number of identifiable markers was the lower ionization rate and signal intensity of lipids by the DART source compared to ESI source used in the previous study (Schütz et al., 2021). Therefore, only the seven relevant markers with high intensities could be fragmented.

The boxplots of the identified markers in Fig. 5 show clearly different signal intensities between DG (34:1), DG (36:3) and DG (36:2) which show particularly high signal intensities in the sample from Peru and thus enabled a differentiation from the other countries, which explains the high classification result of 95.7% of data evaluation using RF algorithm (see Fig. 3).

TG (54:5) could be detected with comparatively high signal intensities in the samples from Germany, France and Hungary, which is why this compound was well suited to separate these samples from the other countries. This effect is probably based on the fact that these Central European countries have a similar continental climate (Saini et al., 2019). The result of the RF classification also implies similarities between these three countries (Fig. 3). Furthermore, the German and French samples show similar marker concentrations for the DGs (34:1), (36:3) and (36:2) as well as TG (50:1), (52:3) and FA (18:1). The French samples come from the department of Alsace, which borders directly on Germany. This proximity of the countries can also be seen in the comparable signal intensities of the markers DG (36:3) and TG (52:3). However, it was possible to distinguish between the two regions or countries with the other TGs, especially with TG (54:5).

Other DGs like DG (34:2) showed comparatively high signal intensities in the Spanish samples. Despite the many varieties used and the resulting high varietal diversity of this sample class, there was not a wide distribution of the signal intensities in the individual samples, which shows the low influence of the varieties on the model. In general, DGs are associated with stress in plant growth and might thus be linked to the geographical origin because temperature and precipitation vary depending on the location. Presumably, their suitability as marker substances for distinguishing geographical origins is also derived from this biological property (Eichmann & Lass, 2015; Upchurch, 2008). TGs serve as storage lipids and can also be affected by stress (Upchurch, 2008). The TGs (50:1) and (52:3) occurred in high concentrations in the samples from Germany, France, Hungary and partly Peru. These TGs and the corresponding FAs were probably also conspicuous due to plant stress parameters (Upchurch, 2008). Furthermore, the samples from the USA had a large variance in the relative signal intensities for TG (50:1). This circumstance can probably attribute to the fact that the US American samples came from different states, some of which were very far away from each other. Therefore, the contents of samples from the USA vary more than in countries with a smaller area. Despite greater fluctuations, TG (50:1), TG (54:5) are suitable for separating these countries.

DART-MS instruments offer the advantage ionizing FAs in positive mode, with high signal intensities (Cajka et al., 2013). The difference in signal intensities of an FA is shown in Fig. 5 based on the signal m/z 263.2588. This m/z was assigned to oleic acid as $[RCO]^+$-fragment, as it has already been described in the literature (Cajka et al., 2013; Vaclavik et al., 2009). The MS^2 spectrum of MG (18:1) also shows the same mass, which confirmed the assignment. High signal intensities of FA (18:1) could be detected especially in samples from Ukraine and Spain, indicating that FAs are also suitable for origin determination. FAs have already been described in the literature as markers for the geographical origin of grain maize and the concentration is influenced by environmental stress (Chebrolu et al., 2016; Harrigan et al., 2007; Hou et al., 2016).

3.3. Comparison of the DART-MS and the LC-ESI-MS measurements

In previous work, a classification of maize has already been carried out using almost the same data set measured with LC-ESI-MS (Schütz et al., 2021). In addition, a part of the samples was analyzed by FT-NIR. In this study, 101 samples from Ukraine, Peru, Slovakia, Spain and the USA were classified and were assigned with an accuracy of 95%. The accuracy of the individual classes ranged from 66.3% for Slovakia to 100.0% for Peru and the USA. In between were the Ukraine with an accuracy of 98.1% and Spain with 98.9%. However, comparability with the DART-MS dataset is limited, as fewer samples and countries were analyzed.

Furthermore, isotope ratio mass spectrometry (IRMS) is another common technique to detect the geographical origin. When harvesting grain maize, care is taken to ensure that the residual moisture content is as low as possible, or this is adjusted by drying immediately after

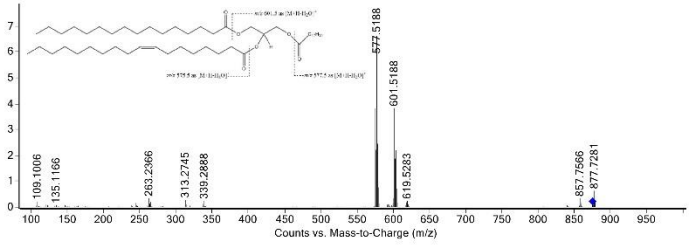


Fig. 4. MS^2 production scan of a precursor with m/z 874.7843, fragmented with a collision energy of 20 eV using the DART-qToF platform. Based on the fragments obtained, the signal TG (52:3) could be assigned.

Table 3
Important signals in grain maize to determine the geographical origin. The signals were identified with the held of MS^2 spectra and the DART-qToF platform. The relevant fragments of the DGs and TGs are listed with their mean decrease accuracy calculated with random forest algorithm.

Tentative compound	Proposed formula	Adduct	Theoretical *m/z*	Measured *m/z*	Delta *m/z*	Relevant Fragments	Mean decrease accuracy
DG 34:2	$C_{37}H_{68}O_5$	$[M + H–H_2O]^+$	575.5034	575.5014	0.0020	239.2360; 263.2366	4.0
DG 34:1	$C_{37}H_{70}O_5$	$[M + H–H_2O]^+$	577.5190	577.5221	0.0031	239.2361; 265.2523	4.3
DG 36:3	$C_{39}H_{70}O_5$	$[M + H–H_2O]^+$	601.5190	601.5333	0.0143	263.2364; 265.2523	3.3
DG 36:2	$C_{39}H_{72}O_5$	$[M + H–H_2O]^+$	603.5347	603.5312	0.0035	265.2534	3.8
TG 50:1	$C_{53}H_{100}O_6$	$[M + NH_4]^+$	850.7858	850.7826	0.0032	313.2716; 577.5157	1.7
TG 52:3	$C_{55}H_{100}O_6$	$[M + NH_4]^+$	874.7858	874.7843	0.0015	263.2366; 313.2745; 577.5188	1.1
TG 54:5	$C_{57}H_{100}O_6$	$[M + NH_4]^+$	898.7858	898.7838	0.0020	263.2359; 339.2886; 601.5186	2.1

harvesting (Huang et al., 2012). Due to the loss of water during drying, the isotope ratios are shifted, as lighter isotopes evaporate faster than heavier ones, IRMS is only suitable for the analysis of storable maize to a limited extent (Cappa et al., 2003).

Therefore, LC-MS was chosen as the benchmark method, as it was applicable to grain maize and was also performed with a large data set from seven countries. The sample extraction using DART-MS was performed with the same modified Bligh and Dyer method, so that comparability with the LC-MS measurements is ensured (Bligh and Dyer, 1959).

A major advantage of measurement using a DART-MS instrument is the short analysis time of just a few seconds per sample. The LC-ESI-MS method with an analysis time of 30 min per sample resulted in measurement times of several days for the complete data set. With the DART-MS platform, the measurements could be completed within one day, which saved time and staff as well as led to results on the same day. Since the extracts were diluted when measuring the samples using LC-ESI-MS and this pipetting step is not necessary with DART-MS, analysis time and solvent consumption could be reduced by omitting this dilution step. In addition, the reduction in measurement time led to more stable results in mass accuracy since all measurements could be performed shortly after calibration and a gradual loss of ToF calibration could not occur. This means that LC-ESI-MS instruments cannot be used to measure several thousand samples without additional calibrations, references or standards. In contrast, it would be possible to make comparable measurements of thousands of samples using DART-MS, which is due to the fact that 200 or more samples per day can be measured.

In turn, what makes LC-ESI-MS analysis easier for the end user is the availability of proprietary software, e.g. the Agilent MassHunter software package, which was specifically designed for LC-ESI-MS analysis. In contrast, there is little end-user software for DART-MS platforms to date. However, it is expected that specialized software will be developed as DART-MS platforms become more widely used in research and routine applications. In addition, the workflow developed in this work shows how spectra can be analyzed with open-source software, filling a gap in existing research for the end user.

As the sustainability or environmental compatibility of the developed methods has become increasingly important in recent years, the reduction of environmental impact plays a decisive role. The measurement by LC uses about 10 mL of solvent per sample, while the amount of solvent used in the extraction method is comparatively small at 1 mL. Here, the DART-MS approach offers a great advantage, as without chromatography about 90% of the solvent consumption is eliminated. Apart from being environmentally friendly, analysis using DART-MS is less expensive than using LC-MS. A cost comparison of the two methods is shown in Table A4 in Appendix A. The biggest cost factor for an analysis using DART-MS is the price of the DIP-it sampler at 1.90 € per piece. In contrast, the helium consumption per analysis is very low, since only about 10 s of helium flow is required in routine analysis. Besides instrumentation and column wear, solvent volumes have the greatest influence on the total cost of an analysis with an LC-ESI-MS instrument. The price of acetonitrile for LC-ESI-MS analysis is approximately 1.84 € per sample, accounting for 50% of the cost. This effect is particularly noticeable because the price of solvents like methanol and acetonitrile has increased fivefold in the last two years (Merck Germany, 2023). In contrast, there is no difference between the methods with regard to the materials used for the extraction, such as methanol, chloroform, reaction tubes, steel balls and pipette tips. With such an increase in the cost of the solvents, DART-MS measurements, with a price of 2.46 € per sample, are about a third cheaper than analysis by LC-ESI-MS with 3.41 €.

DART-MS has not only some advantages over LC-MS, but also compared to direct infusion (DI). DI has the disadvantage that contamination of the MS equipment occurs more frequently, as the extracts are infused directly into the ESI source (Gross, 2014). To avoid contamination, several blank measurements are usually performed between the samples, which in turn increase the analysis time. The average analysis time is 2–10 min per sample (Lísa et al., 2017; McDougall et al., 2008). Often three blanks are measured after one analysis, which increases the analysis time from 2 to 8 min (Lísa et al., 2017). In addition, solvent consumption during injection and blanks must be considered when using LC systems with a constant flow for sample injection. Furthermore, with DART-MS it is possible to obtain other ions and fragments of the analytes even in MS^1 mode due to the slightly harder ionization conditions. Due to the higher fragmentation rate of the analytes, conclusions about the structure of the molecules can be drawn simultaneously with the measurement, similar to electron ionization (EI).

Based on the LC-ESI-MS measurements using the RF algorithm, an accuracy of 98.5% was achieved with 1022 features. A variable reduction to routine 20 markers led to a final result with an accuracy of 90.5% (Schütz et al., 2021). In contrast, the DART-MS approach did not quite reach the correctness of the LC-ESI-MS analysis with an accuracy of 84.4%, but offers the advantages discussed above. Comparing the accuracy of the one vs. rest models, the LC-ESI-MS approach ranged from 89.3 to 99.2%, while the DART-MS approach is in a similar range between 88.3 and 96.7%. Thus LC-ESI-MS and DART-MS systems achieve comparable results in one vs. rest models and that the verification of a sample in these models is possible with both methods.

When comparing the identified markers, there were overlaps between the LC-ESI-MS and the DART-MS platform. The lipids DG (34:2), DG (34:1) and DG (36:3) were selected and identified as markers with both methods, highlighting the importance of these markers. In addition, four molecules, DG (36:2), TG (50:1), (52:3) and (54:5), which were not previously known as markers, were identified using $DART\text{-}MS^2$ measurements.

4. Conclusion

In this proof of concept study, a new workflow for determining the origin of grain maize using a non-polar extraction in combination with a DART-MS approach was presented.

This cost-effective and environmentally friendly mass spectrometry-based method offers the advantage of short analysis times to obtain a complete fingerprint of food and feed samples. For the first time, the capability of DIP-it samplers as a high throughput method was demonstrated by the analysis of 200 samples, which could be classified by a RF data evaluation. With the described method, an accuracy of 84.4% could be obtained, demonstrating sufficient classification based on the

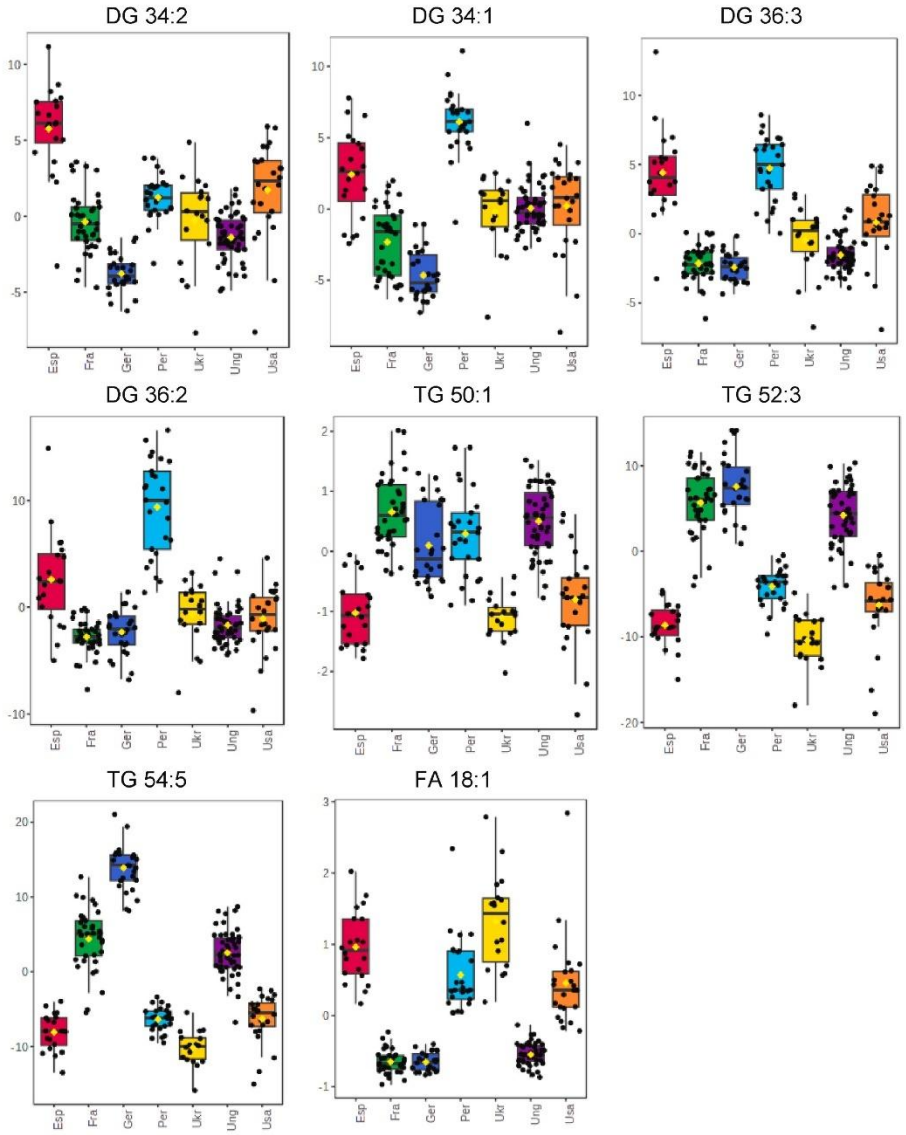


Fig. 5. Normalized concentration ratios of important marker compounds showing a dependence on the geographic origin of the maize samples.

measured fingerprints and the RF algorithm.

The evaluation strategy developed with msConvert, MZmine 2 and R is based purely on open source software and offers the advantage of being freely accessible and well comprehensible. In addition, RF algorithm was shown to be suitable for the evaluation of data with no chromatographic separation like the chosen DART-MS platform.

In further experiments, markers for the classification of the geographical origin of maize were identified by means of DART-MS^2 measurements and, in addition to new markers, already known markers were verified. The lipid markers belong to the substance classes of MGs, DGs and TGs. By means of the described workflow and a continuously evolving database, a contribution can be made to develop further strategies for authentication of food.

Funding

This study was performed within the project "Food Profiling – Development of analytical tools for the experimental verification of the origin and identity of food". This project (funding reference number: 2,816,500,914) was supported by means of the Federal Ministry of Food

and Agriculture (German: Bundesministerium für Ernährung und Landwirtschaft, BMEL) by a decision of the German parliament (German: Bundestag). Project support is provided by the Federal Institute for Agriculture and Food (German: Bundesanstalt für Landwirtschaft und Ernährung, BLE) within the scope of the program for promoting innovation.

CRediT authorship contribution statement

Felix Schmauder: Writing – original draft, Validation, Investigation, Formal analysis. **David Schütz:** Writing – review & editing, Validation, Data curation. **Marina Creydt:** Writing – review & editing, Supervision, Methodology. **Janet Riedl:** Writing – review & editing, Conceptualization. **Markus Fischer:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The authors thank LAVES Stade and Oldenburg, Silvie Kramer from Cargill Germany, the Universidad de Trujillo in Peru, Antoni López Querol from Institute of Agrifood Research and Technology in Lleida and P. Cotty from USDA Agricultural Research Service for providing samples. We further want to thank the organizers of the varietal trials in the German regions Bavaria (Dr. Joachim Eder, LfL Bayern, Institut für Pflanzenbau und Pflanzenzüchtung, Freising), Brandenburg (Dr. Gert Barthelmes, LELF Brandenburg, Teltow) and North Rhine-Westphalia (Norbert Erhardt, Landwirtschaftskammer NRW, Münster) for providing samples from German grain maize varietal trials.

We thank Prof. Dr. Stephan Seifert and Marie-Sophie Müller for proof reading and expertise as well as Rasmus Bøg Hansen for his expertise.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2024.110489.

References

Achten, E., Schütz, D., Fischer, M., Fauhl-Hassek, C., Riedl, J., & Horn, B. (2019). Classification of grain maize (Zea mays L.) from different geographical origins with FTIR spectroscopy—a suitable analytical tool for feed authentication? *Food Analytical Methods, 12*, 2172–2184.

Angelis, E. de, Pilolli, R., Bejjani, A., Guagnano, R., Garino, C., Arlorio, M., & Monaci, L. (2021). Optimization of an untargeted DART-HRMS method envisaging identification of potential markers for saffron authenticity assessment. *Foods, 10*(6), 1238.

Arora, M., Zambrzycki, S. C., Levy, J. M., Esper, A., Frediani, J. K., Quave, C. L., Fernández, F. M., & Kamaleswaran, R. (2022). Machine learning approaches to identify discriminative signatures of volatile organic compounds (VOCs) from bacteria and fungi using SPME-DART-MS. *Metabolites, 12*(3), 232.

Aszyk, J., Byliński, H., Namieśnik, J., & Kot-Wasik, A. (2018). Main strategies, analytical trends and challenges in LC-MS and ambient mass spectrometry–based metabolomics. *TrAC, Trends in Analytical Chemistry, 108*, 278–295.

Avula, B., Smillie, T. J., Wang, Y.-H., Zweigenbaum, J., & Khan, I. A. (2015). Authentication of true cinnamon (Cinnamon verum) utilising direct analysis in real time (DART)-QToF-MS. *Food Additives & Contaminants: Part A, 32*(1), 1–8.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*, 197–227.

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology, 37*(8), 911–917.

Cajka, T., Danhelova, H., Zachariasova, M., Riddellova, K., & Hajslova, J. (2013). Application of direct analysis in real time ionization–mass spectrometry (DART–MS) in chicken meat metabolomics aiming at the retrospective control of feed fraud. *Metabolomics, 9*, 545–557.

Cappa, C. D., Hendricks, M. B., DePaolo, D. J., & Cohen, R. C. (2003). Isotopic fractionation of water during evaporation. *Journal of Geophysical Research: Atmospheres, 108*(D16).

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M.-Y., Paulse, C., Creasy, D., … Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology, 30*(10), 918–920. https://doi.org/10.1038/nbt.2377

Chebrolu, K. K., Fritschi, F. B., Ye, S., Krishnan, H. B., Smith, J. R., & Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics, 12*, 1–14.

Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data* (Vol. 110). Berkeley: University of California, 1-12.

Cody, R. B., Laramée, J. A., Nilles, J. M., & Durst, H. D. (2005). Direct analysis in real time (DART) mass spectrometry. *Japan Electron Optics Laboratory News, 40*(1), 8–12.

Colsch, B., Fenaille, F., Warnet, A., Junot, C., & Tabet, J.-C. (2017). Mechanisms governing the fragmentation of glycerophospholipids containing choline and ethanolamine polar head groups. *European Journal of Mass Spectrometry, 23*(6), 427–444. https://doi.org/10.1177/1469066717731668

Creydt, M., Arndt, M., Hudzik, D., & Fischer, M. (2018). Plant metabolomics: Evaluation of different extraction parameters for nontargeted UPLC-ESI-QTOF-mass spectrometry at the example of white Asparagus officinalis. *Journal of Agricultural and Food Chemistry, 66*(48), 12876–12887.

Creydt, M., & Fischer, M. (2022). Food authentication: Truffle species classification by non-targeted lipidomics analyses using mass spectrometry assisted by ion mobility separation. *Molecular Omics, 18*(7), 616–626.

Creydt, M., Hudzik, D., Rurik, M., Kohlbacher, O., & Fischer, M. (2018). Food authentication: Small-molecule profiling as a tool for the geographic discrimination of German white asparagus. *Journal of Agricultural and Food Chemistry, 66*(50), 13328–13339.

Deutscher Wetterdienst. (2023a). Climate diagramm budapest. https://www.dwd.de/DWD/klima/beratung/ak/ak_128430_di.pdf.

Deutscher Wetterdienst. (2023b). Climate diagramm Madrid. https://www.dwd.de/DWD/klima/beratung/ak/ak_082220_di.pdf.

Deutscher Wetterdienst. (2023c). Climate diagramm san antonio, Texas. https://www.dwd.de/DWD/klima/beratung/ak/ak_722530_di.pdf.

Deutscher Wetterdienst. (2023d). Climate diagramm straßburg. https://www.dwd.de/DWD/klima/beratung/ak/ak_071900_di.pdf.

Eichmann, T. O., & Lass, A. (2015). DAG tales: The multiple faces of diacylglycerol—stereochemistry, metabolism, and signaling. *Cellular and Molecular Life Sciences, 72*, 3931–3952.

Fahy, E., Sud, M., Cotter, D., & Subramaniam, S. (2007). LIPID MAPS online tools for lipid research. *Nucleic Acids Research, 35*(suppl_2), W606–W612.

Frick, A. A., & Weyermann, C. (2019). An untargeted lipidomic approach for qualitative determination of latent fingermark glycerides using UPLC-IMS-QToF-MS E. *Analyst, 144*(11), 3590–3600.

Godzien, J., Ciborowski, M., Martínez-Alcázar, M. P., Samczuk, P., Kretowski, A., & Barbas, C. (2015). Rapid and reliable identification of phospholipids for untargeted Metabolomics with LC–ESI–QTOF–MS/MS. *Journal of Proteome Research, 14*(8), 3204–3216.

Golbraikh, A., & Tropsha, A. (2002). Beware of q2. *Journal of Molecular Graphics and Modelling, 20*(4), 269–276.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis–a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta, 879*, 10–23.

Gross, J. H. (2014). Direct analysis in real time—a critical review on DART-MS. *Analytical and Bioanalytical Chemistry, 406*, 63–80.

Hajslova, J., Cajka, T., & Vaclavik, L. (2011). Challenging applications offered by direct analysis in real time (DART) in food-quality and safety analysis. *TrAC, Trends in Analytical Chemistry, 30*(2), 204–218.

Harrigan, G. G., Stork, L. G., Riordan, S. G., Reynolds, T. L., Ridley, W. P., Masucci, J. D., MacIsaac, S., Halls, S. C., Orth, R., & Smith, R. G. (2007). Impact of genetics and environment on nutritional and metabolite components of maize grain. *Journal of Agricultural and Food Chemistry, 55*(15), 6177–6185.

Hou, Q., Ufer, G., & Bartels, D. (2016). Lipid signalling in plant responses to abiotic stress. *Plant, Cell and Environment, 39*(5), 1029–1048.

Hrbek, V., Rektorisova, M., Chmelarova, H., Ovesna, J., & Hajslova, J. (2018). Authenticity assessment of garlic using a metabolomic approach based on high resolution mass spectrometry. *Journal of Food Composition and Analysis, 67*, 19–28.

Huang, H., Faulkner, D. B., Berger, L. L., & Eckhoff, S. R. (2012). Harvest date influence on dry matter yield and moisture of corn and stover. *Transactions of the ASABE, 55*(2), 593–598.

Khodaei, D., Javanmardi, F., & Khaneghah, A. M. (2021). The global overview of the occurrence of mycotoxins in cereals: A three-year survey. *Current Opinion in Food Science, 39*, 36–42.

Kim, H. J., Baek, W. S., & Jang, Y. P. (2011). Identification of ambiguous cubeb fruit by DART-MS-based fingerprinting combined with principal component analysis. *Food Chemistry, 129*(3), 1305–1310.

Kim, H. J., Seo, Y. T., Park, S., Jeong, S. H., Kim, M. K., & Jang, Y. P. (2015). DART–TOF–MS based metabolomics study for the discrimination analysis of geographical origin of Angelica gigas roots collected from Korea and China. *Metabolomics, 11*, 64–70.

Lian, R., Wu, Z., Lv, X., Rao, Y., Li, H., Li, J., Wang, R., Ni, C., & Zhang, Y. (2017). Rapid screening of abused drugs by direct analysis in real time (DART) coupled to time-of-flight mass spectrometry (TOF-MS) combined with ion mobility spectrometry (IMS). *Forensic Science International, 279*, 268–280.

Lísa, M., Cífková, E., Khalikova, M., Ovčačíková, M., & Holčapek, M. (2017). Lipidomic analysis of biological samples: Comparison of liquid chromatography, supercritical fluid chromatography and direct infusion mass spectrometry methods. *Journal of Chromatography A, 1525*, 96–108.

McDougall, G., Martinussen, I., & Stewart, D. (2008). Towards fruitful metabolomics: High throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry. *Journal of Chromatography B, 871*(2), 362–369.

Merck Germany. (2023). Acetonitrile hypergrade for LC-MS LiChrosolv®. https://www.merckmillipore.com/DE/de/product/Acetonitrile,MDA_CHEM-100029.

Novotná, H., Kmiecik, O., Gałązka, M., Krtková, V., Hurajová, A., Schulzová, V., Hallmann, E., Rembialkowska, E., & Hajšlová, J. (2012). Metabolomic fingerprinting employing DART-TOFMS for authentication of tomatoes and peppers from organic and conventional farming. *Food Additives & Contaminants: Part A, 29*(9), 1335–1346.

Núñez, O., Moyano, E., & Galceran, M. T. (2004). High mass accuracy in-source collision-induced dissociation tandem mass spectrometry and multi-step mass spectrometry as complementary tools for fragmentation studies of quaternary ammonium herbicides. *Journal of Mass Spectrometry, 39*(8), 873–883.

Pang, Z., Chong, J., Zhou, G., Lima Morais, D. A. de, Le, C., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research, 49*(W1), W388–W396.

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics, 11*(1), 395. https://doi.org/10.1186/1471-2105-11-395

Razali, M. T. A., Zainal, Z. A., Maulidiani, M., Shaari, K., Zamri, Z., Mohd Idrus, M. Z., Khatib, A., Abas, F., Ling, Y. S., & Rui, L. L. (2018). Classification of raw stingless bee honeys by bee species origins using the NMR-and LC-MS-based metabolomics approach. *Molecules, 23*(9), 2160.

Saini, M. K., Capalash, N., Kaur, C., & Singh, S. P. (2019). Targeted metabolic profiling indicates differences in primary and secondary metabolites in Kinnow Mandarin (C. nobilis× C. deliciosa) from different climatic conditions. *Journal of Food Composition and Analysis, 83*, Article 103278.

Schütz, D., Achten, E., Creydt, M., Riedl, J., & Fischer, M. (2021). Non-targeted LC-MS metabolomics approach towards an authentication of the geographical origin of grain maize (Zea mays L.) samples. *Foods, 10*(9), 2160.

Schütz, D., Riedl, J., Achten, E., & Fischer, M. (2022). Fourier-transform near-infrared spectroscopy as a fast screening tool for the verification of the geographical origin of grain maize (Zea mays L.). *Food Control, 136*, Article 108892.

United States Department of Agriculture. (2023). Grain: World markets and trade: 2023/24 grain production exceeds consumption. https://apps.fas.usda.gov/psdonline/circulars/grain.pdf.

Upchurch, R. G. (2008). Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnology Letters, 30*, 967–977.

Vaclavik, L., Cajka, T., Hrbek, V., & Hajslova, J. (2009). Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Analytica Chimica Acta, 645*(1–2), 56–63.

Wang, K., Xu, L., Wang, X., Chen, A., & Xu, Z. (2021). Discrimination of beef from different origins based on lipidomics: A comparison study of DART-QTOF and LC-ESI-QTOF. *LWT, 149*, Article 111838.

Westerhuis, J. A., van Velzen, E. J. J., Hoefsloot, H. C. J., & Smilde, A. K. (2008). Discriminant Q 2 (DQ 2) for improved discrimination in PLSDA models. *Metabolomics, 4*, 293–296.

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., & Lee, B. L. (2022). Hmdb 5.0: The human metabolome database for 2022. *Nucleic Acids Research, 50*(D1), D622–D631.

Xu, Y., Zhang, D.-Y., Meng, X.-Y., Liu, X., Sheng, S., Wu, G.-H., Wang, J., & Wu, F.-A. (2017). Generic DART-MS platform for monitoring the on-demand continuous-flow production of pharmaceuticals: Advancing the quantitative protocol for caffeates in microfluidic biocatalysis. *Journal of Pharmaceutical and Biomedical Analysis, 137*, 243–251.

Zhou, L., Zhao, M., Ennahar, S., Bindler, F., & Marchioni, E. (2012). Determination of phosphatidylethanolamine molecular species in various food matrices by liquid chromatography–electrospray ionization–tandem mass spectrometry (LC–ESI–MS 2). *Analytical and Bioanalytical Chemistry, 403*, 291–300.

6.3 Herkunftsauthentifizierung von Haselnüssen (*Corylus avellana* L.) mittels DART-MS.

Der entwickelte Arbeitsablauf zeigte, dass es möglich ist, die Herkunft von Erdbeeren und Körnermais als Vertreter von wasser- und kohlenhydratreichen Lebensmitteln zu bestimmen. Nun wurde der Arbeitsablauf auf ein fettreiches Lebensmittel, die Haselnuss, übertragen, um herauszufinden, ob eine Herkunftsbestimmung mittels DART-MS-basierter Metabolomanalyse bei Haselnüssen möglich ist. In diesem Datensatz waren nur europäische Länder vorhanden, sodass eine Untersuchung eines engeren geographischen Raumes möglich war.

Die Analyse der Spektren von 178 Haselnussproben durch eine ungerichtete PCA zeigte erste Trends zur Trennung der Proben anhand ihrer Herkunft. Die anschließende Klassifizierung mittels *Random Forest* ergab eine Richtigkeit von 93,2%. Durch Identifizierung wichtiger *features* mittels DART-MS/MS Analytik konnten 16 Marker aus den Klassen der Di- und Triacylglyceride sowie der Tocopherole identifiziert werden. Unterschiede in den Konzentrationen der Marker konnten mittels eines Vergleiches bestätigt werden und wurden auf unterschiedliche Sättigungsgrade der Fettsäuren zurückgeführt. Diese Unterschiede in den Fettsäuremustern konnten mit der Anpassung der Pflanze an verschiedene Klimata assoziiert werden.

Hinweis zum Copyright:

This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

URL: https://www.sciencedirect.com/science/article/pii/S0308814624039153

DOI: https://doi.org/10.1016/j.foodchem.2024.142265

Food Chemistry 467 (2025) 142265

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (*Corylus avellana* L.)

Felix Schmauder, Marina Creydt, Markus Fischer *

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

ARTICLE INFO

Keywords:
Food fraud
Adulteration
Authenticity
Lipidomics
Mass spectrometry
Metabolomics
Chemometrics
Hazelnuts

ABSTRACT

Direct analysis in real time mass spectrometry (DART-MS) is a novel method for the authentication of food and feed that represents a serious alternative to established methods. This study aims to analyze hazelnuts from different origins and identify potential marker metabolites using a high-resolution DART-MS platform and a non-targeted metabolomics approach. To investigate the suitability of DART-MS for authenticating the origin of foods with a high fat content, 172 hazelnut samples from 5 countries were analyzed. Data evaluation using principal component analysis (PCA) and Random Forest-based classification led to an accuracy of 93.2 %, demonstrating the high valence of the DART-MS approach for verifying the origin of hazelnuts. In addition, 16 marker metabolites were identified and revealed the importance of di- and triacylglycerols for the authentication of hazelnuts. These results demonstrate the high suitability of DART-MS based analysis as a rapid, cost-effective, and environmentally friendly approach for food authentication.

1. Introduction

Hazelnuts are a high-priced raw material for a wide range of food products. Due to global trade, annual prices depend heavily on the success of the harvest. Depending on crop failure, import prices can fluctuate by up to 50 % compared to the previous year (Statistisches Bundesamt, 2024). There are also different prices, which depend on the region of cultivation and the variety. For example, the Italian hazelnuts"Nocciola del Piemonte", "Nocciola Romana" and "Nocciola di Giffoni" are subject to the EU seal "Protected Geographical Indication" (PGI)(European Comission, 2024). In 2022, Türkiye was the most important country for hazelnut production, with a yield of 765,000 tons. With less than 100,000 tons, other important countries were Italy, Azerbaijan, the USA, Chile and Georgia. In addition, France and Germany were relevant local producers for the European market (FAO, 2023).

This diverse global market with large price differences makes hazelnuts susceptible to counterfeiting. To prevent this food fraud, a wide range of analytical approaches have been tested to authenticate hazelnuts. In the context of high-tech methods, the geographical origin was determined using liquid chromatography-mass spectrometry (LC-MS) from a non-polar extract, which led to very promising results (97.4 % accuracy, 5 origins). Subsequently, a targeted method was developed based on the marker compounds that had been identified (Klockmann et al., 2016; Klockmann et al., 2017). In addition, gas chromatography ion mobility-mass spectrometry (GC-IMS-MS) was used to detect volatile organic compounds (VOC) and use them for the determination of geographical origin. This approach, taking the VOCs into account, led to an accuracy of 86 % considering three countries of origin (Sammarco, Bardin, et al., 2023). In addition, isotope ratio mass spectrometry (IRMS), which is the current standard method in routine analysis for determining the origin, was used to distinguish 96 hazelnut samples. Differences in isotope distribution were found between faraway countries, but neighboring countries in the Black Sea region showed very similar isotope distributions (Sammarco, Rossi, et al., 2023). A simple approach was used by determining the origin of hazelnuts in three Turkish regions based on the fatty acid profile. Gas chromatography combined with a flame ionization detector (GD-FID) was used for this purpose, and a sufficient result of 86.2 % was achieved (Tüfekci & Karataş, 2018). Using proton nuclear magnetic resonance (^{1}H NMR) and measuring polar extracts of 262 hazelnut samples sourced from 5 geographical origins, an accuracy of 96.0 % was achieved (Bachmann et al., 2018). By means of near-infrared spectroscopy (NIR), four origins could be distinguished with an accuracy of 89–100 % and five origins with an accuracy of 90.6 % (Sammarco, Dall'Asta, & Suman, 2023; Shakiba et al., 2022). Using a remarkably large data set of 1200 samples,

* Corresponding author.
E-mail address: markus.fischer@uni-hamburg.de (M. Fischer).

https://doi.org/10.1016/j.foodchem.2024.142265
Received 21 August 2024; Received in revised form 20 November 2024; Accepted 24 November 2024
Available online 26 November 2024
0308-8146/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

which was also investigated by NIR, an accuracy of 95.5 % was achieved. However, the dataset only contained three origins (Moscetti et al., 2015). Most high-tech methods such as GC–MS, LC-MS or NMR are characterized by long measurement times of 15–30 min per sample, which makes it difficult to use these methods when rapid analysis is required. Furthermore, the methods are associated with an increased use of chemicals in the form of solvents or derivatization reagents.

Advances in the speed and environmental friendliness of analyses can be achieved through the use of direct analysis in real time mass spectrometry (DART-MS) couplings (Schmauder et al., 2024). This comparatively new analysis technique makes it possible to introduce a sample directly into the DART ion source of a mass spectrometer without the need for chromatographic separation. The time required for an analysis is reduced enormously to less than one minute per sample. To implement this technically, the DART ion source uses excited helium atoms to ionize small molecules from sample extracts, which are then drawn into the vacuum of a MS. By analyzing as many metabolites as possible, conclusions can be drawn about the geographical origin of the samples.

Other studies on origin authentication using DART ionization have been carried out before, but only a limited number used a sample size in the three-digit range. These included a study of the origin of 188 *Angelica gigas* roots (Kim et al., 2015) and another study that focused on determining the origin of beef (Wang et al., 2021), as well as a study we published ourselves that aimed to verify the origin of grain maize samples (Schmauder et al., 2024). Further examples of food analysis using DART-MS can be found in the recent review paper by Wang (Wang, 2024).

In this study, hazelnuts were analyzed for the first time using a DART-MS-based lipidomics approach to examine a large number of samples of a fat-rich plant-based food. The data set of 172 samples covers the most important supplier countries as well as countries known for low-cost goods that are often used for adulteration. Other countries that may become more important for hazelnut production in the future due to climate change were also included (Ahmadov, 2024; An et al., 2020). Since the dataset includes many countries and samples, it is ideal for drawing comparisons between different analytical methods.

Some samples have already been analyzed in an earlier study to determine the origin of hazelnuts using a high-resolution LC-MS approach. The use of positive ionization mode in combination with non-polar extraction proved to be superior to other measurement options (Klockmann et al., 2016). This knowledge was used here to develop a fast method based on DART-MS. In the present study the procedure of a non-polar extraction is applied to DART-MS and thus fingerprints of the medium and non-polar components of hazelnuts can be investigated. Hereafter, the samples were classified according to their origin, using a supervised classification, to determine the accuracy of the prediction. Furthermore, it is particularly important to identify the features that were important for classification, which were then fragmented in DART-MS^2 experiments to enhance the comprehension of the features structure. Standards of the possible compounds were also measured in order to enable a clear assignment and to gain further insights into the mechanisms of the DART ionization, which is still poorly understood. The elucidation of the key features enabled a comparison between the markers identified by DART-MS and LC-MS. The DART-MS approach used in this study, which was carried out in a similar way to the LC-MS methods, aimed to identify similarities and differences between both techniques. In addition, in view of the numerous studies on the origin of hazelnuts, we compared the performance of DART-MS with all current analytical approaches. As a result, a thorough assessment of the advantages and limitations of this metabolomics-based method was conducted.

2. Materials and methods

2.1. Chemicals

Helium was purchased from Westfalen Gas (Münster, Germany). Ultrapure water was produced by a Merck Direct-Q 3 UV-R (Merck, Darmstadt, Germany) Millipore water purification system. The extraction solvents chloroform (HPLC grade) and methanol (LC-MS grade) were purchased from Carl Roth GmbH (Karlsruhe, Germany).

Butylated hydroxytoluene, glycerin-trioleate, monoacylglycerol (MG) 18:1, diacylglycerol (DG) 16:0; 18:1, DG 18:1; 18:1 and α- as well as γ-tocopherol were purchased from Sigma-Aldrich (Taufkirchen, Germany) in MS-quality.

2.2. Hazelnut samples

A total of 172 samples from the harvest years 2014–2016 were investigated. One sample consisted of an independent batch of 1.0 kg hazelnut kernels with shell (testa) or 1.5 kg of unshelled hazelnuts. No technical replicates were used. Of these samples, 50 originated from France, 15 from Georgia, 26 from Germany, 41 from Italy and 40 from Türkiye. The regions considered in each country were the relevant cultivation regions: Midi-Pyrénées and Aquitaine in France; Guria, Samegrelo and Imereti in Georgia; Bavaria in Germany; Campania, Piedmont and Lazio in Italy as well as Akçakoca, Ordu and Samsun in Türkiye. In addition, the broadest possible range of varieties was included in the data set in order to minimize the influence of varieties on the data set. Further information on the samples, such as variety, harvest year and regional origin within a country, can be found in the electronic appendix in Table S1.

Samples were provided from different sources: (i) The majority of samples were authentic and provided by hazelnut-producing companies as well as commercial importers, and for these samples both the country of origin and the place of cultivation are known. (ii) In addition, 40 semi-authentic samples from the confectionary industry were included in the dataset. The authenticity of these samples was ensured by manufacturer documentation, and of which only the country of cultivation is known. Of these samples, 6 were Georgian, 9 Italian, and 25 Turkish. The postharvest processing, which included drying, storage, and cracking was performed by the suppliers under realistic commercial conditions. After shipment to the research center, samples were stored at −40 °C to prevent further adulteration of the metabolome. The unshelled samples were cracked, and only the hazelnut kernels with testa were used for further investigation.

2.3. Sample processing

For sample processing, the unshelled hazelnuts were peeled and then frozen in liquid nitrogen. Grinding was done with a Grindomix GM 300 (Retsch, Haan, Germany) and the same amount of dry-ice to protect the samples from friction heat. Part of the ground samples were freeze-dried using a Gamma 1–20 freeze-dryer (Christ, Osterode am Harz, Germany) (Klockmann et al., 2016).

Sample extraction was done by weighing 50 ± 0.5 mg into a 2 mL safe-lock tube (Eppendorf, Hamburg, Germany) and adding two steel balls (THS, Mühlacker, Germany). Pre-cooled extraction agents used were 250 µL chloroform and 500 µL methanol, which were added to the sample. Metabolite extraction and disruption of the cells were performed with a ball mill (BeadRuptor24, Omni International, Kennesaw, USA, 3.1 m/s for 1 min). After the first extraction step, 500 µL ultrapurified water and 250 µL chloroform were added, and the ball milling was carried out for a second time for 2 min. The polar and non-polar phases were separated by centrifugation for 20 min (Eppendorf 5420R, Hamburg, Germany, 12,700 rpm, 4 °C). 400 µL of the lower non-polar phase were transferred to a new tube, and the centrifugation was repeated for 7 min. The remaining supernatant was stored at 4 °C and

measured within 24 h.

2.4. Mass spectrometry measurements

The DART-MS measurements were performed on a qToF mass spectrometer (Agilent 6560 IM-QTOF-MS, Agilent Technologies, Santa Clara, USA, which was initially tuned and calibrated for the detection of smaller masses (*m/z* 50–1700). For this step, an ESI ion source (Dual JetStream, Agilent Technologies, Santa Clara, USA) and the Agilent Technologies ESI tuning mix were used. After the tuning procedure, the DART source (Ion Sense, Saugus, USA) was installed, which was equipped with a Dip-it sample holder and automatic linear rail. The DART-source was positioned in front of the MS by means of a vapor interface (Ion Sense, Saugus, USA). The vapor interface was used to transfer the ions produced by the DART ion source into the MS. The necessary pre-vacuum in the vapor interface was generated by a pre-vacuum pump (Vacuubrand MZ 2NT, 7 mbar, Wertheim, Germany).

Samples were placed in the linear rail with a Dip-it sampler, which was dipped for 3 s into the sample and placed in the automatic rail. The DART ion source was operated at 350 ± 3 °C, the heater wait time was 50 s, the pulse duration was 5 s and the grid voltage was 350 V. As ionization gas, helium was chosen, whereas nitrogen was used as standby gas. Each measurement began with 50 s of heating followed by an air blank, for which the first Dip-it sampler position was left empty. Then the sample extracts were measured, and this procedure was repeated for all 172 samples. The used mass range was *m/z* 100–1000, and the spectral rate was 10 Hz. After every 30 measurements, the same sample extract was analyzed to check repeatability.

2.5. Marker identification

Important *m/z* ratios were identified by chemometric analysis using MS^2 experiments. For this purpose, standards for the possible analytical classes to be determined were used. About 5 mg of a standard was weighed into a 2 mL safe-lock tube, and 1 mL of chloroform was added to dissolve the substance. The MS^2 experiments were carried out according to the same procedure as the sample measurements (see chapter 2.4), with a collision energy of 20 eV and a spectral rate of 25 Hz. If no sufficient fragmentation could be achieved, the collision energy was increased to 40 eV. Samples from different countries were measured until a spectrum of good quality was obtained. The latter means, above all, that fragments with sufficiently high signal intensities were obtained.

2.6. Data handling and Chemometrics

Data acquisition of the samples and the MS^2 spectra was performed using the Agilent MassHunter Data Acquisition 9.00 software (Agilent Technologies, Santa Clara, USA). To view and evaluate the MS^1 and MS^2 spectra, the MassHunter Qualitative Analysis 10.0 software (Agilent Technologies, Santa Clara, USA) was used.

Processing of the sample spectra was started with converting the .d files to mzML files using MsConvert without additional filters. Feature finding and alignment were done with MZmine version 2.53 performing the following steps: Feature finding (noise level 10 counts), ADAP chromatogram builder (retention time limit 0.50–0.80, minimum group size 3, group intensity threshold 150, minimum highest intensity 150, *m/z* tolerance 0.01), deisotoping (*m/z* tolerance 0.01, retention time deviation 0.2 min), and alignment (*m/z* tolerance 0.01, retention time deviation 0.2 min). The data table obtained after alignment was filtered, so that only features occurring in at least 48 samples remained, so that a feature must be present in just over a quarter of the samples. Gap-filling was performed with an *m/z* deviation of 0.005 and a retention time tolerance of 0.2 min. The resulting final table was exported in .csv format.

The software MetaboAnalyst 5.0 was used for further analysis. After the data were subjected to median normalization and mean centering, principal component analysis (PCA) was used to analyze for visible outliers, but no suspicious samples were found. In order to filter out constant background features from the laboratory air and solvents, an ANOVA was performed with a cut-off value of 0.05 and the number of features was reduced from 1449 to 799.

All following analyses were done using R 4.3.3 and R-Studio 2023.12.1. PCA and visualization were performed with the package factoextra (v 1.0.7). Random Forest (RF) was performed using the implementation of the ranger package (v 0.16.0) in the mlr3 package (v 0.17.2). The cut-off value was 1/k with k = 5 (number of classes), the number of trees was increased to 1000, and the default settings for mtry and nodesize of the ranger package were used. For validation of the RF classification, 10-fold cross validation (CV) was used and repeated 100 times to minimize the influence of chance. The percentage values of the assigned samples per country are the means of all 100 repetitions of the cross-validation.

3. Results and discussion

3.1. Unsupervised analysis

First, a PCA, on the basis of 799 features, was performed. The score plots and the corresponding loading plots of PC1, PC2 and PC4 are shown in Fig. 1. Additional score plots are displayed in Fig. S1 in the electronic appendix.

The score plot of PC1 *vs.* PC2 (Fig. 1A) shows a large spatial separation of the samples. Upon coloring the samples according to their origin, it was noticeable that the German samples (green), for example, were clearly separated from all other samples by negative values on the PC1 and positive values on the PC2. Samples from France (red) were also well separated with negative loadings on PC1 and PC2, but a wider spread of the samples was observed. This can be explained by the fact that the French samples represent the largest class, and numerous regions and varieties were included in this dataset (Table S1). The Italian and Turkish samples (respectively blue and purple) were located very centrally in the scores plot, but a clustering of the samples was still evident. Samples from Georgia (brown) demonstrated high loadings on PC1 and were spatially separated, but had larger overlapping areas with the Turkish samples. It was therefore interesting to consider the loadings that caused the spatial separation in PC1 and PC2. Fig. 1C and D show the loadings on PC1 and PC2 plotted according to the *m/z* of the features. The PC1 was mainly influenced by features with a lower *m/z* ratio between *m/z* 100–350 and *m/z* 500–600. In contrast, higher loadings of *m/z* 550 and above had a minor influence on PC1. Considering the second PC, high loadings were observed for features with an *m/z* in the range of *m/z* 500–600 and *m/z* 850–920. These regions were traced back to the classes of DGs (*m/z* 500–600) and TGs (*m/z* 850–920) using database search. Due to the partial spatial separation by PC1 and PC2, these areas were determined to be responsible for the differentiation.

Moreover, the scores plot of PC1 *vs.* PC4 (Fig. 1B) shows a separation in two different clusters: France and Germany, as well as Georgian, Italian, and Turkish samples. These two clusters could be divided into Central European/Mediterranean countries with France and Germany as well as Southern European/Mediterranean countries with Georgia, Italy and Türkiye. The Central European samples showed negative values on PC1 and positive values on PC4. In contrast, the southern European/Mediterranean samples were positively loaded on PC1 as well as negative on PC4. Within these two clusters, there was a partial separation of the samples according to their origin, with overlapping areas between the Georgian, Italian and Turkish samples in particular. Considering the loading plot of PC4, the ranges of *m/z* 550–650 and *m/z* 850–920 showed high positive and negative loadings. Thus, as mentioned above, DGs and TGs could be responsible for the separation, whereas the most features between *m/z* 100–300 showed lower loadings on PC4 and thus differentiation.

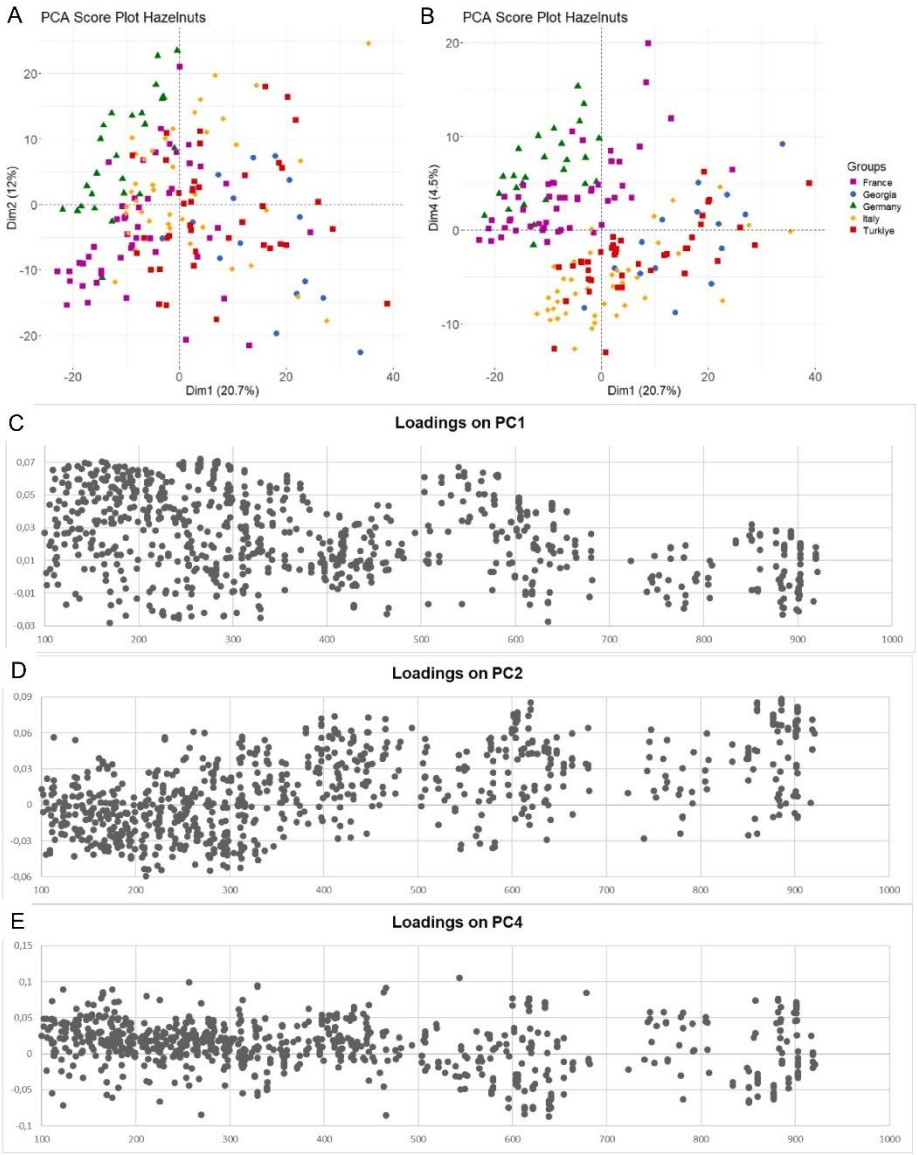


Fig. 1. PCA of all hazelnut samples, measured with the DART-MS approach. PC1 *vs.* PC2 (A) and PC1 *vs.* PC4 (B) are shown, with samples colored according to the geographical origin from France (purple), Georgia (blue), Germany (green), Italy (orange), and Türkiye (red). Furthermore, the loadings of the *m/z* of all features on PC1 (C), PC2 (D) and PC4 (E) are displayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. *Random forest classification*

The spatial separation of the samples according to origin, which could already be determined using unsupervised methods, was then continued using a supervised method to obtain a complete separation of the samples.

For this purpose, an RF analysis was performed with all 799 features and a 10-fold CV with 100 repetitions. The results obtained are reported in Table 1, leading to an overall accuracy of 93.2 %. Samples from France demonstrated an accuracy rate of 92.7 %, and there were only a few misclassifications. 3.8 % of the French samples were classified as German, which is due to the geographical proximity of the two

Table 1
Confusion matrix of classification results in percent of the hazelnut samples calculated with the Random Forest algorithm and validated with 100 times repeated 10-fold cross-validation. The classification was performed with samples from France, Georgia, Germany, Italy, and Türkiye. In addition, the total number of samples used for analysis is given.

		Response class					
		France	Georgia	Germany	Italy	Türkiye	#
True class	France	92.7	1.7	3.8	1.8	0.0	50
	Georgia	2.7	89.0	0.0	8.4	0.0	15
	Germany	0.2	0.0	99.8	0.0	0.0	26
	Italy	2.4	3.6	0.0	91.8	2.3	41
	Türkiye	0.0	6.9	0.0	0.4	92.6	40

countries. In addition, the French sample set was very diverse, as there were many varieties and regions. As already observed in the PCA (see Chapter 3.1), this diversity made the class more inhomogeneous, which on the one hand leads to a more realistic model, but also increases the difficulty of classification and differentiation. Furthermore, the Georgian samples provided a good classification rate with 89.0 %, but 8.4 % of the Georgian samples were classified as Italian. Georgia represented the smallest class with 15 samples, while Italy represented the second-largest class with 41 samples, which may have led to a greater weighting of the Italian samples in the multivariate classification and thus to some misclassifications of small classes. The German class showed excellent classification results with 99.8 %. Only 0.2 % of the German samples were classified as French, which confirms the spatial proximity of the classes already discussed for the French samples. Italian samples were correctly assigned with a very good accuracy of 91.8 %, which demonstrated the good suitability of the model for detecting the economically important Italian hazelnuts. In the classes of Türkiye, France, and Georgia, all countries with a Mediterranean climate, misclassifications between 2.3 and 3.6 % occurred. Furthermore, 92.6 % of correct classifications were found for the class of the Turkish samples, demonstrating a very good classification of this group of samples. However, 6.9 % of the samples were assigned to the Georgian samples. This confusion was to be expected, as Türkiye and Georgia are neighboring countries. Furthermore, hazelnut cultivation is concentrated in Türkiye and Georgia near the Black Sea coast, which leads to very similar climates and thus similar influences on the metabolome. Overall, the good to excellent classification rates of the individual countries of 89.0–99.8 % showed that the samples were classified very well, leading to a very good overall accuracy of 93.2 %.

When classifying this data set, it was also necessary to consider whether there were other differences among the sample classes that were not related to origin. This concerns the harvest years and the varieties of the samples. As listed in Table S1 of the electronic appendix, samples from the 2014–2016 vintages were present in all classes. Only the Georgian class lacks samples from the year 2016. This is a minor limitation, but the impact on classification was considered to be very low. The influence of varieties on classification was more difficult to assess, as the variety in hazelnut cultivation is an essential part of a region's authenticity. *Nocciola Romana*, for example, is registered as a protected designation of origin. This protection of *Nocciola Romana* applies to both the prescribed variety and the specific growing area in Viterbo and Rome, making the variety an integral part of the authentic origin, which is also protected (Council Regulation EC No 510/2006, 2006). Consequently, the separation of variety and origin could not be completely ruled out in this study.

The importance of the variables for classification was then analyzed based on the variable importance of the RF and ranked based on the mean decrease in accuracy (MDA). A table of the 100 most important variables for classification can be found in Table S2 of the electronic appendix. When considering the essential variables, it was noticeable that three mass ranges in particular were crucial for the classification. The m/z 550–650 range was associated to DGs, and the range m/z 820–920 could be assigned to TGs. These two mass ranges were already identified as significant using PCA (see Chapter 3.1) and the assumption from the unsupervised analysis was confirmed. In addition, features in the range of m/z 410–470 proved to be the most important for the classification. The four most significant features alone, m/z 431.3522; 465.4090; 417.3861 and 447.3878, are located in this mass range. Through a database search, these signals could be assigned to the class of tocopherols, which represented another important class of markers.

3.3. Identification and interpretation of relevant markers

In order to identify as many relevant features as possible, reference substances were ordered, one for each molecule class. These were measured using DART-MS^2 experiments, and the corresponding MS^2 spectra can be found in Fig. S3-S23 of the electronic supplementary material. The characteristic fragments of these standards are shown in Table 2. A few markers could not be identified as the DART-ionization provides lower intensities for some markers, and therefore sufficient fragment spectra could not be obtained.

First, the fragmentation patterns of α- and γ-tocopherol were analyzed, which showed the characteristic fragments of m/z 165 and m/z 151. The fragments obtained were also observed by LC-ESI-MS^2 (Giusepponi et al., 2019).

Based on the fragmentation patterns revealed, the four masses with an m/z 431.3884, 447.3833, 417.3618, and 463.3788, which contributed to the classification of the samples, were assigned to the class of tocopherols. Using the DART source, these features could be detected as $[M + H]^+$ ions. The fragmentation patterns of the masses from the samples can be found in Table 3. The mass m/z 431.3811 was identified as α-tocopherol, while γ-tocopherol was identified as m/z 417.3618. In addition, the characteristic fragment of α-tocopherol of m/z 165.0867 was also observed in two other features, which were assigned to hydroxy-tocopherol with m/z 447.3863 and di-hydroxy-tocopherol with m/z 463.3999.

DG 34:1 and DG 36:2 were measured as standards for the class of DGs. Both DGs ionized as $[M + H]^+$, $[M + H - H_2O]^+$ and $[M + NH_4]^+$ ions. The formation of $[M + NH_4]^+$ adducts has already been observed in other studies and is probably due to the fact that ammonium ions are often present in the atmosphere of chemical laboratories (Gross, 2014; Schmauder et al., 2024). The fragmentation of the DGs mainly led to the corresponding MGs with fragmentation of a fatty acid (FA) with an m/z of 313.2687 for MG 16:0 and m/z 339.2838 for MG 18:1. In addition, during fragmentation of $[M + H]^+$ ion of DG 34:1 (m/z 595.5209), the $[M + H\text{-}H_2O]^+$ ion was additionally detected, which shows that the fragmentation can also lead to the loss of water. While DG 34:1 showed two fragments based on the cleavage of the two FA 18:1 and 16:0, as expected, DG 36:2 showed a simpler fragmentation pattern since only FA 18:1 is bound in the DG, resulting in only one fragment ion (m/z 339.2894) of MG 18:1.

The relevant classification features belonging to the class of the DGs were DG 32:0, DG 34:1, DG 36:2, DG 36:1 and DG 36:4. This relatively high number of relevant marker compounds underlines the importance of this class for the classification result. It is notable that several DGs in the form of different adduct ions were relevant for the classification result. For example, DG 36:2 was ranked as the $[M + H]^+$ ion with an MDA of 0.806 and as the $[M + H - H_2O]^-$ ion with an MDA of 0.208 in the most important variables, demonstrating the importance of both ions for the result. By identifying 22 features, 16 markers were thus found, since in some cases, two or three different adduct ions of the same molecule were relevant for the result.

TG 54:3 was selected as a reference substance for the class of TGs and was also detected twice as $[M + H]^-$ and $[M + NH_4]^-$ ions. In both cases, the fragmentation led to the cleavage of a fatty acid and detection of the DG 36:2 (m/z 603.5247) which could be detected as an $[M + H - H_2O]^+$ ion. With TG 38:3, TG 50:1, 52:2, TG 54:3, TG 54:4 and TG 54:5, important markers were found within the class of TGs. These markers

Table 2
Fragmentation behavior of different standard substances for non-polar authenticity analysis, measured using a DART-MS platform.

Standard	Formula	Adduct	Theoretical m/z	Measured m/z	Delta m/z	Relevant fragments in MS^2 mode
α-Tocopherol	$C_{29}H_{51}O_2$	$[M + H]^-$	431.3884	431.3801	0.0083	165.0871
γ-Tocopherol	$C_{28}H_{48}O_2$	$[M + H]^-$	417.3727	417.3655	0.0072	151.0715
FA 18:1	$C_{18}H_{34}O_2$	$[M + H]^-$	283.2631	283.2586	0.0045	107.0823; 121.0978; 135.1133; 149.1285
DG 34:1	$C_{37}H_{70}O_5$	$[M + NH_4]^+$	612.5561	612.5475	0.0086	313.2687; 339.2838; 577.5088
DG 34:1	$C_{37}H_{70}O_5$	$[M + H]^+$	595.5196	595.5209	0.0087	313.2687; 339.2838; 577.5088
DG 34:1	$C_{37}H_{70}O_5$	$[M + H - H_2O]^+$	577.5190	577.5109	0.0081	313.2687; 339.2838; 577.5088
DG 36:2	$C_{39}H_{72}O_5$	$[M + NH_4]^+$	638.5718	638.5627	0.0091	265.2526; 339.2894
DG 36:2	$C_{39}H_{72}O_5$	$[M + H]^+$	621.5452	621.5362	0.0090	265.2526; 339.2894
DG 36:2	$C_{39}H_{72}O_5$	$[M + H - H_2O]^+$	603.5347	603.5261	0.0086	265.2526; 339.2894
TG 54:3	$C_{57}H_{104}O_6$	$[M + H]^+$	885.7906	885.7773	0.0133	603.5247
TG 54:3	$C_{57}H_{104}O_6$	$[M + NH_4]^+$	902.8171	902.8040	0.0131	603.5247

Table 3
Important signals to determine the geographical origin of hazelnuts. Identification was performed by means of MS^2 spectra obtained by the DART-MS platform. The features are listed with their relevant fragments as well as the mean decrease in accuracy, which was calculated using Random Forest classification.

Tentative compound	Proposed formula	Adduct	Theoretical m/z	Measured m/z	Delta m/z	Relevant fragments	Mean decrease accuracy
α-Tocopherol	$C_{29}H_{50}O_2$	$[M + H]$	431.3884	431.3811	0.0073	165.0867	3.186
α-OH-Tocopherol	$C_{29}H_{50}O_3$	$[M + H]^-$	447.3833	447.3863	0.0030	165.0867	2.288
γ-Tocopherol	$C_{28}H_{48}O_2$	$[M + H]$	417.3727	417.3618	0.0109	151.0709	1.664
α-OH-OH-Tocopherol	$C_{29}H_{50}O_4$	$[M + H]^-$	463.3788	463.3999	0.0211	165.0864	1.440
DG 32:0	$C_{35}H_{68}O_5$	$[M + H - H_2O]^+$	551.5030	551.4943	0.0091	239.2324	0.265
DG 34:1	$C_{37}H_{70}O_5$	$[M + NH_4]^-$	612.5561	612.5574	0.0013	313.2674; 339.2826	0.644
DG 34:1	$C_{37}H_{70}O_5$	$[M + H]^-$	595.5196	595.5198	0.0002	313.2671; 339.2821	0.380
DG 34:1	$C_{37}H_{70}O_5$	$[M + H - H_2O]^+$	577.5190	577.5093	0.0097	313.2821; 339.2828	0.405
DG 36:2	$C_{39}H_{72}O_5$	$[M + H]^-$	621.5452	621.5470	0.0018	339.2828	0.806
DG 36:2	$C_{39}H_{72}O_5$	$[M + H - H_2O]^+$	603.5347	603.5249	0.0098	339.2843	0.208
DG 36:1	$C_{39}H_{74}O_5$	$[M + NH_4]^-$	640.5874	640.5721	0.0153	339.2823; 341.3963	0.238
DG 36:1	$C_{39}H_{74}O_5$	$[M + H - H_2O]^+$	605.5503	605.5506	0.0003	337.2681; 339.2813	0.175
DG 36:4	$C_{39}H_{68}O_5$	$[M + H]^-$	617.5139	617.5167	0.0028	265.2477; 339.2830	0.158
DG 36:4	$C_{39}H_{68}O_5$	$[M + H - H_2O]^+$	599.5034	599.5063	0.0029	339.2819	0.183
TG 38:3	$C_{41}H_{72}O_6$	$[M + NH_4]^-$	678.5672	678.5550	0.0122	379.2774; 381.2931	0.396
TG 50:1	$C_{53}H_{100}O_6$	$[M + NH_4]$	850.7858	850.7864	0.0006	551.4932; 577.5095	0.922
TG 50:1	$C_{53}H_{100}O_6$	$[M + H]$	833.7593	833.7442	0.0151	551.4940; 577.5098	0.581
TG 52:2	$C_{53}H_{102}O_6$	$[M + NH_4]$	876.8015	876.8030	0.0015	577.5095; 603.5363	0.301
TG 54:3	$C_{57}H_{104}O_6$	$[M + H]^-$	885.7906	885.7919	0.0013	603.5353	0.262
TG 54:4	$C_{57}H_{102}O_6$	$[M + NH_4]^-$	900.8015	900.8027	0.0012	339.2829; 603.5251	0.270
TG 54:5	$C_{57}H_{100}O_6$	$[M + H]^-$	881.7593	881.7453	0.0140	601.5089; 603.5253	0.202
DGTS 36:2	$C_{46}H_{85}NO_7$	$[M + H]^-$	764.6404	764.6647	0.0243	603.5230	0.129

Abbreviations: DG. diacylglycerol; DGTS. diacylglyceryl-O-4′-(*N*,*N*,*N*-trimethyl)-homoserine; FA. fatty acid; MG. monoacylglycerol; PE. glycerophosphoethanolamine; PC. glycerophosphocholine; TG. triacylglycerol.

ionized only as $[M + NH_4]^-$ adducts, $[M + H - H_2O]^-$ adducts were not observed. The best marker compound among the TGs was TG 50:1, which could be detected as an $[M + NH_4]^+$ ion with an MDA of 0.922 and as an $[M + H]^+$ ion with an MDA of 0.581. All other identified TGs were ranked with MDAs of 0.202–0.396.

Furthermore, diacylglyceryl-O-4′-(*N*,*N*,*N*-trimethyl)-homoserine (DGTS) 36:2 was identified from the class of betaine lipids by the neutral loss of *m/z* 161 with an MDA of 0.129, DGTS 36:2 contributed to a slightly better classification.

When comparing the classes of the identified markers, tocopherols were the most relevant with MDAs of 1.440–3.186, indicating a high impact of tocopherols on the classification result. Furthermore, DGs with MDAs between 0.158 and 0.806 and TGs with 0.202–0.922 showed about the same influence, but the DGs and TGs were the largest class with 17 identified features, showing that the TGs and DGs generally exert a pronounced effect on the classification result. The identified DGTS had a relative lower MDA of 0.129, but showed that the ionization and identification of betaine lipids by DART-MS was possible. A total of 22 features were traced back to 16 marker substances that are associated with four different substance classes.

Fig. 2 shows boxplots of all 16 identified markers, separated according to the five different origins of the samples. Particularly, for the tocopherols (Fig. 2A-D), there were noticeable differences in signal intensities between the samples of different origins. For example, α-OH-OH-tocopherol could be detected in the German samples with comparatively high signal intensities, while low intensities were present in the Georgian, Italian, and Turkish samples. On the other hand, γ-tocopherol showed especially high signal intensities in the Turkish samples. α- and γ-tocopherol act as antioxidants in both plants and humans, helping to prevent the oxidation of lipids in cell membranes. The higher γ-tocopherol content in hazelnuts from Türkiye makes them less susceptible to lipid oxidation (Górnaś et al., 2022). The results of the high influence of tocopherols were consistent with other studies on geographical origin, where, for example, tocopherols were relevant markers for the origin of coffee beans and soy seeds (Ghosh et al., 2021; Ismaiel et al., 2023). Furthermore, DG 32:0, DG 34:1, DG 36:2, DG 36:1, TG 50:1, TG 52:2 and FA 18:1 (Fig. 2E-H,K,L,P) demonstrated higher signal intensities in the Georgian, Italian and Turkish samples, whereas lower intensities occurred in the French and German samples. Therefore, the mono- and di-unsaturated DG and TGs could be used to separate the samples from Central Europe, including Germany and France, from the southern European/Mediterranean countries. In contrast, the lipids DG 36:4, TG 38:3, TG 54:4 and TG 54:5 (Fig. 2I,J,M,N), which contained higher levels of unsaturated FAs, showed higher signal intensities in the Central European samples. DGs and TGs serve as energy storage compounds for plants and are synthesized or degraded depending on energy

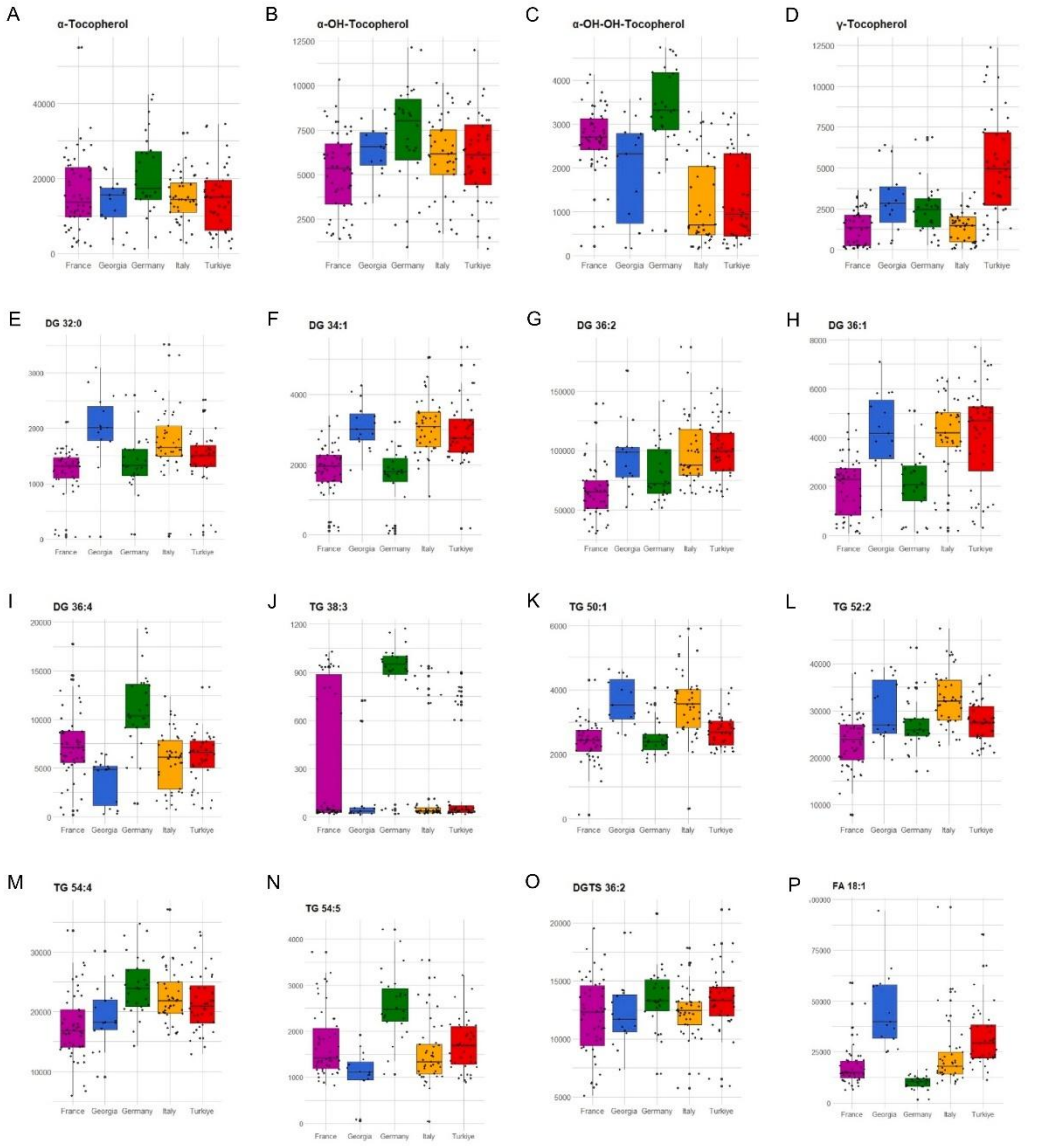


Fig. 2. Boxplots of identified marker compounds to distinguish the origin of hazelnuts. The boxplots show the relative concentration after mean normalization and standardization. The colour code reflects the origin of the samples from France (purple), Georgia (blue), Germany (green), Italy (orange), and Türkiye (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

requirements. The production and breakdown of DGs and TGs are influenced by internal and external factors, which leads to the formation of various derivatives of these chemical classes (Wang et al., 2021). These differences between saturated and polyunsaturated FAs were also found to be relevant in other studies, for example, for the authentication of seeds and nuts (Boukid et al., 2019; Lösel et al., 2024). In addition, higher saturated fat levels in warmer countries could be linked to the plant's response to heat and drought stress. For example, polyunsaturated fatty acids contributed to tolerance at lower temperatures, while low levels of unsaturated fats increased resistance to high temperatures (Upchurch, 2008). DGTS 36:2 (Fig. 2O) showed only a minor difference between Germany and Georgia. Betaine lipids, such as DGTS,

play a major role in the lipid metabolism of FA 18:1 and serve as substrates for desaturation, which confirms the correlation between the saturation levels of lipids and the differentiation of origin described above (Vogel & Eichenberger, 1992). However, these could contribute slightly to the result of the multivariate classification. Therefore, the main differences between the individual countries were traced back to the different tocopherols as well as the different fat compositions.

3.4. Comparison of DART-MS with LC-ESI-MS measurements

The hazelnut dataset for analysis using DART-MS has already been analyzed and classified using a non-targeted LC-ESI-MS approach (Klockmann et al., 2016). It was therefore appropriate to compare these two methods.

Measurements of 196 hazelnut samples from five different origins using LC-ESI-MS and classification with support-vector-machine (SVM) as well as leave-one-out cross-validation (LOOCV) led to an accuracy of 97.4 % (Klockmann et al., 2016). The comparison with the 93.2 % accuracy of the DART-MS approach (see chapter 3.2) showed a slightly better classification accuracy for LC-MS. It should be noted that the results of DART-MS were validated using 10-fold cross-validation and those of LC-MS using LOOCV. In a LOOCV, the fact that the test set contains only one sample can lead to more parallel models than a 10-fold CV. Furthermore, chromatographic separation using LC-MS may contribute to a better ionization, since in a 24-min measurement, as developed by Klockmann et al., fewer molecules have to be ionized simultaneously, thus reducing ion suppression effects (Klockmann et al., 2016).

Time saving is an advantage of DART-MS for high-throughput analysis, providing results within minutes, whereas the measurement of a complete data set with LC-MS can extend over several days. In addition to the rapid analysis using DART-MS, the elimination of chromatography means that no solvents had to be used during analysis. The DART-MS approach therefore results in a more environmentally friendly analysis, as organic solvents were only required for extraction, not for measurement. Extraction required about one milliliter of solvent, while about 12 mL of solvent were used per sample for liquid chromatography. Consequently, the use of DART-MS reduced the amount of solvent required by around 92 %. Depending on how LC-MS measurements are carried out, DART-MS measurements are about one-third cheaper, mainly due to the costs of the solvents for the LC-MS analysis.

The sustainability of DART-MS measurements was tested using the Analytical GREEnness metric approach and software (AGREE) tool and compared with an LC-MS measurement (Pena-Pereira et al., 2020). The DART-MS approach achieved a sustainability score of 0.61 and an LC-MS measurement of 0.48. The higher score of the DART-MS based measurements demonstrated being more sustainable than the established LC-MS approach. The results of the analysis are shown in Fig. S2 of the supplementary information.

On the other hand, an unsustainable aspect of using DART ion sources is the use of helium as an ionization gas. Helium is obtained as a by-product of natural gas extraction and is therefore not sustainable (Olafsdottir & Sverdrup, 2020). This problem of a possible helium shortage is countered by use of helium recycling plants (Jones et al., 2023). In addition, the consumption of a DART ion source is very low at 2.5 l/min and nitrogen or argon may be used as an alternative ionization gas (Gross, 2014).

A comparison of the most important metabolites identified by LC-MS and DART-MS revealed major similarities: The metabolites identified by LC-MS included tocopherols, DGs and TGs. In addition, phospholipids from the classes of phosphatidylethanolamines (PEs) and glycerophosphocholines (PCs) were detected. Except for PCs and PEs, markers from the same substance classes were identified using DART-MS. The fact that phospholipids could be identified by LC-MS but not by DART-MS is due to the different ionization processes that occur with DART and the ESI sources otherwise used for LC-MS measurements. In previous studies, phospholipid standards were used to show that PEs and PCs already fragmented during ionization, and the phosphate group was split off. As a result, only the DGs resulting from fragmentation can be detected, and not the phospholipids themselves. (Schmauder et al., 2024). For this reason, no PCs or PEs could be identified as markers using DART-MS. However, an increased number of DGs could be identified with 5 DGs. Whether an ion represents an in-source fragmentation of a phospholipid or a DG cannot be conclusively clarified due to the same m/z of both ions. Considering the number of metabolites identified using the DART-MS approach, 4 tocopherols, 5 DGs, 6 TGs and one DGTS were identified. In contrast, γ-tocopherol, 5 PCs, 3 PEs, 4 DGs and 7 TGs were identified using LC-MS. γ-tocopherol, DG 34:1, DG 36:2 and TG 38:3 were identified by both methods.

3.5. Comparison of the results with other metabolomics approaches

In addition to the LC-MS approach described in Chapter 3.4, other analytical platforms were used to authenticate hazelnuts. An overview can be found in Table 4. Considering the accuracy of the classifications, the described LC-MS approach was superior to all other technologies with 97.4 %, followed by an NMR-based approach with 96 % accuracy. Both studies investigated the non-volatile metabolome (Bachmann et al., 2018; Klockmann et al., 2016). This result was in line with expectations, as LC-MS and NMR are established methods for authenticating food. In addition to the accuracy of the classification, repeatability and interpretability must also be taken into account. NMR is a very robust method with good reproducibility, which makes it possible to compare data sets recorded over long periods of time. This aspect is a major disadvantage that occurs in the context of LC-MS analyses. The advantages of LC-MS analyses lie primarily in the lower detection limits, which enable a higher number of metabolites to be detected (Consonni & Cagliani, 2019; Lolli & Caligiani, 2024).

An alternative to LC-MS or NMR is the investigation of the volatile metabolome by gas chromatography coupled mass spectrometry (GC–MS). For example, a headspace GC–MS approach for four classes using an independent test data set achieved an accuracy of 86 % and gained further insights into the aroma profile of hazelnuts (Sammarco, Bardin, et al., 2023). In contrast to the investigation of the non-volatile metabolome, the GC–MS approach of the volatile metabolome showed a lower accuracy. Furthermore, it is also noteworthy that the simpler analysis of the fatty acid profile using GC-FID was also suitable for authentication and allowed the classification of three regions with an accuracy of 86.2 %.

All methods, LC-MS, NMR, GC–MS and GC-FID, had a measurement time of 20–30 min in common. In contrast, the DART-MS approach described above offers the advantage of very short analysis times, as the measurement time could be reduced to less than one minute. Furthermore, the amounts of chemicals used were reduced, while large amounts of organic solvents were required for LC-MS, large quantities of liquid helium for NMR, and in the case of fatty acid distribution (GC–MS,-FID), toxic derivatization reagents. Only the analysis using a headspace approach with GC–MS is comparably environmentally friendly and chemical-friendly like the analysis using DART-MS. The accuracy of 93.2 % of the DART-MS approach proved to be superior to the analysis of the volatile metabolome and fatty acid distribution. However, a small test data set was used for the volatile metabolome analysis, which makes comparison by means of the CV difficult.

Considering other rapid methods, NIR spectroscopy has been widely used for origin determination. Classification results of 90.6–96.3 % were obtained in several studies, although the number of samples and countries varied. The only study, also conducted with five classes, resulted in an accuracy of 90.6 %, showing that the DART-MS approach was superior to the NIR approach (Shakiba et al., 2022). Higher accuracies achieved by different NIR approaches can be explained by a small number of groups in the classification (Moscetti et al., 2015; Sammarco, Dall'Asta, & Suman, 2023). A smaller number of groups increases the

Table 4
Overview of studies already conducted to determine the origin of hazelnuts.

	Approach	Number of samples	Number of classes	Classification method	Validation method	Accuracy [%]
Klockmann et al.	LC-MS	195	5	SVM	LOOCV	97.4
Bachmann et al.	NMR	262	5	Random Subspace	Testset	96.0
Moscetti et al.	FT-NIR	1200	3	SMVDA	CV	95.5
Shakiba et al.	FT-NIR	233	5	Random Subspace	RNCV	90.6
Sammarco et al.	FT-NIR	57	4	PLS-DA	Testset	96.3
Sammarco et al.	GC–MS	107	4	PLS-DA	Testset	86
Tüfekci and Karaç	GC-FID	87	3	LDA	CV	86.2
This study	DART-MS	172	5	Random Forest	CV	93.2

Abbreviations: CV. cross-validation, LDA. linear discriminant analysis, LOOCV. leave-one-out CV, PLS-DA. partial least square discriminant analysis, RNCV. repeated-nested CV, SVM. support-vector-machine.

probability of a correct classification *a priori* and thus leads to better classification results. Furthermore, the publication by Moscetti et al. is based on a large dataset of three origins with 1200 samples, which led to an excellent classification rate due to the reduced number of classes and the high number of samples. The different dataset is therefore only partially suitable for comparison, which was why the results of Shakiba et al. were preferred for comparison. Studying the application of NIR and DART-MS approaches in the laboratory, both methods were fast, with one minute per sample, and required minimal or no chemical use. While NIR had an advantage due to the simpler sample preparation without extraction, the DART-MS approach is superior in the identification of markers and the interpretability of the data. In addition, DART-MS was used to identify detailed marker compounds for classification purposes, whereas NIR only permits the narrowing of wavenumbers and traceability to molecular classes such as lipids, in accordance to the existing literature.

Overall, the DART-MS approach used proved to be well suited for authentication when a rapid method without chromatography as well as an environmentally friendly analysis with the possibility of marker identification is required.

4. Conclusion

In the present study, the possibilities of DART-MS for determining the origin of hazelnuts were investigated. For this purpose, 172 hazelnut samples from 6 important growing countries were analyzed using a non-targeted lipidomics approach. The classification of the samples, which was carried out with a Random Forest model and evaluated by a 10-fold cross-validation, resulted in a very good accuracy of 93.2 %. The correct classifications of the 6 individual countries were between 89.0 and 99.8 %, which highlights the high suitability of DART-MS for origin classification. Furthermore, a ranking of significant features was obtained, and by measuring standard substances, 22 features were traced back to 16 important marker substances. In particular, the identified features belong mainly to the classes of tocopherol derivatives as well as DGs and TGs. A good literature base on hazelnut authentication allowed the comparison of DART-MS with a wide range of alternative methods for origin authentication. This comparison showed that the accuracy of the DART-MS approach was between the already established low- and high-tech methods. Compared to other high-tech methods, DART-MS offers the possibility of faster analysis times, cost-efficient measurements, and environmentally friendliness, making DART-MS a serious emerging option for origin authentication.

CRediT authorship contribution statement

Felix Schmauder: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marina Creydt:** Writing – review & editing, Methodology, Conceptualization. **Markus Fischer:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

We thank Sven Klockmann and René Bachmann for the acquisition of the samples and Julian Hoppe for his help with sample preparation as well as Johannes Brockelt and Timothy Kehrer for proof reading.

Funding

The idea of this project arose in the context of the project "Food Profiling - Development of analytical tools for the experimental verification of the origin and identity of food". This Project (Funding reference number: 2816500914) is supported by means of the Federal Ministry of Food and Agriculture (BMEL) by a decision of the German Bundestag (parliament). Project support is provided by the Federal Office for Agriculture and Food (BLE) within the scope of the program for promoting innovation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.142265.

References

Ahmadov, A. (2024). The impact of climate change on hazelnut cultivation. *Turkish Journal of Food and Agriculture Sciences, 6*, 106–115.
An, N., Turp, M. T., Türkeş, M., & Kurnaz, M. L. (2020). Mid-term impact of climate change on hazelnut yield. *Agriculture, 10*, 159.
Bachmann, R., Klockmann, S., Haerdter, J., Fischer, M., & Hackl, T. (2018). 1H NMR spectroscopy for determination of the geographical origin of hazelnuts. *Journal of Agricultural and Food Chemistry, 66*, 11873–11879.
Boukid, F., Abbattangelo, S., Carini, E., Marseglia, A., Caligiani, A., & Vittadini, E. (2019). Geographical origin discrimination of pistachio (*Pistacia vera* L.) through combined analysis of physical and chemical features. *European Food Research and Technology, 245*, 143–150.
Consonni, R., & Cagliani, L. R. (2019). The potentiality of NMR-based metabolomics in food science and food authentication assessment. *Magnetic Resonance in Chemistry, 57*, 558–578.
Council Regulation (EC) No 510/2006. (2006). COUNCIL REGULATION (EC) No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural COUNCIL REGULATION (EC) No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. https://eur-lex.europa.eu/legal-conte

nt/EN/TXT/PDF/?uri=CELEX:32006R0510&qid=1731598995469 Accessed 14.11.2024.
European Comission. (2024). EAmbrosia the EU geographical indications register. https://ec.europa.eu/agriculture/eambrosia/geographical-indications-register/ Accessed 21.03.2024.
FAO. (2023). Hazelnut production: Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL. Accessed 21.03.2024.
Ghosh, S., Zhang, S., Azam, M., Qi, J., Abdelghany, A. M., Shaibu, A. S., … Htway, H. T. P. (2021). Seed tocopherol assessment and geographical distribution of 1151 Chinese soybean accessions from diverse ecoregions. *Journal of Food Composition and Analysis, 100*, Article 103932.
Giusepponi, D., Galarini, R., Barola, C., Torquato, P., Bartolini, D., Moretti, S., Saluti, G., Gioiello, A., Libetta, C., & Galli, F. (2019). LC-MS/MS assay for the simultaneous determination of tocopherols, polyunsaturated fatty acids and their metabolites in human plasma and serum. *Free Radical Biology and Medicine, 144*, 134–143.
Górnaś, P., Baškirovs, G., & Siger, A. (2022). Free and esterified tocopherols, tocotrienols and other extractable and non-extractable tocochromanol-related molecules: Compendium of knowledge, future perspectives and recommendations for chromatographic techniques, tools, and approaches used for tocochromanol determination. *Molecules, 27*, 6560.
Gross, J. H. (2014). Direct analysis in real time—A critical review on DART-MS. *Analytical and Bioanalytical Chemistry, 406*, 63–80.
Ismaiel, L., Fanesi, B., Kuhalskaya, A., Barp, L., Moret, S., Pacetti, D., & Lucci, P. (2023). The determination of Triacylglycerols and tocopherols using UHPLC–CAD/FLD methods for assessing the authenticity of coffee beans. *Foods, 12*, 4197.
Jones, A. T., Down, R. B., Lawson, C. R., Keeping, D., & Kirichek, O. (2023). Carbon footprint of helium recovery systems. *Low Temperature Physics, 49*, 967–970.
Kim, H. J., Seo, Y. T., Park, S., Jeong, S. H., Kim, M. K., & Jang, Y. P. (2015). DART–TOF–MS based metabolomics study for the discrimination analysis of geographical origin of Angelica gigas roots collected from Korea and China. *Metabolomics, 11*, 64–70.
Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., & Fischer, M. (2016). Food fingerprinting: Metabolomic approaches for geographical origin discrimination of hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS. *Journal of Agricultural and Food Chemistry, 64*, 9253–9262.
Klockmann, S., Reiner, E., Cain, N., & Fischer, M. (2017). Food targeting: Geographical origin determination of hazelnuts (*Corylus avellana*) by LC-QqQ-MS/MS-based targeted metabolomics application. *Journal of Agricultural and Food Chemistry, 65*, 1456–1465.
Lolli, V., & Caligiani, A. (2024). How NMR contributes to food authentication: Current trends and perspectives. *Current Opinion in Food Science, 101200*.
Lösel, H., Arndt, M., Wenck, S., Hansen, L., Oberpottkamp, M., Seifert, S., & Fischer, M. (2024). Exploring the potential of high-resolution LC-MS in combination with ion mobility separation and surrogate minimal depth for enhanced almond origin authentication. *Talanta, 271*, Article 125598.
Moscetti, R., Radicetti, E., Monarca, D., Cecchini, M., & Massantini, R. (2015). Near infrared spectroscopy is suitable for the classification of hazelnuts according to protected designation of origin. *Journal of the Science of Food and Agriculture, 95*, 2619–2625.
Olafsdottir, A. H., & Sverdrup, H. U. (2020). Assessing the past and future sustainability of global helium resources, extraction, supply and use, using the integrated assessment model WORLD7. *Biophysical Economics and Sustainability, 5*, 6.
Pena-Pereira, F., Wojnowski, W., & Tobiszewski, M. (2020). AGREE—Analytical GREEnness metric approach and software. *Analytical Chemistry, 92*, 10076–10082.
Sammarco, G., Bardin, D., Quaini, F., Dall'Asta, C., Christmann, J., Weller, P., & Suman, M. (2023). A geographical origin assessment of Italian hazelnuts: Gas chromatography-ion mobility spectrometry coupled with multivariate statistical analysis and data fusion approach. *Food Research International, 171*, Article 113085.
Sammarco, G., Dall'Asta, C., & Suman, M. (2023). Near infrared spectroscopy and multivariate statistical analysis as rapid tools for the geographical origin assessment of Italian hazelnuts. *Vibrational Spectroscopy, 126*, Article 103531.
Sammarco, G., Rossi, M., Suman, M., Cavanna, D., Viotto, L., Pettenà, P., Dall'Asta, C., & Iacumin, P. (2023). Hazelnut products traceability through combined isotope ratio mass spectrometry and multi-elemental analysis. *JSFA Reports, 3*, 633–645.
Schmauder, F., Schütz, D., Creydt, M., Riedl, J., & Fischer, M. (2024). Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS). *Food Control, 110489*.
Shakiba, N., Gerdes, A., Holz, N., Wenck, S., Bachmann, R., Schneider, T., Seifert, S., Fischer, M., & Hackl, T. (2022). Determination of the geographical origin of hazelnuts (*Corylus avellana* L.) by near-infrared spectroscopy (NIR) and a low-level fusion with nuclear magnetic resonance (NMR). *Microchemical Journal, 174*, Article 107066.
Statistisches Bundesamt. (2024). Entwicklung der Einfuhrpreise von Haselnuessen in Deutschland in den Jahren 2009 bis 2023. https://de.statista.com/statistik/daten/studie/385791/umfrage/entwicklung-der-einfuhrpreise-von-haselnuessen-in-deutschland/ Accessed 21.03.2024.
Tüfekci, F., & Karataş, Ş. (2018). Determination of geographical origin Turkish hazelnuts according to fatty acid composition. *Food Science & Nutrition, 6*, 557–562.
Upchurch, R. G. (2008). Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnology Letters, 30*, 967–977.
Vogel, G., & Eichenberger, W. (1992). Betaine lipids in lower plants. Biosynthesis of DGTS and DGTA in *Ochromonas danica* (Chrysophyceae) and the possible role of DGTS in lipid metabolism. *Plant and Cell Physiology, 33*, 427–436.
Wang, K., Xu, L., Wang, X., Chen, A., & Xu, Z. (2021). Discrimination of beef from different origins based on lipidomics: A comparison study of DART-QTOF and LC-ESI-QTOF. *LWT, 149*, Article 111838.
Wang, Y. (2024). Recent advances in the application of direct analysis in real time-mass spectrometry (DART-MS) in food analysis. *Food Research International, 114488*.

6.4 Herkunftsauthentifizierung von Erdbeeren (*Fragaria* × *ananassa*) mittels FT-NIR Spektroskopie.

Eine weitere verbreitete Methode für die schnelle Analytik von Authentifizierungsfragen ist die FT-NIR Spektroskopie. Um einen Vergleich mit dem in der vorherigen Publikation gewählten DART-MS-basierten Ansatz vorzunehmen, wurden insgesamt 198 Erdbeerproben, die aus vier unterschiedlichen Ländern stammten, mit zwei verschiedenen NIR-Spektrometern untersucht. Dafür wurden ein größeres *benchtop* Gerät und ein einfacheres mobiles NIR-Gerät eingesetzt und die Spektren der beiden Geräte für einen Vergleich herangezogen.

Die Auswertungen umfassten ein Zwei-Klassen-Modell, in dem die deutschen Proben von den ausländischen Proben unterschieden wurden, sowie ein Vier-Klassen-Modell, in dem die Proben ihrem Ursprungsland zugeordnet wurden. Der Vergleich der NIR-Geräte zeigte, dass das eingesetzte mobile NIR-Gerät für die Unterscheidung von deutschen und ausländischen Erdbeeren geeignet ist und eine Richtigkeit von 84,0% lieferte. Darüber hinaus konnte mit dem größeren *benchtop* Gerät eine höhere Richtigkeit von 91,9% erzielt werden, was auf die höhere Anzahl an Messpunkten zurückgeführt werden konnte. Im Kontrast zum *benchtop* Gerät ist die Anwendung des mobilen NIR-Gerätes auf differenziertere Fragestellungen, wie die Unterscheidung der vier Länder, nur limitiert möglich. Es wurde mittels Klassifizierung durch eine LDA eine Richtigkeit von 71,8% erzielt. Dagegen war es mittels des laborbasierten Gerätes möglich die Länder mit einer Richtigkeit von 80,5% zu unterscheiden. Dieses Ergebnis verdeutlicht, dass das laborbasierte NIR-Spektrometer bei herausfordernden Fragestellungen dem kleineren mobilen NIR-Gerät überlegen ist.

Hinweis zum Copyright:

You are free to: Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

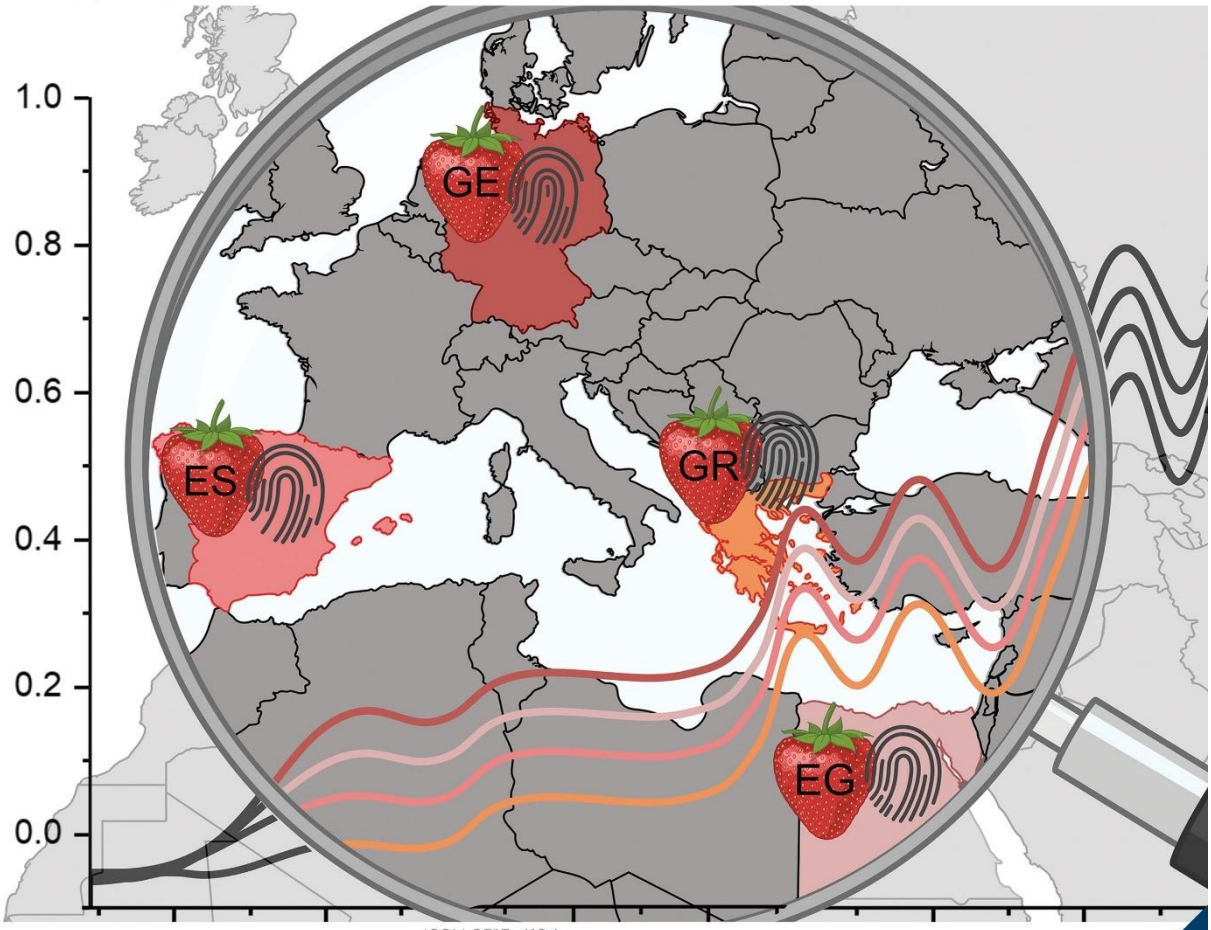
URL: https://pubs.rsc.org/en/content/articlehtml/2024/mo/d4mo00161c

DOI: https://doi.org/10.1039/D4MO00161C

Volume 21
Number 1
January 2025
Pages 1–102

Molecular Omics

rsc.li/molomics



ISSN 2515-4184

ROYAL SOCIETY OF CHEMISTRY

RESEARCH ARTICLE
Markus Fischer *et al.*
Competing technologies: determining the geographical origin of strawberries (*Fragaria* × *ananassa*) using laboratory based near-infrared spectroscopy compared to a simple portable device

Indexed in Medline!

Molecular Omics

ROYAL SOCIETY OF CHEMISTRY

RESEARCH ARTICLE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Mol. Omics*, 2025, 21, 7

Competing technologies: determining the geographical origin of strawberries (*Fragaria* × *ananassa*) using laboratory based near-infrared spectroscopy compared to a simple portable device†

Johannes Brockelt, ‡ Felix Schmauder,‡ Kim Brettschneider, Marina Creydt, Stephan Seifert and Markus Fischer *

The application and development of fast and simple screening methods for the authentication of foods has increased continuously in recent years. A widely used analytical technique is Fourier transform near-infrared spectroscopy (FT-NIR). Despite the simple application of FT-NIR analysis, the analyses are usually carried out on benchtop devices in the laboratory. However small, inexpensive and mobile NIR devices could be used on-site. Despite the simple use of FT-NIR analysis, the examinations are usually carried out on a stationary benchtop device in a laboratory. However, in order to be able to perform the application directly on site, the application of small, cost-effective and mobile NIR devices for food analysis is crucial. In this study, both, a benchtop NIR instrument and a handheld NIR device with a lower resolution and analyzed wavenumber range were applied for the differentiation of strawberries from different geographical origins. Distinguishing German and non-German strawberries using linear discriminant analysis (LDA) yielded an accuracy of 91.9% and 84.0% using the benchtop and the handheld devices, respectively. Relevant variables could be assigned to lipids, carbohydrates and proteins. Overall, our study demonstrated for the first time that analyzing the geographical origin of strawberries using NIR spectroscopy is also possible by means of a handheld device.

Received 18th August 2024,
Accepted 21st November 2024

DOI: 10.1039/d4mo00161c

rsc.li/molomics

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1. Introduction

Fourier transform near-infrared spectroscopy (FT-NIR) is an established method for simple and rapid authentication of food as it is a robust technique, which requires minimal sample preparation. It has proven to be a comprehensive analytical tool in omics technology and particularly in metabolomics. Compared to other spectroscopic techniques, NIR spectroscopy has limited resolution despite its broad applicability. In omics research, NIR spectroscopy is therefore often used as a complementary technology in combination with high-resolution techniques such as LC-MS.[1] Furthermore, it is an environmentally friendly method as no solvents or hazardous chemicals need to be used. For this reason, small, handheld NIR devices that are portable and can be operated with a laptop or smartphone have increasingly come on the market in recent years.[2] In order to assess the performance of this option for mobile analysis, a detailed comparison with already established benchtop devices is necessary. In recent studies, NIR applications have been used to analyse food ingredients, but also to answer more complex questions about authenticity. For example, analyses of the geographical origin of green coffee beans, almonds, soybeans and hazelnuts were successfully carried out.[3–6] For foods with a low water content, it is possible to measure either the whole food or a homogenized and freeze-dried powder. The additional workload of freeze-drying leads to better classification rates.[7] Determining the origin of water rich samples without freeze-drying, on the other hand, has not been possible up to now, as the large water bonds lead to unspecific spectra with little information.[8,9] For this reason, the established method of sample application was chosen to enable a comparison between different NIR devices. All of these example studies were performed on Benchtop NIR devices, which normally require a laboratory with a stable temperature. However, there are already a few comparative studies in the area of food

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany.
E-mail: markus.fischer@uni-hamburg.de; Fax: +49 40 42838-4342; Tel: +49 40 42838-4359

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4mo00161c

‡ Authors contributed equally to this work

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

authentication that were carried out with handheld devices. The first publication dealt with the adulteration of South African honey and compared several handheld NIR devices based on their classification performance, but only a small number of samples were used.[10] In addition, the performance of different handheld NIR devices was compared for the analysis of the adulteration of coriander seeds with salt, sawdust and starch with regression models.[11] Furthermore, handheld NIR and benchtop devices were applied to distinguish different truffle varieties with 100% accuracy, which was to be expected as the differences in truffle varieties are large, and also other analytical techniques achieved very good results.[12–15] Moreover, a trend to separate thirty truffle samples according to their geographical origin was observed.[12] Recently, a study was published in which FT-NIR analyses of Spanish ham from different types of pigs and feeds were compared using handheld and benchtop instruments. Sixty ham samples from two ham types were compared and achieved classification rates of 50–92% depending on the NIR device.[16] In all four comparative studies, different handheld devices were used and compared with the established benchtop devices. However, as far as we are aware, there are no studies that have used a large sample set to perform an in-depth comparison between a benchtop and a handheld NIR device for authentication of food regarding geographical origin.

Strawberries are a food that is grown and traded worldwide, with the price depending on their origin. It is therefore important to have a fast and cost-effective authentication process in place to differentiate between different geographical origins in order to prevent and detect food fraud. Important strawberry-growing countries for the European market are Spain, Greece, Germany and Egypt. Germany in particular has a high strawberry consumption and imports 127 000 tons of strawberries annually, making it an important target market for strawberry exporters.[17]

To date, almost all available handheld NIR devices measure reflection scattering, which is particularly suitable for powders and granulates.[18,19] The measurement of liquid samples, for which an NIR device based on transmission mode would be necessary, is not part of this study, as handheld devices based on transmission measurements have not yet achieved any commercial significance.[9] Furthermore, freeze-drying can avoid the negative influence of water in the NIR spectrum, which often superimposes relevant signals.[7]

The focus of this work is on the detailed comparison of a benchtop and a handheld NIR device using a dataset of 198 strawberries from four different countries (German 110 samples, Egyptian 22 samples, Greek 21 samples and Spanish 45 samples). In addition, the measuring ranges of the two NIR devices are analysed with regard to their different information content. For data evaluation, a comparison of the two NIR devices is carried out using principal component analysis (PCA) and a classification is performed using linear discriminant analysis (LDA). In addition, an analysis of variance (ANOVA) was carried out for each variable from the datasets in order to identify the most relevant wavelengths that enable the sample groups to be distinguished. In this way, both NIR devices can be compared on the basis of their spectra, classification rate and information content of the significant variables.

2. Materials and methods

2.1. Sample acquisition

A total of 198 strawberry samples were used for the study. Of those samples 110 German, and 22 Egyptian, 21 Greek and 45 Spanish. The acquisition of the German samples was carried out directly in the field, which guarantees that this is authentic sample material. For this purpose, 500 g of strawberries were collected from various rows in a field and from various plants, to cover the variance of one location as broad as possible. The samples of Egypt, Greek and Spain were provided by traders and distributors. In order to achieve a robust comparison of the two NIR devices, a sample pool was selected that covers various exogenous parameters such as different harvest years from 2021–2024, different varieties and regions (Table S1 in the ESI†), a summary of the information is shown in Table 1. To determine the influence of the water content of intact, homogenized and freeze-dried sample material in the NIR spectrum, 12 randomly selected strawberry samples were used.

2.2. Sample preparation

Sample preparation was started by removing the stems, quartering fruits and flash freezing it in liquid nitrogen. The quartered strawberries were stored at −20 °C until further processing. To homogenize the samples, they were mixed with the same amount of dry ice in a knife mill (Grindomix GM 300, Retsch, Haan, Germany). The following grinding program was used: first pre-grinding with the blade running backwards to protect it for 1 minute at 1000 rpm. Then the main grinding process took place for a total of 2 minutes at 4000 rpm. The dry ice was evaporated for approx. Three days by storage at −20 °C. In order to apply the NIR devices to strawberries, all samples were converted into a powder using freeze-drying. Freeze-drying was carried out by using a means of a freeze-dryer (Beta 2-8 LSCplus, Martin Christ, Osterode am Harz, Germany). This procedure has already been used in a large number of NIR studies and improves classification, even for matrices with low water content.[7] The samples were freeze-dried for a total of four days, with the samples being stirred using a glass rod on the third day to allow the water to escape completely. After the freeze-drying the samples were ground using a mortar (Haldenwanger, Waldkrainburg, Germany) and stored at −80 °C until

Table 1 Summary of the meta-information of the analyzed samples

Origin		Harvest year		Cultivation		Variety	
Egypt	22	2021	12	Organic	37	Clery	12
Germany	110	2022	71	Conventional	161	Asia	7
Greece	21	2023	107			Fortuna	6
Spain	45	2024	8			Verdi	5
						Alegro	4
						Malwina	4

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

further analysis. The sample preparation was identical for both NIR devices.

2.3. Near-infrared spectroscopy

The analysis of the 198 freeze-dried strawberries was performed according to standardized workflows of our group.[20] A total of 1.25 ± 0.10 g of the freeze-dried sample material was weighed directly into glass vials (52.0 mm × 22.0 mm × 1.2 mm, Nipro Diagnostics Germany GmbH, Ratingen, Germany) and tempered to 22 °C ± 2 °C. For analysis with a benchtop spectrometer, an instrument from Bruker with an integration sphere, which was controlled with the OPUS software (TANGO, Bruker Optics, Bremen, Germany) was used. Per spectrum 50 scans were recorded in a wavenumber range of 11 550–3950 cm^{-1} with a resolution of 4 cm^{-1}. Five technical replicates were measured per sample.

The novel NIR scanner "SenoCorder" from Senorics (Dresden, Germany) was used for the analysis of the 198 strawberry samples with a handheld NIR device. This device uses a Sensosense chip that emits the NIR radiation generated by two wolfram lamps and detects the reflected radiation from the sample. The device uses narrow-band sensors made of an organic, radiation-absorbing material, thus requiring only single wavenumbers and no monochromator.[21] There are 16 pixels located on the chip, which enables the analysis of 16 different wavenumbers simultaneously and measurements within 2 seconds. Hence, the data generated by this device contains 16 values (spacing between 30 and 36 nm) in the wavenumber range from 8446–5938 cm^{-1} (second overtone), as it has been shown that the most relevant information for such studies is in this range.[7,22] Due to the large distance between the measuring points, the handheld NIR has a low resolution with a S/N ratio of 4500 : 1. The NIR Scanner has automatic temperature control, which makes it ideal for robust use outside the laboratory. Ten technical replicates were measured for each sample to cover the sample's variance as best as possible. Between measurements of the replicates, the glass vial was shaken thoroughly and levelled evenly.

In addition to the freeze-dried samples, 12 intact strawberry samples and the same samples in homogenized form were measured with both NIR devices. For this purpose, NIR spectra were recorded from each individual strawberry at 5 points on the shell. The homogenized, non-freeze-dried strawberries were analysed by transferring 2.0 g of each into a glass vial. The NIR measurements of the 12 homogenized and freeze-dried samples were carried out in the same way as that of the freeze-dried samples.

2.4. Determination of the water content

To ensure that the freeze-drying process was carried out correctly, the water content of 12 strawberry samples was determined before and after freeze-drying using a drying balance (SMART 6, CEM Corporation, Matthews, USA). An intact strawberry from each sample was used to analyse the dry matter and resulting water content in the strawberry. To determine the water content of freeze-dried strawberries, 1.25 g of each sample processing stages was dried with the drying scale and the resulting water content was calculated.

2.5. NIR spectra pre-processing

For the data of the benchtop device, multiplicative scatter correction (MSC) was first applied to normalize the data and reduce scattering. After that, the data was binned, meaning that five adjacent variables were averaged into one bin, resulting in a reduction of the 3720 variables to 744 NIR wavenumber bins as binning reduces background noise. The formation of buckets does not necessarily result in the loss of information, as neighbouring variables are strongly correlated.[5] In the final step, the arithmetic mean of the five spectra of a sample was generated. For pre-processing of the data from the handheld NIR device, MSC and mean centering of 10 spectra of each sample was performed. The entire data pre-processing was performed using Unscrambler 11 (Aspen Technology Inc., Bedford, MA, USA) and MATLAB R2020b (The MathWorks Inc., Natick, MA, USA).

2.6. Statistical analysis

For the application of multivariate approaches, R (V.4.3.3) and R-Studio (V 2023.12.1) were used. PCA was performed with scaling using the package factoextra (V 1.0.7). For LDA the packages mlr3 (V 0.17.2), mlr3verse (V 0.2.8) and mlr3filters (V 0.7.1) were used. All classification models were validated with a 10-fold cross validation, which was repeated 100 times. The results of the 10-fold cross-validation and 100 repetitions were averaged. In the first run of a 10-fold cross-validation, the model is trained with 90% of the data set and tested with the remaining 10%. In the second run, the next 10% of the data set is used as test dataset. This continues until all data has been used as test dataset for validation. This procedure ensures that the method has been validated with different test data sets. To identify the important variables, an analysis of variance was used with the ANOVA-filter function of the package mlr3, which includes a correction for multiple testing. All variables with a p-value of 0.05 and lower were considered as significant.

3. Results and discussion

3.1. NIR spectra of strawberries

In order to determine the differences in the data, obtained from the two different NIR spectrometers, the averaged spectra were investigated. The measured water contents in Table S2 (ESI†) shows that freeze-drying has been carried out correctly. Fig. 1A shows the FT-NIR spectra of the freeze-dried strawberry samples measured with the FT-NIR benchtop device in the wavenumber range of 11 550–3950 cm^{-1}.

The spectra show signals in the range between 8600 cm^{-1} and 8200 cm^{-1}, which could be assigned to the second overtone of C–H stretching vibrations and the second overtone of the HC–CH stretching vibration, *e.g.*, of unsaturated fatty acids. Additionally, signals between 5800 and 5600 cm^{-1} could be associated with aliphatic hydrocarbons, potentially resulting

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

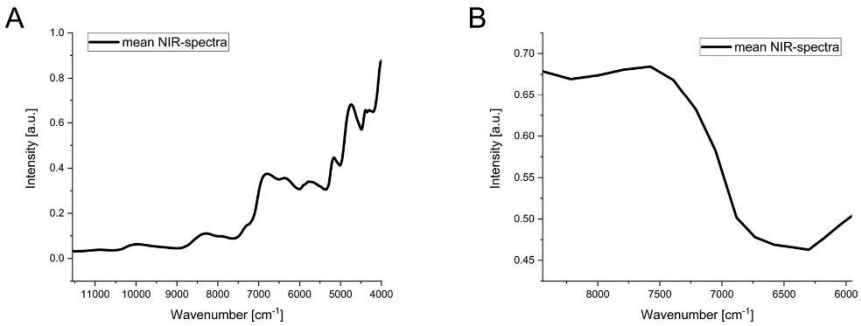


Fig. 1 NIR-spectra of freeze-dried strawberries obtained by the benchtop (A) and handheld (B) device.

from the first overtone of the symmetric CH_2 vibration and C–H stretching. Furthermore, bands at 4350 cm^{-1} could be caused by the second overtone of the C–H stretching and bending vibration of C–H_2. Bands between 7000 and 6000 cm^{-1} are also visible and could be associated to the first overtone of the C–H stretching vibration and the third overtone of the C=O stretching vibration. In addition, they could also be assigned to the first overtone of the O–H stretching vibration and the N–H stretching vibration of proteins, and the first overtone of the symmetric and asymmetric O–H stretching vibrations. These absorption bands are characteristic of amines, amides and proteins. The signal at ~5200 cm^{-1} could be assigned to the O–H stretching vibration and the H–O–H bending vibration of water. Bands at 4800 cm^{-1} could be assigned to O–H, C–O and C = O–O stretching vibrations as well as to a combination band of $CONH_2$, which is typical for carbohydrates and proteins.[7,23]

Fig. 1B shows the NIR spectra of freeze-dried strawberries obtained by the handheld device covering the spectral range between 8500 cm^{-1} and 6000 cm^{-1}, which can be associated to the second overtone of O–H, N–H and the C–H stretching vibrations of proteins, water and lipids 23. Since the handheld NIR device only records 16 data points with a comparatively small wavenumber range, a detailed interpretation of the spectra was only possible to a limited extent. For this reason, interpretation of detailed information about the chemical profile of strawberries in the spectra is only possible with the benchtop FT-NIR.

3.2. Principal component analysis

Principal component analysis was applied to the datasets obtained by the benchtop and handheld device containing 744 and 16 variables, respectively. The scores of the principal components (PCs) 1 and 2, 1 and 3, as well as the loadings of PC 1, PC 2 and PC 3 are shown in Fig. 2 for the data obtained by the benchtop and handheld device.

For the analysis of the data of the benchtop device, the predominantly negative values of PC 3 (Fig. 2C) can be used to separate the German samples from the non-German samples with predominantly positive values. Looking at the same plot with the colouring of the non-German samples according to specific countries (Fig. 3C), it can be seen that many of the non-German samples with negative values for the scores of the third principal components are Egyptian samples.

The differentiation between German and Egyptian samples is therefore difficult on the basis of PCA, while the separation of German from Spanish and Greek samples is clearly possible by the scores of the third principal component. The loadings of this principal component (Fig. 2I) shows high values in the spectral regions 8600–8300 cm^{-1}, 6900–6600 cm^{-1}, 5900–5200 cm^{-1} and 4900–4800 cm^{-1} associated with lipids, proteins and carbohydrates (see previous section) leading to the conclusion that different molecule classes are responsible for this differentiation. The scores of the first two PCs (Fig. 2 and 3A), which account for more than 80% of the explained variance, however, do not allow a distinction to be made between the samples of the different geographical origins. This plot shows a broad distribution of the German samples, which could be explained by a comparatively high diversity of the German samples. This could be caused by the broad geographical origin of the samples within Germany since the samples originated from 13 different federal states or by the great diversity of the samples in terms of taxonomic varieties since the German samples contained 41 different varieties.

The results of the PCA applied to the dataset of the NIR handheld device are shown in Fig. 2B, D, F, H and J. Most of the German samples show positive values for the first PC and negative values for the second PC, while non-German samples are characterized by the opposite (Fig. 2B). In contrast to the data from the benchtop device, the main variances of the dataset, adding up to about 79%, are therefore relevant for the specific characterization of the German samples, while PC 3 does not contain any relevant information (Fig. 2D). Again, however, when the non-German samples are coloured according to their country of origin, it is evident that Egyptian samples in particular have similar scores to German samples, and that Greek and Spanish show quite similar spectra. Thus, the PCA of the data from the handheld device also allows a distinction to be made between the two groups of German and Egyptian and

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

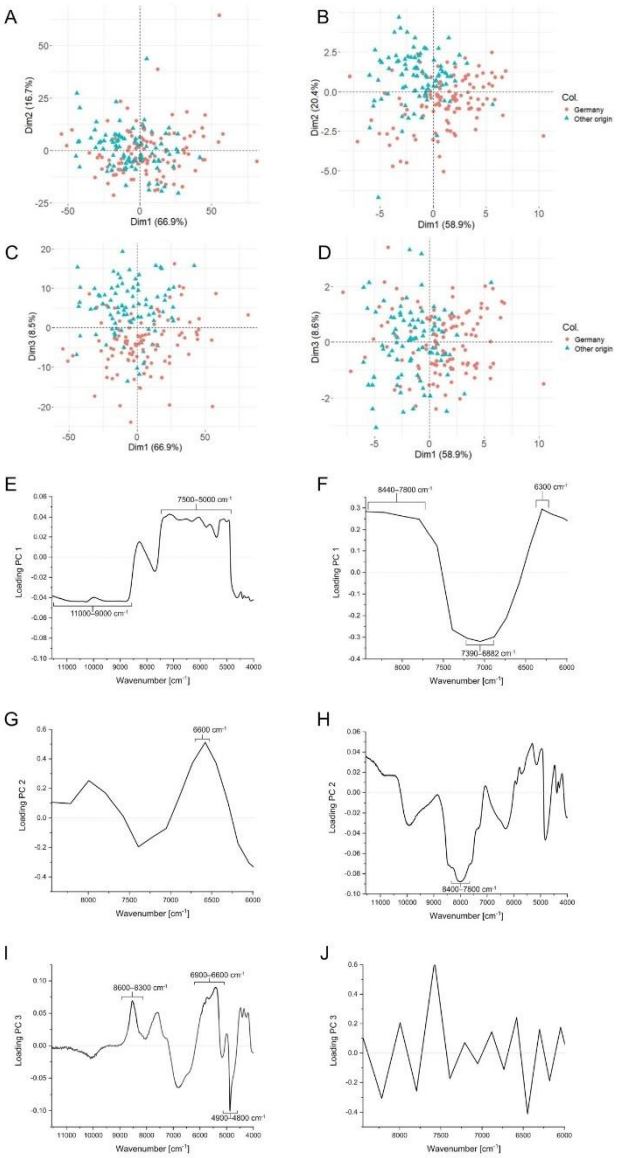


Fig. 2 Results of the PCA conducted on the NIR spectra obtained by the benchtop (A + C + E + G + I) and handheld (B + D + F + H + J) device. The scores of the first two principal components (A + B), the scores of the first and third principal component (C + D), as well as the loadings (E + F + G + H + I + J) of the first, second and third principal component are shown, with the labelled maxima and minima. The scores were coloured according to the origin of the samples from Germany or other countries (Egypt, Greece and Spain).

Spanish and Greek samples rather than a precise distinction between the spectra of the individual countries. The loadings show that the spectral regions 8440–7800 cm^{-1}, 7390–6882 cm^{-1} and around 6300 cm^{-1} (Fig. 2F) assigned to aliphatic

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

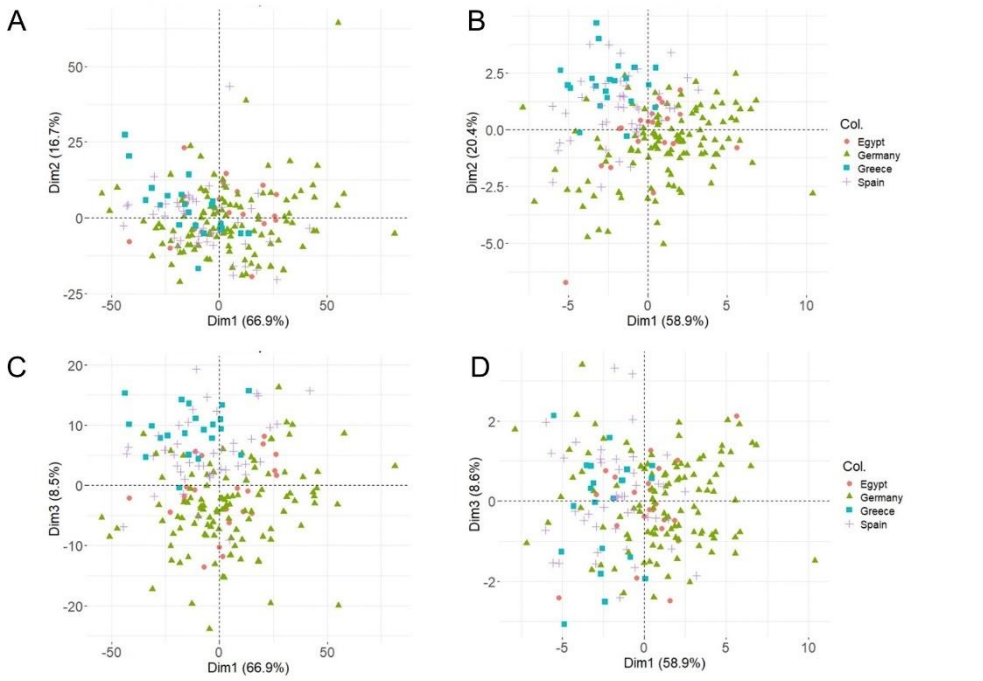


Fig. 3 Results of the PCA conducted on the NIR spectra of the benchtop (A + C) and handheld (B + D) device. The scores (A + B) of the first two principal components as well as the first and third principal component (C + D) are shown and the scores were coloured according to the origin of the samples from Egypt, Germany, Greece and Spain.

Table 2 Results of the classification for the differentiation of German and non-German samples analysed with the benchtop and handheld device. An LDA with 10-fold cross validation and 100 repetitions was applied and the mean values of all repetitions in percent are shown together with the total number of samples

		Response class			
			Germany [%]	Other origin [%]	Number of samples
True class	Benchtop	Germany [%]	93.6	6.4	110
		Other origin [%]	10.1	89.9	88
	Handheld	Germany [%]	83.7	16.3	110
		Other origin [%]	15.6	84.4	88

hydrocarbons and proteins (see previous section), and around 6600 cm^{-1} (Fig. 2H) assigned to proteins have a great influence on the first and second PC, respectively. The analyses of the samples with the handheld device therefore indicate that proteins in particular are relevant for differentiating between German/Egyptian and Spanish/Greek samples.

When using unsupervised PCA, no complete separation between all the strawberry samples of the different origins could be obtained, neither by using the data of the benchtop nor the handheld device. Hence, a supervised approach was applied for classification in the following sections.

3.3. Differentiation of German and non-German strawberries

For differentiation of German and non-German samples, LDA was applied, and the results are shown in Table 2. The classification based on data of the benchtop device resulted in an accuracy of 91.9% with a correct classification rate of 93.6 and 89.9% for German and non-German samples, respectively. For the differentiation based on the data of the handheld device, an accuracy of 84.0% was obtained and 83.7 and 84.4% of the German and non-German samples were correctly classified, respectively. As expected, the benchtop device achieves a better accuracy for the identification of German samples, but

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

the results of the much simpler and cheaper handheld device are also comparatively good.

3.4. Differentiation by country of origin

Subsequently, the potential of NIR to distinguish specific geographical origins was analysed and four-class models were trained with LDA to distinguish between strawberry samples from the four different countries. The results are shown in Table 3. The accuracy for the benchtop device was 80.5%. Egyptian and German samples were classified with an accuracy of 79.2% and 93.3%, respectively, while 16.4% of the Egyptian samples were wrongly assigned to the German group. Therefore, mainly Egyptian samples were classified as German, while misclassifications were rare for the other two classes. The reason for this could be that German samples are overrepresented in the dataset and in these cases machine learning methods often tend to incorrectly assign samples to these groups. However, the fact that approximately 5% of the German samples are classified as Egyptian also indicates a similarity between the NIR spectra of these two classes, which was also observed in the PCA in the previous section. The accuracies of the Greek and Spanish samples were 42.4 and 69.1%, respectively, with frequent misclassification of 49.2% and 24.4% between those two classes. This could be explained by the similarity of the sample composition caused by similar climate conditions of the Spanish and Greek growing regions as Spanish samples originated mainly from the Huelva region and the Greek samples from the Pyrgos region. Both regions are located directly on the Mediterranean and at approximately 37° north latitude, and are characterized by a very similar maritime climate. The similar climatic conditions in Spain and Greece go hand in hand with similar cultivation methods. On large farms, strawberries are grown using irrigation systems that are very much geared towards large production volumes. In contrast, Germany and Egypt have positioned themselves as producers of high-quality strawberries, which in turn leads to similar growing methods. Several studies have already shown that similar exogenous influences such as climatic conditions can lead to similar plant metabolomes.[24–26] For example, Hou *et al.* demonstrated that high temperatures change the glycerolipid composition of soybeans by reducing the proportion of polyunsaturated fatty acids.[27–29]

The results from the handheld device (Table 3B) show a total classification accuracy of 71.8% and 52.7, 82.7, 55.3 and 57.1% of the Egyptian, German, Greek and Spanish samples were correctly assigned. Hence, the accuracies are obviously lower than for the analysis of the data obtained by the benchtop device. The results of the handheld device, however, show the same misclassification patterns, as German and Egyptian and Greek and Spanish samples were often misclassified among each other. It can therefore be concluded that the data from the handheld device are limited for the differentiation of strawberry samples from individual countries of origin.

Overall, the reason for the comparatively low classification with regard to individual countries of origin could be that the variance of the individual groups is not sufficiently represented in the model. This could be particularly the case if samples from Spain and Greece, which are very similar, are to be differentiated and could improve if an even larger number of samples is used to train the models. This is supported by the fact that a distinction between German and non-German samples, which is based on a much larger number of samples, is possible with higher accuracy. In addition, the misclassifications of 21.3 and 44.1% between the Greek and Spanish classes could be caused by the similar climatic and geographic conditions, as discussed above. In addition, the dataset was also checked using a completely independent test dataset. However, fully independent test datasets are considered a high standard for method verification. The results of the independent test datasets for both NIR devices can be found in the electronic appendix in Tables S5 and S6 (ESI†). With the benchtop device, 95.0% and 80.0% accuracy were achieved for the two- and four-class models, respectively. For the NIR handheld device, the accuracies were 85.0% and 70.0%. Due to the total number of samples, the test datasets are only based on a smaller number of samples, which is a limitation of validation using test datasets. These accuracies were comparable to the validation using repeated cross validation which demonstrated that both validation methods lead to similar results. In the independent test datasets, the same confusions between the Greek and Spanish samples and Egyptian and German samples were

Table 3 Results of the classification for the differentiation by country of origin analysed with the benchtop and handheld device. An LDA with 10-fold cross validation and 100 repetitions was applied and the mean values of all repetitions in percent are shown together with the total number of samples

			Response class				
			Egypt	Germany	Greece	Spain	Number of samples
A							
True class	Benchtop	Egypt	79.2	16.4	0.0	4.4	22
		Germany	4.9	93.3	0.4	1.5	110
		Greece	0.5	7.9	42.4	49.2	21
		Spain	1.9	4.8	24.2	69.1	45
B							
True class	Handheld	Egypt	52.7	36.1	0.0	11.2	22
		Germany	10.3	82.7	0.5	6.5	110
		Greece	0.0	0.6	55.3	44.1	21
		Spain	5.5	16.0	21.3	57.1	45

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

found, which were already determined during the cross validation and have already been discussed. Due to the better comparability of the cross-validation, this was used for the class comparison.

3.5. Selection and interpretation of the relevant variables

In order to identify the variables relevant for the classification, an ANOVA was performed, and the results are reported in Tables S2 and S3 (ESI†). Fig. 4A and B show the averaged spectra with the respective variables marked that have a *p*-value of less than 0.05 and are therefore relevant for the classification of geographical origin.

The benchtop FT-NIR spectra (Fig. 4A) shows two relevant spectral regions between 7250 and 6500 cm^{-1}, as well as between 5250 and 4250 cm^{-1}. The fact that these spectral regions are indeed relevant for the differentiation of geographical regions is also confirmed by the spectra of the individual countries shown in Fig. 4C, since the averaged spectra of German and Egyptian samples show higher intensities in the range between 7250 and 6500 cm^{-1}, while the spectra of Spanish and Greek samples show higher intensities in the range between 4750 and 4250 cm^{-1}. Of these two relevant regions, the handheld device only detects the region around 7000 cm^{-1}, which is also marked as relevant in Fig. 4C. The fact that no information about the spectral range between 5250 and 4250 cm^{-1} is obtained by this device is probably the reason for the lower accuracy of the differentiation of the geographical origins.

In order to examine the class differences in more detail, representative variables of the different datasets were analysed. Boxplots of those variables are shown in Fig. 5 for the data obtained by the benchtop device. The variable at 4340.82 cm^{-1} (Fig. 5A), which could be assigned to aliphatic hydrocarbons, *e.g.* from lipids, generally shows lower intensities for German samples and higher intensities for non-German samples. However, the distinction is not entirely accurate as German and non-German samples also show high and low intensities respectively. Considering the intensity values of the individual countries (Fig. 5D), it becomes clear that many Egyptian samples have low intensity values, just like the German samples. In addition, it is evident from this figure that the Spanish samples show slightly lower intensities than the Greek samples, but there is also a large overlap between the intensities of these two groups. The boxplot of this variable thus shows a relatively clear distinction between the two groups of samples from Germany/Egypt and Spain/Greece, but also overlaps within these groups. The same can be observed for the variable at 6902.04 cm^{-1} (Fig. 5F) which was assigned to proteins, as

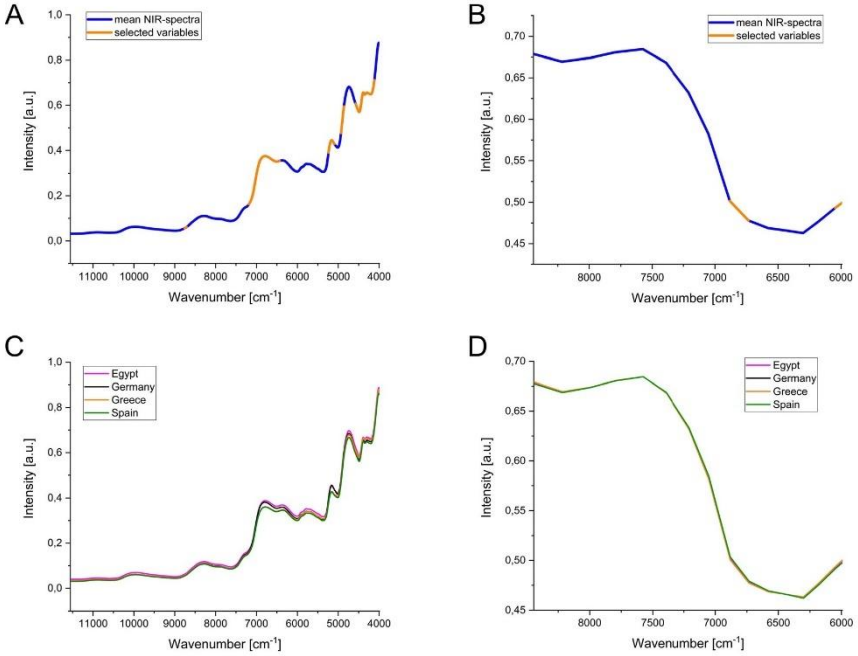


Fig. 4 Mean NIR-spectra obtained by the benchtop FT-NIR device (A) and the handheld NIR device (B) with highlighted via ANOVA selected significant variables (orange). In addition, the spectra of the benchtop FT-NIR device (C) and the handheld NIR device (D) averaged by country are shown and the relevant area are expanded in detail.

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

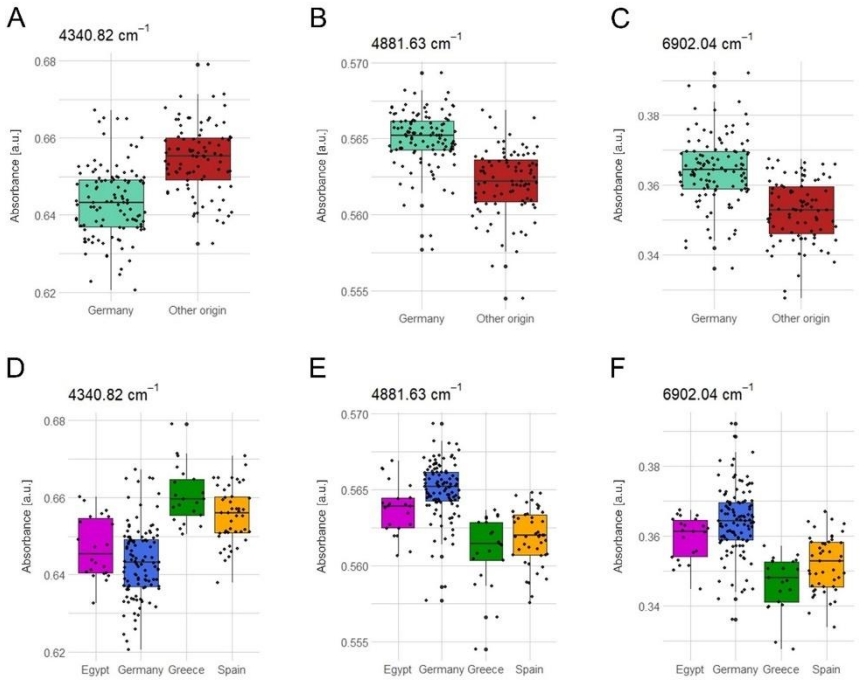


Fig. 5 Boxplots of the variables at 4340.82, 4881.62 and 6902.04 cm^{-1}, which were selected for the differentiation of the different geographical origins of strawberries using the benchtop device. The respective intensities for the two-class (A–C) and four-class models (D–F) are shown.

the former group shows comparatively low intensities and the latter group high intensities, and is also evident from the classification results in the previous sections and the results of the PCA in Section 3.2.

For other foods, it has already been shown that the protein band is relevant for classification. For example, an accuracy of 90.0% was achieved in grain maize solely on the basis of the origin classification using the protein band.[30] This is possible because samples from individual countries can differ significantly based on their protein and fat content. In the case of strawberries, it has already been shown that different protein concentrations are present for different cultivation methods, while the variety has only a limited influence. As the cultivation methods vary in different countries and the peak season takes place in different months, this could have an impact on the content of different proteins and fats. In addition, different NaCl contents of the soil can affect the amino acid metabolism. In strawberries, it has already been shown that increased NaCl stress is associated with a low protein content. The soil therefore has an influence on amino acid metabolism and protein profiles.[31]

For the analysis of the data obtained by the handheld device, boxplots of the two variables with the lowest *p*-values at 5938 cm^{-1} and 6734 cm^{-1} are shown in Fig. 6A–D. These variables, which could be assigned to proteins, also allow a comparatively clear distinction between Egyptian and German and Spanish and Greek samples. The variable a 5938 cm^{-1} (Fig. 6A) is characterized by relatively high values for the first group and relatively low values for the second group, while the variable at 6734 cm^{-1} (Fig. 6B) shows the opposite. Considering these variables, the comparatively successful separation of German and non-German samples using the handheld device, is comprehensible.

Hence since no clearer distinction between the different classes can be recognized based on the variables of the data obtained by the benchtop device, the better performance seems to be achieved mainly by a combination of different variables also including those from the spectral range that is not covered by the handheld device. However, it should be noted that a limitation of the dataset is that only 12 samples from Germany were available for 2021, so that a classification based on the harvest year cannot be completely ruled out. To minimize this influence, additional samples were taken from all four countries, including Germany, in the remaining three years. Another parameter that can influence the classification is the cultivation of organic strawberries. Only 37 strawberries from Spain that were produced for the German market are organic. In contrast, there were no organic samples from Egypt

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

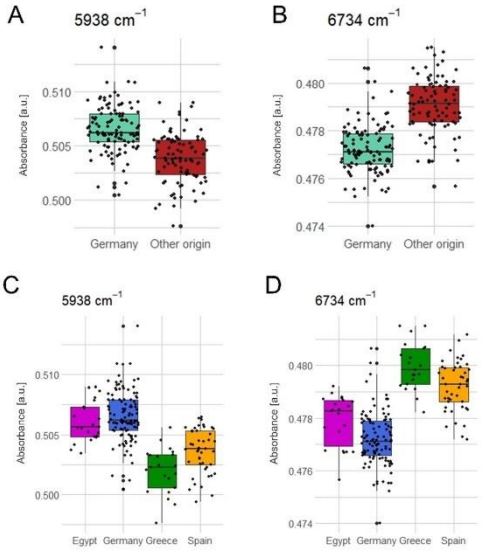


Fig. 6 Boxplots of the variables at 5938 and 6734 cm^{-1}, which were selected for the differentiation of the different geographical origins of strawberries using the handheld device. The respective intensities for the two-class (A and B) and four-class models (C and D) are shown.

and Greece. The cultivation of organic food is of great importance in Germany and is therefore part of a standard market sample dataset. However, it cannot be ruled out that samples are categorized as German or Spanish due to their organic cultivation. However, the low risk of confusion between German and Spanish samples suggests that this effect does not exist.

The variety is only known in about half of the samples, with the most common varieties being Clery with 12, Asia with 7 and Fortuna with 6 samples. The difference in variety could also be another parameter that influences the classification. For example, the varieties Clery and Asia are only present in German samples, while Fortuna is present in Egyptian, Greek and Spanish samples. Nevertheless, the varieties are adapted to local climatic conditions and harvest times, so that it is not possible to obtain samples of the same varieties from all regions. The varieties are therefore part of the regional authenticity of the samples, but an influence of the different varieties on the classification cannot be excluded and can only be minimized by obtaining as many varieties as possible. Comparing the results of the origin classification with the results of other water-rich matrices, some similarities can be seen. For example, asparagus shows a similar tendency towards classification with 88.5% accuracy in 5 classes. This classification model is performing better than the 4-class model shown here with 80.5%, but asparagus is a more homogeneous matrix. Only a few varieties of asparagus are grown commercially and the cultivation methods are very similar between countries, which leads to a lower variance and therefore better classification.

Classification between Korean and non-Korean soybeans was also successful with a correctness of 95.9%. This model shows similar classification rates to the 91.9% classification rate of German and non-German strawberries shown here. Furthermore, the protein and fat bands of the NIR spectrum were recognized as relevant for the classification of soybeans. Freeze-drying was also used for less water-rich foods such as soybeans so that the relevant protein and fat bands were not obscured by the water bands. This shows that the processing of the samples, the results and detection of relevant wavelengths are consistent with previous studies. Other established methods for determining the origin of food, such as LC and GC coupled with mass spectrometry, have a higher resolution and therefore potentially distinguish better between countries.[32,33] However, these methods are associated with greater technical effort, such as higher consumption of chemicals.[34] For this reason, the trend towards chemical-free analysis such as FT-NIR is necessary in terms of sustainability. The NIR method developed here could be utilized as a rapid screening method in routine laboratories to check for suspicious samples. In addition, developed method could also be used for incoming goods inspection. It is particularly well suited for the application of a fast and environmentally friendly method.

Overall, the results show that for a rough differentiation between German and non-German strawberries, a handheld device measuring only a few variables is sufficient to perform a quick and initial analysis on site. However, for a more precise classification of the geographical origin of strawberries, a benchtop device with higher resolution and a wider spectral range should be used.

4. Conclusion

In this study, the classification of strawberries according to their geographical origin was analysed with NIR spectroscopy using a benchtop and a handheld device. The separation of German and non-German samples showed an accuracy of 91.7% and 84% for the benchtop and handheld device, respectively, while the differentiation between individual countries of origin showed classification accuracies of 80.5% and 71.8%, respectively. A more detailed analysis of the relevant variables showed that the better performance of the benchtop instrument was due to both higher resolution and a larger spectral range analysed. The handheld device offers the advantage of lower acquisition costs and a laboratory-independent use, while the benchtop NIR device proved to be superior for the classification of complex questions. However, based on our results, it seems possible to use the handheld device for a quick and rough on-site analysis to determine whether the strawberries originate from Germany or not.

Author contributions

Conceptualization, J. B., F. S. and M. F.; methodology, J. B. and F. S.; software, J. B. and F. S. Validation, J. B. and F. S.; formal analysis, J. B., F. S., K. B. and S. S.; investigation, J. B., F. S., and

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

M. C.; resources, S. S. and M. F.; data curation, J. B. and F. S.; writing – original draft preparation, J. B. and F. S.; writing – review & editing, K. B., S. S., M. C., and M. F.; visualization, J. B. and F. S.; supervision, M. F.; project administration, M. F.; funding acquisition, M. F., S. S.

Data availability

The datasets of the NIR spectroscopy were created as part of the current study are available in the Zenodo repository at: **https://doi.org/10.5281/zenodo.13331169**.

Conflicts of interest

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

Acknowledgements

This IGF Project (01IF22909N) of the research association FEI, Godesberger Allee 125, 53175 Bonn, is/was funded as part of the program for promoting joint industrial research (IGF) by the Federal Ministry for Economic Affairs and Climate Protection based on a resolution of the German Bundestag. The authors thank Robin Dammann, Lina Cvancar, Nils Wax, Marie-Sophie Müller and Christian Ahlers for their help with sampling and helpful discussions.

Notes and references

1 H. Lösel, J. Brockelt, F. Gärber, J. Teipel, T. Kuballa, S. Seifert and M. Fischer, Comparative Analysis of LC-ESI-IM-qToF-MS and FT-NIR Spectroscopy Approaches for the Authentication of Organic and Conventional Eggs, *Metabolites*, 2023, **13**, 882.
2 J. Müller-Maatsch and S. M. van Ruth, Handheld devices for food authentication and their applications: A review, *Foods*, 2021, **10**, 2901.
3 A. Giraudo, S. Grassi, F. Savorani, G. Gavoci, E. Casiraghi and F. Geobaldo, Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis, *Food Control*, 2019, **99**, 137–145.
4 B. Richter, M. Rurik, S. Gurk, O. Kohlbacher and M. Fischer, Food monitoring: Screening of the geographical origin of white asparagus using FT-NIR and machine learning, *Food Control*, 2019, **104**, 318–325.
5 J. H. Lee, J. M. An, H. J. Kim, H. C. Shin, S. H. Hur and S. H. Lee, Rapid discrimination of the country origin of soybeans based on FT-NIR spectroscopy and data expansion, Food Analytical, *Methods*, 2022, **15**, 3322–3333.
6 N. Shakiba, A. Gerdes, N. Holz, S. Wenck, R. Bachmann, T. Schneider, S. Seifert, M. Fischer and T. Hackl, Determination of the geographical origin of hazelnuts (Corylus avellana L.) by Near-Infrared spectroscopy (NIR) and a Low-Level Fusion with nuclear magnetic resonance (NMR), *Microchem. J.*, 2022, **174**, 107066.
7 M. Arndt, M. Rurik, A. Drees, K. Bigdowski, O. Kohlbacher and M. Fischer, Comparison of different sample preparation techniques for NIR screening and their influence on the geographical origin determination of almonds (Prunus dulcis MILL.), *Food Control*, 2020, **115**, 107302.
8 H. Lösel, N. Shakiba, S. Wenck, P. Le Tan, M. Arndt, S. Seifert, T. Hackl and M. Fischer, Impact of Freeze-Drying on the determination of the geographical origin of almonds (Prunus dulcis Mill.) by near-infrared (NIR) spectroscopy, Food Analytical, *Methods*, 2022, **15**, 2847–2857.
9 L. Wang, D.-W. Sun, H. Pu and J.-H. Cheng, Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments, *Crit. Rev. Food Sci. Nutr.*, 2017, **57**, 1524–1538.
10 A. Guelpa, F. Marini, A. Du Plessis, R. Slabbert and M. Manley, Verification of authenticity and fraud detection in South African honey using NIR spectroscopy, *Food Control*, 2017, **73**, 1388–1396.
11 C. McVey, U. Gordon, S. A. Haughey and C. T. Elliott, Assessment of the analytical performance of three near-infrared spectroscopy instruments (benchtop, handheld and portable) through the investigation of coriander seed authenticity, *Foods*, 2021, **10**, 956.
12 C. Kappacher, B. Trübenbacher, K. Losso, M. Rainer, G. K. Bonn and C. W. Huck, Portable vs. Benchtop NIR-sensor technology for classification and quality evaluation of black truffle, *Molecules*, 2022, **27**, 589.
13 M. Creydt and M. Fischer, Food authentication: Truffle species classification by non-targeted lipidomics analyses using mass spectrometry assisted by ion mobility separation, *Mol. Omics*, 2022, **18**, 616–626.
14 K. Losso, H. Wörz, C. Kappacher, S. Huber, T. Jakschitz, M. Rainer and G. K. Bonn, Rapid quality control of black truffles using Direct Analysis in Real Time Mass Spectrometry and Hydrophilic Interaction Liquid Chromatography Mass Spectrometry, *Food Chem.*, 2023, **403**, 134418.
15 T. Segelke, S. Schelm, C. Ahlers and M. Fischer, Food authentication: Truffle (Tuber spp.) species differentiation by FT-NIR and chemometrics, *Foods*, 2020, **9**, 922.
16 M. Hernández-Jiménez, I. Revilla, A. M. Vivar-Quintana, J. Grabska, K. B. Beć and C. W. Huck, Performance of benchtop and portable spectroscopy equipment for discriminating Iberian ham according to breed, *Curr. Res. Food Sci.*, 2024, 100675.
17 *Statistisches Bundesamt, Import von Erdbeeren*, **https://www-genesis.destatis.de/genesis/online**, (accessed 31.03-2024).
18 H. Cen and Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, *Trends Food Sci. Technol.*, 2007, **18**, 72–83.
19 M. C. Pasikatan, J. L. Steele, C. K. Spillman and E. Haque, Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials, *J. Near Infrared Spectrosc.*, 2001, **9**, 153–164.

Open Access Article. Published on 22 November 2024. Downloaded on 2/18/2025 10:13:42 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

20 A. Drees, J. Brockelt, L. Cvancar and M. Fischer, Rapid determination of the shell content in cocoa products using FT-NIR spectroscopy and chemometrics, *Talanta*, 2023, **256**, 124310.
21 H. Yan, M. de Gea Neves, I. Noda, G. M. Guedes, A. C. Silva Ferreira, F. Pfeifer, X. Chen and H. W. Siesler, Handheld near-infrared spectroscopy: State-of-the-art instrumentation and applications in material identification, food authentication, and environmental investigations, *Chemosensors*, 2023, **11**, 272.
22 M. Arndt, A. Drees, C. Ahlers and M. Fischer, Determination of the geographical origin of walnuts (Juglans regia L.) using near-infrared spectroscopy and chemometrics, *Foods*, 2020, **9**, 1860.
23 J. Workman and L. Weyer, *Practical guide and spectral atlas for interpretive near-infrared*, CRC, 2012.
24 Deutscher Wetterdienst, *Climate Table Athen*, Attika, Greece, **https://www.dwd.de/DWD/klima/beratung/ak/ak_167160_kt.pdf**, (accessed 19 March 2024).
25 Deutscher Wetterdienst, *Climate Table of Huelva*, Andalusia, Spain, **https://www.dwd.de/DWD/klima/beratung/ak/ak_083830_kt.pdf**, (accessed 19 March 2024).
26 D.-M. Ma, S. V. S. Gandra, R. Manoharlal, C. La Hovary and D.-Y. Xie, Untargeted metabolomics of Nicotiana tabacum grown in United States and India characterizes the association of plant metabolomes with natural climate and geography, *Front. Plant Sci.*, 2019, **10**, 470291.
27 R. F. Wilson, Seed composition, *Soybeans: improvement, production, and uses*, 2004, vol. 16, pp. 621–677.
28 G. Hou, G. R. Ablett, K. P. Pauls and I. Rajcan, Environmental effects on fatty acid levels in soybean seed oil, *J. Am. Oil Chem. Soc.*, 2006, **83**, 759–763.
29 R. G. Upchurch, Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress, *Biotechnol. Lett.*, 2008, **30**, 967–977.
30 D. Schütz, J. Riedl, E. Achten and M. Fischer, Fourier-transform near-infrared spectroscopy as a fast screening tool for the verification of the geographical origin of grain maize (Zea mays L.), *Food Control*, 2022, **136**, 108892.
31 A. J. Keutgen and E. Pawelzik, Quality and nutritional value of strawberry fruit under long term salt stress, *Food Chem.*, 2008, **107**, 1413–1420.
32 P. Zhong, X. Wei, X. Li, X. Wei, S. Wu, W. Huang, A. Koidis, Z. Xu and H. Lei, Untargeted metabolomics by liquid chromatography-mass spectrometry for food authentication: A review, *Compr. Rev. Food Sci. Food Saf.*, 2022, **21**, 2455–2488.
33 G. Sammarco, D. Bardin, F. Quaini, C. Dall'Asta, J. Christmann, P. Weller and M. Suman, A geographical origin assessment of Italian hazelnuts: Gas chromatography-ion mobility spectrometry coupled with multivariate statistical analysis and data fusion approach, *Food Res. Int.*, 2023, **171**, 113085.
34 M. Tobiszewski, Metrics for green analytical chemistry, *Anal. Methods*, 2016, **8**, 2993–2999.

 This journal is © The Royal Society of Chemistry 2025

7 Zusammenfassende Diskussion

In dieser Arbeit wurde die Möglichkeit des Einsatzes der DART-MS-basierten Metabolomanalyse für die Herkunftsbestimmung von Lebensmitteln untersucht und zusätzlich ein Vergleich zu einer NIR-basierten Methode gezogen. Im ersten Abschnitt wurde ein Arbeitsablauf für die DART-MS-basierte *non-targeted* Metabolomanalyse entwickelt. Dieser Arbeitsablauf wurde dann auf drei unterschiedliche Lebensmittelmatrices, als Vertreter für wasserreiche, kohlenhydratreiche und fettreiche Lebensmittel angewendet. Im letzten Schritt wurde die Herkunftsanalyse von Erdbeeren mittels FT-NIR Spektroskopie durchgeführt, um zwei schnelle analytische Ansätze vergleichen zu können.

7.1 Herkunftsauthentifizierung von Erdbeeren

Jährlich werden weltweit circa 9 Millionen Tonnen Erdbeeren geerntet, die in Deutschland zu den beliebtesten Obstsorten zählen (Bundesanstalt für Landwirtschaft und Ernährung 2024; Quarshi et al. 2023). Da Deutschland seinen Erdbeerbedarf nicht selbst decken kann, wurden im Jahr 2023 circa 118.600 t frische Erdbeeren hauptsächlich aus Spanien, Griechenland, Polen, den Niederlanden und Ägypten importiert (Statistisches Bundesamt 2024). In den verschiedenen Lieferländern existieren unterschiedliche Preisstrukturen bei der Erzeugung, was zu deutlichen Preisunterschieden führt. Dies macht es lukrativ, die Herkunft von Erdbeeren zur Gewinnmaximierung zu verfälschen, was einerseits eine Verbrauchertäuschung darstellt sowie anderseits den redlichen Lebensmittelunternehmen schadet (Nassar et al. 2020). Dieser Umstand macht die Entwicklung neuer analytischer Methoden für den Herkunftsnachweis von Lebensmitteln notwendig. Dabei diente in der vorliegenden Arbeit die Erdbeere als Beispiel, um die Herkunftsauthentifizierung von wasserreichen Lebensmitteln mittels eines DART-MS-basierten Ansatzes zu erproben.

7.1.1 Methodenoptimierung

Um das beste Vorgehen zur Identifizierung der Herkunft von Erdbeeren zu ermitteln, wurden die Metaboliten aus 62 gefriergetrockneten Proben auf zwei verschiedene Weisen extrahiert. Einmal mittels einer polaren und einmal mittels einer unpolaren Extraktionsmethode, um sowohl die polar als auch die unpolar extrahierbaren Metabolite zu analysieren. Hierfür wurden Methanol als polares Extraktionsmittel und die von Creydt et al. modifizierte Extraktion nach Bligh and Dyer als unpolare Extraktionsmethode verwendet (Creydt et al. 2018a; Bligh und Dyer 1959). Polare wässrige Extraktionsmethoden wurden verworfen, da wässrige Extrakte aufgrund der längeren Verdampfungszeit des Lösungsmittels, die Applikation in der DART-Ionenquelle erschwerten. Die erhaltenen Extrakte wurden mittels der in Abschnitt 4.5.2.2.2 vorgestellten DART-qToF-MS-Kopplung, sowohl im positiven als auch im negativen

Ionisierungsmodus gemessen. Aus den zwei verschiedenen Extraktionsmethoden und den zwei Ionisierungsmodi ergaben sich somit vier verschiedene Kombinationen.

Die Gesamtzahl der detektierten *features* betrug für die Kombination polar-positiv 469 und für die polar-negative Kombination 554. Im Gegensatz dazu wurden bei den unpolaren Extraktionen 1118 *features* bei der positiven und 1563 bei der negativen Ionisation detektiert. Somit ergaben sich bei den unpolaren Extraktionen zwei- bis dreimal mehr detektierte *features* als bei der polaren Extraktion. Die multivariate Auswertung, eine *Random Forest* Klassifizierung, ergab Richtigkeiten der unpolar-positiven Kombination von 72,1% und der unpolar-negativen von 69,4%. Die polaren Extraktionen ergaben mittels positiver Ionisation eine Richtigkeit von 63,5% und 62,9% im negativen Ionisationsmodus. Diese Ergebnisse zeigen, dass die unpolare Extraktionsmethode in Bezug auf die Klassifizierungsgenauigkeit besser abschnitt als die polare Extraktionsmethode. Dieses Ergebnis stimmt ebenfalls mit der bereits diskutierten Anzahl der detektierten *features* überein, sodass die unpolare Extraktion für das weitere Vorgehen ausgewählt wurde. Ebenfalls wurde die positive Ionisation ausgewählt, da diese in der Klassifikation der Proben am besten abschnitt und die Kombination aus positiver Ionisation und unpolarer Extraktion sich bereits in der Herkunftsanalytik von weiteren Lebensmitteln mittels LC-MS als geeignet erwiesen hat (Creydt et al. 2018b; Schütz et al. 2021; Klockmann et al. 2016).

In einem nächsten Schritt wurde das Probenaufgabesystem der DART-Ionenquelle optimiert, was mittels der Quick-Strip- oder der Dip-it-Methode durchgeführt werden kann. Für diese Experimente wurde ein Datensatz aus 200 Erdbeerproben verwendet, mit dem Ziel, deutsche und nicht-deutsche Erdbeeren voneinander zu unterscheiden. Mittels des Quick-Strip-Ansatzes ließen sich 605 *features* ionisieren und diese führten zu einer Klassifizierungsgenauigkeit von 86,5%, während das Dip-it-Probenaufgabesystem 597 *features* erfasste und eine Klassifizierungsgenauigkeit von 87,0% ergab. Da die Anzahl der *features* nahezu identisch war und die Klassifizierungsgenauigkeit nur um 0,5% abwich, wurde darauf geschlossen, dass es keinen deutlichen Unterschied bei der Klassifizierung mit den verschiedenen Probenaufgabensystemen gab. Bei Verwendung der Quick-Strip-Karte müssen die Extrakte pipettiert werden, während bei der Dip-it-Methode nur die Glasstäbe in den Probenextrakt getaucht werden müssen. Da bei der Dip-it-Methode der Schritt des Pipettierens entfällt, ist die Dip-it-Methode der schnellere und einfachere durchzuführende Ansatz und wurde deshalb für das weitere Verfahren ausgewählt. Darüber hinaus zeigt dieses Ergebnis der Methodenoptimierung, dass eine Unterscheidung zwischen deutschen und nicht-deutschen Proben mit der entwickelten Methode mit vielversprechender Genauigkeit von 87,0% möglich ist.

7.1.2 Authentifizierung von Erdbeeren

Die in Abschnitt 7.1.1 entwickelte Methode wurde nun auf einen größeren marktrealen Datensatz, bestehend aus 207 Proben, angewendet. Die Proben stammten aus den wichtigsten Erzeugerländern für den europäischen Markt: Spanien, Griechenland, Ägypten, Polen, die Niederlande sowie Deutschland.

Zunächst wurde eine unüberwachte PCA durchgeführt, um Trends und potenzielle Ausreißer innerhalb des Datensatzes zu identifizieren. Die *Score Plots* der ersten drei Hauptkomponenten sind in Abbildung 04 dargestellt und zeigen, dass erste Ansätze zur Separierung der geographischen Herkunft der Proben im Datensatz vorhanden sind.

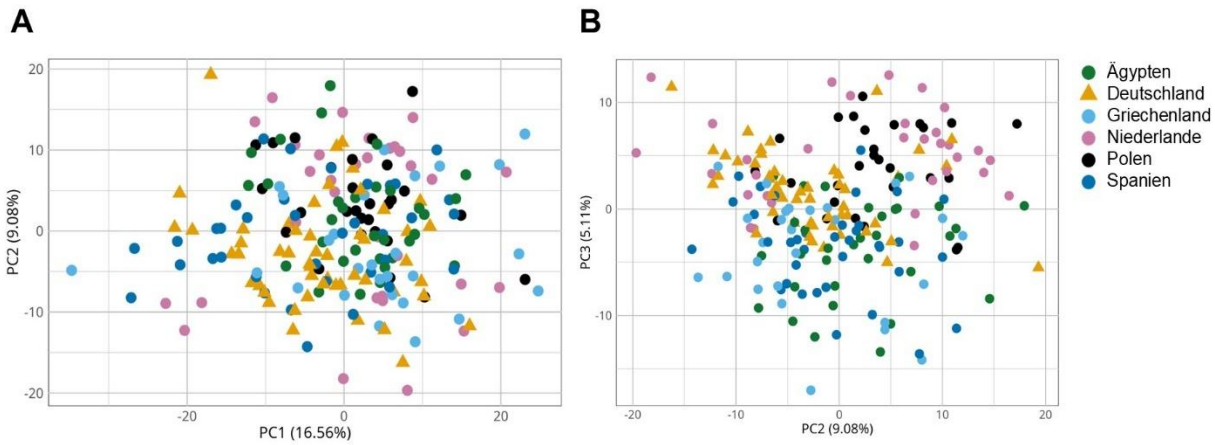


Abbildung 04: Ergebnisse der Hauptkomponentenanalyse: *Score Plots* der ersten und zweiten (A) sowie der zweiten und dritten Hauptkomponente (B) für die DART-MS-Daten der Erdbeeren. Die Proben wurden entsprechend ihrem Herkunftsland unterschiedlich eingefärbt.

Bei Betrachtung des *Score Plots* der ersten und zweiten Hauptkomponente (A) zeigt sich keine Auftrennung der Proben. Eine Clusterung nach der Herkunft ist erst im *Scores Plot* der zweiten und dritten Hauptkomponente (B) sichtbar, jedoch werden auch große Überlappungsbereiche der verschiedenen Herkünfte deutlich, was keine klare Trennung aller Herkünfte mittels der PCA zulässt. Nur einzelne Herkünfte wie beispielsweise die niederländischen und die griechischen Proben lassen sich mittels dieses ungerichteten Verfahrens deutlich voneinander unterscheiden.

Daraufhin wurde mit dem *Random Forest* ein gerichtetes Klassifikationsverfahren zur Unterscheidung der sechs Herkünfte angewendet. Es ergab sich eine Richtigkeit für die Klassifikation der Länder von 63,8%. Die Ergebnisse der einzelnen Länder sind Tabelle 01 genauer aufgeschlüsselt.

Tabelle 01: Konfusionsmatrix mit den prozentualen Richtigkeiten der Herkunftsbestimmung von 207 Erdbeerproben. Die Klassifikation wurde mittels des *Random Forest* durchgeführt. Es ist die Anzahl der vermessenen Proben (#) angegeben.

		Vorhergesagte Klasse [%]						
		Ägypten	Deutschland	Griechenland	Niederlande	Polen	Spanien	#
Wahre Klasse [%]	Ägypten	66,7	3,0	15,2	3,0	12,1	0,0	33
	Deutschland	2,0	78,4	3,9	5,9	5,9	3,9	51
	Griechenland	3,7	7,4	48,1	0,0	11,1	29,6	27
	Niederlande	3,1	9,4	12,5	59,4	9,4	6,3	32
	Polen	21,4	0,0	0,0	14,3	64,3	0,0	28
	Spanien	0,0	5,6	30,6	8,3	0,0	55,6	36

Die Richtigkeiten der einzelnen Länder lagen bei 48,1–78,4% an korrekt klassifizierten Proben, was die *a priori*-Wahrscheinlichkeit von 16,7% für alle Klassen deutlich übersteigt. Die beiden Klassen mit der geringsten Klassifizierungsgenauigkeit sind die griechischen Proben mit 48,1% und die spanischen Proben mit 55,6% korrekten Klassifizierungen. Es ist auffällig, dass der Großteil der Fehlzuordnung der Proben aus Griechenland und Spanien zur jeweils anderen Klasse geschieht. Zusätzlich lassen sich Fehlklassifikationen zwischen Polen, den Niederlanden und Deutschland erkennen. Diese Überschneidung lässt sich mit den ähnlichen klimatischen Bedingungen in den Anbaugebieten erklären. Sowohl Griechenland und Spanien als auch die Niederlande, Deutschland und Polen sind Länder mit ähnlichen Klimata und Regenmengen. Durch diese ähnlichen Wachstumsbedingungen in mediterranen oder zentraleuropäischen Klimata entstehen ähnliche Metabolitenprofile, was zu den beobachteten Fehlklassifizierungen führt. Insgesamt zeigen diese Richtigkeiten, dass eine Klassifizierung möglich, aber die Unterscheidung der einzelnen Länder mit ähnlichen Wachstumsbedingungen schwierig ist.

Um diese Unterschiede und Überschneidungen zwischen den Klassen zu untersuchen, wurden wichtige Marker mittels des Boruta-Verfahrens selektiert und mittels DART-MS/MS Experimenten identifiziert. Dies demonstrierte die Möglichkeit mittels DART-MS Identifizierungen von Markern durchzuführen und eine Betrachtung der Signalintensitäten der Marker zu ermöglichen. Somit konnte untersucht werden, ob die Ähnlichkeiten in der Klassifizierung auch auf der Ebene der Marker vorhanden waren. Als Ergebnis der Markerselektion wurden insbesondere Di- und Triacylglyceride als relevante Verbindungen identifiziert. Die Signalintensitäten ausgewählter Markerverbindungen sind in Abbildung 05 in Form von Boxplot-Diagrammen dargestellt.

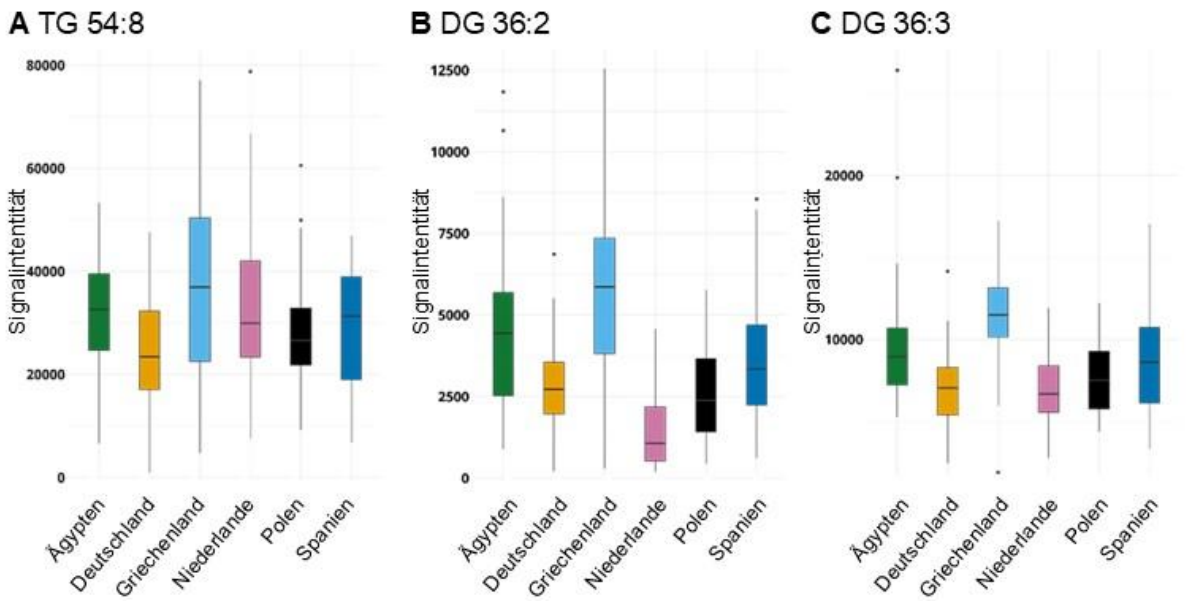


Abbildung 05: Boxplot der Signalintensitäten von drei mittels Boruta selektierten Markerverbindungen (A: Triacylglyceride 54:8; B: Diacylglycerid 36:2, C: Diacylglycerid 36:3) zur Herkunftsbestimmung von Erdbeeren aus den sechs Länder Ägypten, Deutschland, Griechenland, den Niederlanden, Polen und Spanien.

Bei der Betrachtung der Signalintensitäten wurden entsprechende Intensitätsunterschiede zwischen den Proben unterschiedlicher Herkünfte festgestellt, wenn auch für einige Markerverbindungen, wie dem Triacyglycerid 54:8, nur geringe Unterschiede zu beobachten waren. Größere Unterschiede zeigten die Signalintensitäten der Diacylglyceride 36:2 und 36:3, welche in den griechischen, spanischen und ägyptischen Proben höher waren als die der zentraleuropäischen Länder. Derartige Differenzen machten Unterschiede in klimatisch sehr unterschiedlichen Ländern deutlich, demonstrierten aber gleichzeitig auch Überschneidungen in klimatisch sehr ähnlichen Gebieten. Daraus wurde geschlossen, dass grundsätzlich eine Herkunftsunterscheidung der verschiedenen Erdbeerproben möglich ist und geeignete Markerverbindungen identifiziert werden können, es jedoch Einschränkungen gibt, wenn die Erdbeeren aus benachbarten Regionen stammen. Eine ausführliche Analyse der untersuchten Länder zeigte, dass insbesondere zwischen den geographisch verwandten Proben aus der Norddeutschen Tiefebene sowie Proben aus Griechenland und Spanien, erhebliche Ähnlichkeiten zu beobachten sind.

Eine weitere Erklärung für die erschwerte Herkunftsunterscheidung liegt in der Komplexität des Erdbeeranbaus. So ist der Erdbeeranbau wie in Abschnitt 4.1.1 bereits erläutert eine Sonderkultur, bei der unterschiedliche Pflanzsysteme wie Freilandanbau, Folientunnel oder Gewächshäuser zum Einsatz kommen (Oğuz et al. 2022; Akhatou et al. 2016). Insbesondere bei einem Anbau, der nicht im Boden, sondern in Substrat stattfindet, ist der Einfluss des Standortes auf das Metabolom reduziert. Des Weiteren werden Erdbeersetzlinge in Torf herangezogen und danach an ihren Endstandort verbracht (Capocasa et al.). Durch diese

Einflüsse reduziert sich eine markante Ausprägung des Metaboloms zusätzlich. Zusammenfassend war es möglich, die sechs Länder mit einer Richtigkeit von 63,8% und deutsche von importierten Erdbeeren mit 87,0% voneinander zu unterscheiden, auch wenn die Leistungsfähigkeit der Methode aufgrund der erwähnten Faktoren reduziert ist.

7.2 Herkunftsauthentifizierung Körnermais

Körnermais ist eine der weltweit wichtigsten Feldfrüchte und leistet einen bedeutenden Beitrag zur globalen Ernährungssicherheit (United States Department of Agriculture 2025). Als wichtiges Lebens- und Futtermittel ist Körnermais durch das große Handelsvolumen besonders anfällig für betrügerische Herkunftserklärungen. Der Ort des Anbaus beeinflusst nicht nur die ernährungsphysiologische Qualität des Futters, sondern auch den Risikofaktor für die Kontamination mit Mykotoxinen (Kirpich et al. 2018). Darüber hinaus können Ernteausfälle zu lokalen und jährlichen Preisschwankungen beitragen. All diese Faktoren erhöhten die Lukrativität, das Herkunftsland von Körnermais zu fälschen.

Sowohl Verbraucher als auch Landwirte, die Körnermais als Futtermittel beziehen, können von diesen Verfälschungen betroffen sein. Die USA und europäische Länder, wie Frankreich, Deutschland, Ungarn und die Ukraine, tragen zu einem großen Teil zur weltweiten Körnermaisernte bei und wurden daher für den Herkunftsvergleich von Körnermais ausgewählt. Peru wurde als Beispiel für ein südamerikanisches Land gewählt, da dieser Kontinent ebenfalls ein großer Körnermaisproduzent ist (Vásquez et al. 2024). Mittels eines Datensatzes von insgesamt 200 Körnermaisproben konnte die DART-MS-basierte Herkunftsanalyse an einem kohlenhydratreichen Lebensmittel getestet werden. Aus einer vorangegangenen Arbeit zur Herkunftsbestimmung mittels LC-MS konnten die optimierte Extraktion nach Bligh und Dyer und die positive Ionisation übernommen werden (Schütz et al. 2021; Bligh und Dyer 1959). Die Vermessung der Probenextrakte fand mit der in Abschnitt 4.5.2.2.2 erläuterten DART-qToF-MS Kopplung statt und es wurde das Dip-it-Verfahren zur Probenaufgabe verwendet.

Die Ergebnisse wurden zunächst mittels einer ungerichteten Hauptkomponentenanalyse untersucht. Die erhaltenen Ergebnisse sind graphisch in Form von *Score Plots* in Abbildung 06 dargestellt.

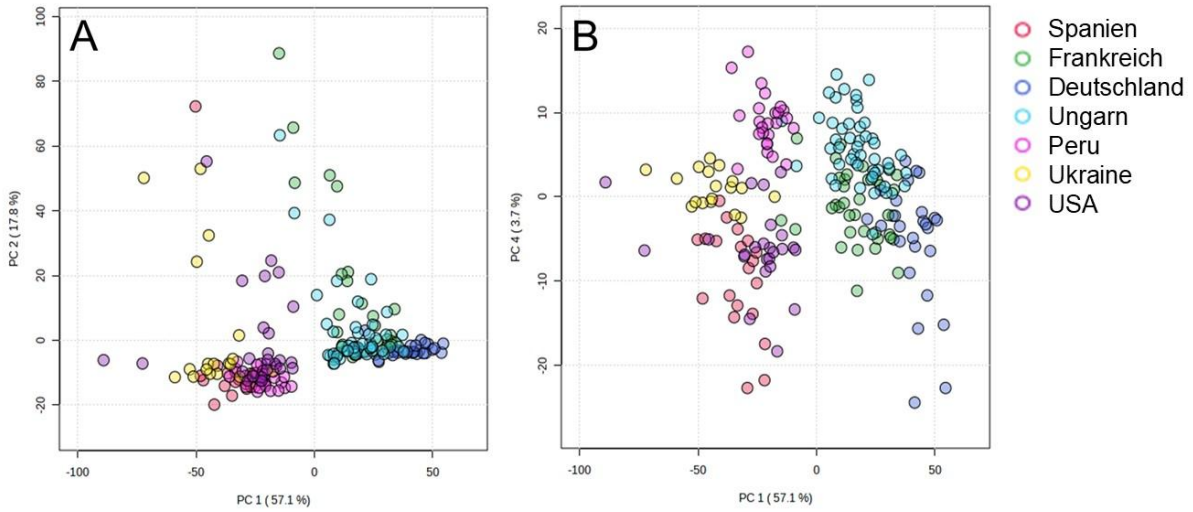


Abbildung 06: Ergebnisse der Hauptkomponentenanalyse am Beispiel von Körnermais: *Score Plots* der ersten und zweiten (A) sowie der ersten und vierten Hauptkomponente (B) für die DART-MS-Daten von Körnermais nach Summen-Normalisierung und Mittelwertzentrierung. Die vermessenen Stichproben wurden entsprechend ihrer geographischen Herkunft unterschiedlich eingefärbt.

Die Hauptkomponenten eins bis vier erklären 57,1; 17,8; 11,6 bzw. 3,7% der Varianz der Daten. Der *Score Plot* der ersten beiden Hauptkomponenten zeigt eine klare Trennung in zwei Cluster. Innerhalb des einen Clusters ordnen sich die Proben aus Frankreich, Ungarn und Deutschland an, in dem anderen Cluster befinden sich die Proben aus Spanien, den USA, der Ukraine und Peru. Innerhalb dieser Cluster sind geringe Tendenzen zur räumlichen Separation der Proben zu erkennen. Bei Betrachtung der Hauptkomponenten eins und vier zeigt sich eine Clusterbildung der Proben anhand ihrer Herkunft. Diese räumliche Abgrenzung der Proben untereinander belegt, dass bereits mittels einer ungerichteten Analyse Trends zur Herkunftsunterscheidung erkennbar sind.

Die anschließend auf *Random Forest*-basierende Klassifizierung ergab eine Richtigkeit von 84,4%, was bei einem Modell mit sieben Klassen eine gute Richtigkeit darstellt. Die Ergebnisse sind in der Konfusionsmatrix in Tabelle 02 genauer aufgeschlüsselt.

Tabelle 02: Konfusionsmatrix mit den prozentualen Richtigkeiten der Herkunftsbestimmung von 200 Körnermaisproben. Die Klassifikation wurde mittels *Random Forest* durchgeführt und es ist die Anzahl der vermessenen Proben (#) angegeben.

		Vorhergesagte Klasse [%]							
		Spanien	Frankreich	Deutschland	Peru	Ukraine	Ungarn	USA	#
Wahre Klasse [%]	Spanien	82,1	0,0	0,0	4,1	0,0	0,0	13,9	20
	Frankreich	0,0	82,8	9,6	0,0	0,0	6,4	1,2	37
	Deutschland	0,0	2,9	92,7	0,0	0,0	4,4	0,0	26
	Peru	3,7	0,0	0,0	95,7	0,0	0,0	0,6	25
	Ukraine	0,0	3,4	10,5	2,7	73,7	0,0	9,7	18
	Ungarn	0,0	15,4	2,4	2,0	0,0	80,1	0,0	50
	USA	6,2	0,0	4,1	4,3	0,6	0,0	84,9	24

Die Proben aus Peru und Deutschland zeigen mit 95,7 bzw. 92,7% die höchsten Richtigkeiten. Bei den Proben aus der Ukraine ist die niedrigste Richtigkeit von 73,7% zu beobachten, was auf die geringe Anzahl von 18 Proben zurückzuführen ist. Bei der Verwendung von multivariaten Klassifikationsverfahren ist es möglich, dass kleine Klassen im Training der Modelle, aufgrund ihrer geringeren Probenanzahl weniger gewichtet werden. Dieser Effekt kann mittels einer Gewichtung der Proben beim Training zum Teil ausgeglichen werden und wurde deshalb auf den vorliegenden Datensatz angewendet (Agusta 2019). Jedoch zeigte sich weiterhin eine geringere Richtigkeit bei den ukrainischen Proben, was darauf schließen lässt, dass eine gewichtete Klassifizierung dieses Problem nicht gänzlich lösen kann, aber eine Klassifizierung der Proben dennoch mit ausreichender Richtigkeit möglich ist. Die ungarischen und französischen Proben wurden mit einer Genauigkeit von 80,1% und 82,8% richtig klassifiziert. Dabei sind die Fehlzuordnungen zwischen Frankreich und Ungarn von 15,4 und 6,4% auffällig, was auf die ähnlichen kontinentalen Klimata der beiden Länder zurückzuführen ist. Beide Länder haben ein kontinentales und warmes Klima mit sehr ähnlichen Jahreswerten für Temperatur und Niederschlag, was die Unterscheidung zwischen diesen beiden Klassen wahrscheinlich erschwert hat (Izsák et al. 2022; Emery et al. 2021; Schütz et al. 2021). Darüber hinaus gab es einige Fehlklassifizierungen zwischen Proben aus Spanien und den USA. Der Grund dafür könnte sein, dass die USA ein recht großes Land sind und die Proben aus verschiedenen südlichen Bundesstaaten gezogen wurden, um einen möglichst großen Teil der USA abzudecken. Dieser Ansatz bietet den Vorteil einer großen Abdeckung des jeweiligen Landes, schafft aber auch eine diffusere Klasse mit größeren Überschneidungen zu anderen Klassen. Diese Ähnlichkeiten zwischen den spanischen und US-amerikanischen Proben wurden bereits in der Hauptkomponentenanalyse in Abbildung 06 ersichtlich.

Zur Identifizierung relevanter Marker wurde eine Rangfolge der Variablenwichtigkeit mittels *Random Forest* erstellt. Da über DART-MS/MS-Messungen nur wenig Literatur zum Fragmentierungsverhalten von unterschiedlichen Metaboliten vorlag, wurden Standards von verschiedenen, im unpolaren Metabolom zu erwartenden Molekülen vermessen. Durch die Aufklärung der Fragmentierungsmuster konnte gezeigt werden, dass Mono-, Di- und Triacylglyceride an den Esterbindungen fragmentiert werden, was eine Analogie zu LC-ESI-MS/MS Messungen darstellt. Zusätzlich wurde bewiesen, dass diese Verbindungen als $[M+NH_4]^+$-Addukte ionisiert werden, obwohl kein Ammonium dem Extraktionsmittel zugegeben wurde. Dieser Effekt stellt ein Spezifikum der Ionisation in der Laboratmosphäre dar, da die in der Laborluft vorhandenen Ammoniakmoleküle für eine Bildung dieser Addukte ausreichen (Gross 2019). Des Weiteren konnte gezeigt werden, dass Phospholipide mittels DART-Ionenquellen nur schwer ionisiert werden können und diese bei der Ionisierung, unter Abspaltung der Phosphatgruppe *in-source* fragmentiert werden.

Mithilfe der abgeleiteten Fragmentierungsmuster wurden sieben Metabolite aus den Lipidgruppen der Di- und Triacylglyceride unter den wichtigen *features*, die für eine Herkunftsbestimmung geeignet sind, mithilfe von MS/MS-Spektren identifiziert. Von diesen Markern sind von drei beispielhaften Vertretern in Abbildung 07 die Signalintensitäten der einzelnen Länder als Boxplot-Diagramme dargestellt.

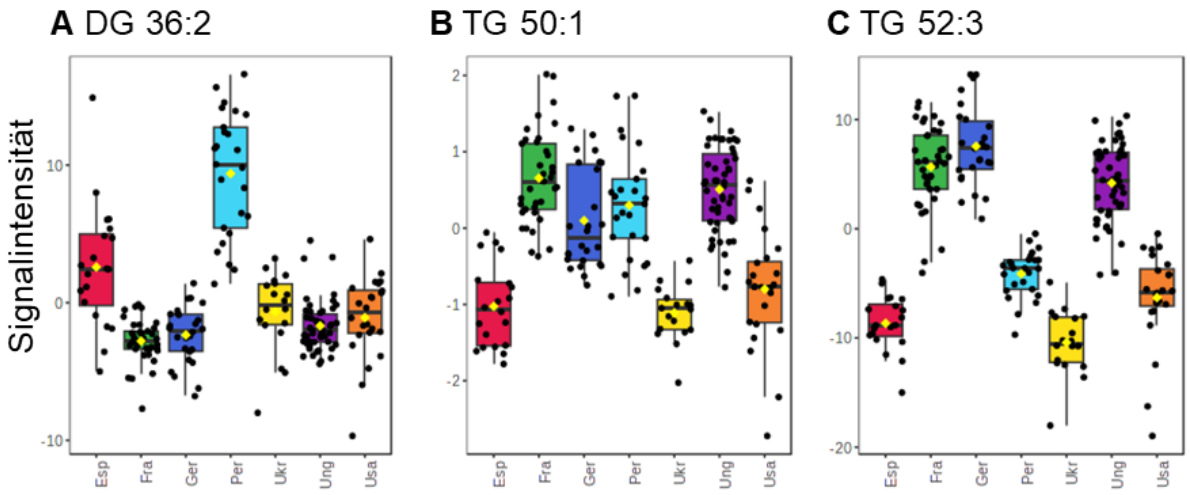


Abbildung 07: Boxplots der Signalintensitäten von drei identifizierten Markerverbindungen (A: Diacylglycerid 36:2; B: Triacylglycerid 50:1, C: Triacylglycerid 52:3) zur Herkunftsbestimmung von Körnermais aus den sieben Länder Spanien, Frankreich, Deutschland, Peru, Ukraine, Ungarn und den USA.

Die Diagramme der identifizierten Marker zeigen deutlich unterschiedliche Signalintensitäten zwischen den Ländern. Insbesondere Diacylglyceride 36:2 weisen besonders hohe Signalintensitäten in den Proben aus Peru auf und können somit eine Unterscheidung zu den anderen Ländern ermöglichen, was das hohe Klassifikationsergebnis von 95,7% erklärte. Des Weiteren liegen in den deutschen und französischen Proben ähnliche Signalintensitäten für die Triacylglyceride 50:1 und 52:3 vor. Die französischen Proben stammen aus dem Departement Elsass, das direkt an Deutschland grenzt, worin eine ähnliche Signalintensität dieser Marker aufgrund der ähnlichen Herkünfte begründet liegt Die Diacylglyceride sind eine Stoffklasse, welche häufig mit verschiedenen Arten von Pflanzenstress wie bspw. Hitze, Trockenheit und Sonneneinstrahlung in Verbindung gebracht wird, was die unterschiedlichen Konzentrationen aufgrund von verschiedenen Einflussfaktoren wie Klima und Niederschlag an den unterschiedlichen Standorten erklären kann (Kue Foka et al. 2020; Fahad et al. 2017). Ebenfalls von unterschiedlichen Pflanzenstressfaktoren beeinflussbar ist der Sättigungsgrad von Triacylglyceriden, weswegen eine Unterscheidung der Herkunft in Zusammenhang mit Pflanzenstress und daraus resultierenden eher gesättigten oder ungesättigten Verbindungen vorgenommen werden kann (Upchurch 2008). Somit konnten mittels DART-MS/MS verschiedene Marker identifiziert werden, welche auf unterschiedliche Einflüsse auf das Pflanzenmetabolom zurückzuführen sind. Insgesamt konnte gezeigt werden, dass ein *non-targeted* Ansatz mittels DART-MS zur Bestimmung der Herkunft von Körnermais

herangezogen werden kann und es möglich war mittels DART-MS/MS Experimenten aussagekräftige Marker zu identifizieren.

7.3 Herkunftsauthentifizierung von Haselnüssen

Die europäische Haselnuss ist ein wichtiges Handelsgut und Rohware für die Lebensmittelindustrie und wird dementsprechend in einer Vielzahl von Ländern angebaut und weltweit gehandelt (Food and Agriculture Organization of the United Nations 2024; Sammarco et al. 2023b).

Der Anbau von Haselnüssen erfolgt in Ländern mit mediterranem Klima wie Frankreich, Italien, die Türkei oder Georgien. Zudem wird derzeit auch der Anbau in Deutschland aufgrund wärmerer Klimata ausgeweitet (Mehlenbacher und Molnar 2021). Insbesondere italienische Haselnüsse, die mit dem EU-Siegel „geschützte geographische Angabe“ beworben werden dürfen, und des daraus resultierenden höheren Preises, weisen ein hohes Potential für Verfälschungen auf. Daher bildet die Haselnuss ein ideales Beispiel, um die Herkunftsbestimmung eines fettreichen Lebensmittels, mittels eines DART-MS-basierten *Metabolomics*-Ansatzes, zu überprüfen. Hierfür wurde ein Datensatz von 172 Haselnussproben aus den fünf Ländern Frankreich, Georgien, Deutschland, Italien und der Türkei verwendet. Aus einer vorangegangenen Untersuchung von Haselnüssen mittels LC-MS wurde die Eignung einer unpolaren Extraktion in Kombination mit einer positiven Ionisation für die Messungen abgeleitet. Um die in dieser Arbeit vorgestellten Datensätze vergleichen zu können, wurde für die Haselnussproben ebenfalls eine modifizierte Extraktion nach Bligh and Dyer verwendet. Die erhaltenen Haselnussextrakte wurden dann durch die in Abschnitt 4.5.2.2.2 beschriebene DART-qToF-Kopplung in einem *non-targeted* Ansatz mittels des Dip-it-Probenaufgabensystems vermessen.

Die Ergebnisse der ungerichteten Hauptkomponentenanalyse, sind in Form von *Score Plots* der ersten, zweiten und vierten Hauptkomponente in Abbildung 08 dargestellt.

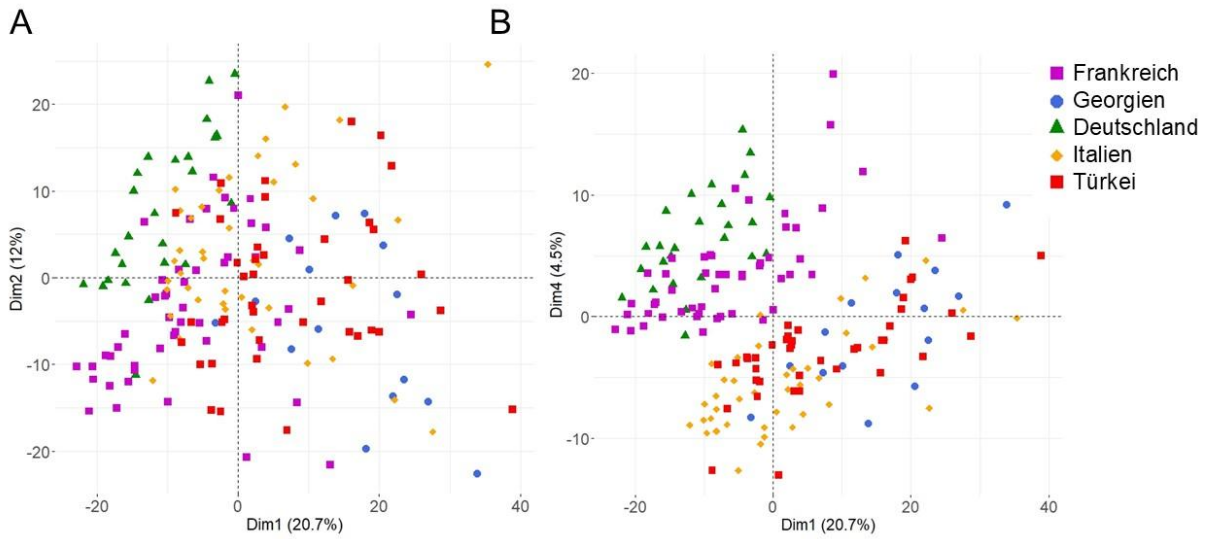


Abbildung 08: Ergebnisse der Hauptkomponentenanalyse zur Herkunftsbestimmung der Haselnussproben: *Score Plots* der ersten und zweiten (A) sowie der ersten und vierten Hauptkomponente (B) für die DART-MS-Daten von Haselnussproben nach Median-Normalisierung und Mittelwertzentrierung. Die Stichproben wurden entsprechend ihrer geographischen Herkünfte unterschiedlich farblich markiert.

Der *Score Plot* der ersten beiden Hauptkomponenten (Abbildung 08A) zeigt eine große räumliche Trennung der Proben. Bei der Einfärbung der Proben nach ihrer Herkunft wird direkt deutlich, dass sich bspw. die deutschen Proben räumlich von den Proben anderer Herkünfte abgrenzen. Ebenfalls weisen die Proben aus den anderen Ländern eine Clusterung auf. Deutlicher wird dieser Trend auch bei Betrachtung der ersten und vierten Hauptkomponenten im *Scores Plot* (Abbildung 08B), in dem eine Trennung in zwei verschiedene Cluster ersichtlich wird. Dabei liegen einerseits die Proben aus Frankreich und Deutschland beieinander sowie andererseits die georgischen, italienischen und türkischen Proben. Aufgrund der Anordnung der Proben besteht der eine Cluster folglich aus den mitteleuropäischen Ländern und der anderen Cluster aus den Proben der südeuropäischen/mediterranen Ländern. Innerhalb dieser beiden Cluster lag zudem teilweise eine Trennung der Proben nach ihrer Herkunft, mit Überschneidungen, insbesondere zwischen den georgischen, italienischen und türkischen Proben, vor. So zeigte sich bereits in der ungerichteten Analyse eine Clusterung der Proben entsprechend ihren geographischen Herkünften.

Im Anschluss wurde eine gerichtete Klassifizierung mittels *Random Forest* durchgeführt. Diese ergab eine sehr gute Richtigkeit von 93,2%. Die erhaltene Konfusionsmatrix der Tabelle 03 zeigt die nach Ländern aufgeschlüsselten prozentualen Richtigkeiten.

Tabelle 03: Konfusionsmatrix mit den prozentualen Richtigkeiten der Herkunftsbestimmung von 172 Haselnussproben aus fünf Ländern. Die Klassifikation wurde mittels *Random Forest* durchgeführt und es ist die Anzahl der vermessenen Proben (#) angegeben.

		Vorhergesagte Klasse [%]					
		Frankreich	Georgien	Deutschland	Italien	Türkei	#
Wahre Klasse [%]	Frankreich	92,7	1,7	3,8	1,8	0,0	50
	Georgien	2,7	89,0	0,0	8,4	0,0	15
	Deutschland	0,2	0,0	99,8	0,0	0,0	26
	Italien	2,4	3,6	0,0	91,8	2,3	41
	Türkei	0,0	6,9	0,0	0,4	92,6	40

Die Richtigkeiten der einzelnen Länder liegen zwischen 89,0% und 99,8% und somit alle in einem sehr guten Bereich. Dennoch fallen einige Fehlklassifikationen auf, wie beispielsweise die Zuordnung von 3,8% der französischen Proben zur deutschen Klasse. Wie schon in der Hauptkomponentenanalyse festgestellt wurde, weisen die deutschen und französischen Proben Ähnlichkeiten auf, welche die geographische Nähe der beiden Länder widerspiegeln. Dieser Effekt zeigt sich auch in der Fehlklassifizierung von 6,9% der türkischen Proben zur georgischen Klasse, da die beiden Herkünfte direkte Nachbarländer mit Haselnussanbau an der Schwarzmeerküste sind. Des Weiteren wurden die georgischen Proben mit 89,0% richtig zugeordnet, aber 8,4% der georgischen Proben wurden als italienisch klassifiziert. Georgien stellte mit 15 Proben die kleinste Klasse dar, während Italien mit 41 Proben die zweitgrößte Klasse darstellte, was zu einer größeren Gewichtung der italienischen Proben in der multivariaten Klassifizierung und damit zu Fehlklassifizierungen von kleinen Klassen zu den eher größeren Klassen führen kann. Insgesamt konnten jedoch alle untersuchten Herkünfte mit sehr guter Richtigkeit klassifiziert werden, auch wenn sich geographische Ähnlichkeiten in den fehlklassifizierten Proben erkennen lassen.

Zur Identifizierung wichtiger Variablen wurden eine Rangfolge der Variablenwichtigkeit mittels *Random Forest* aufgestellt und mittels Datenbanksuche die *features,* den möglichen Metaboliten zugeordnet. Für die Metabolitklassen der Tocopherole, Fettsäuren, Di- und Triacylglyceride wurden die Fragmentierungsmuster mittels DART-MS/MS-Messungen anhand von Standards aufgeklärt. Zusätzlich konnte gezeigt werden, dass die Metabolite während der Ionisierung bis zu drei unterschiedlichen Addukten als $[M+H]^+$, $[M+H–H_2O]^+$ und $[M+NH_4]^+$-Addukte bildeten. Die Fragmentierungsmuster wurden dann auf die MS/MS-Messungen der wichtigsten *features* angewendet und es konnten 16 unterschiedliche Marker identifiziert werden. Eine Auswahl an Signalintensitäten von verschiedenen Markern findet sich in Abbildung 09.

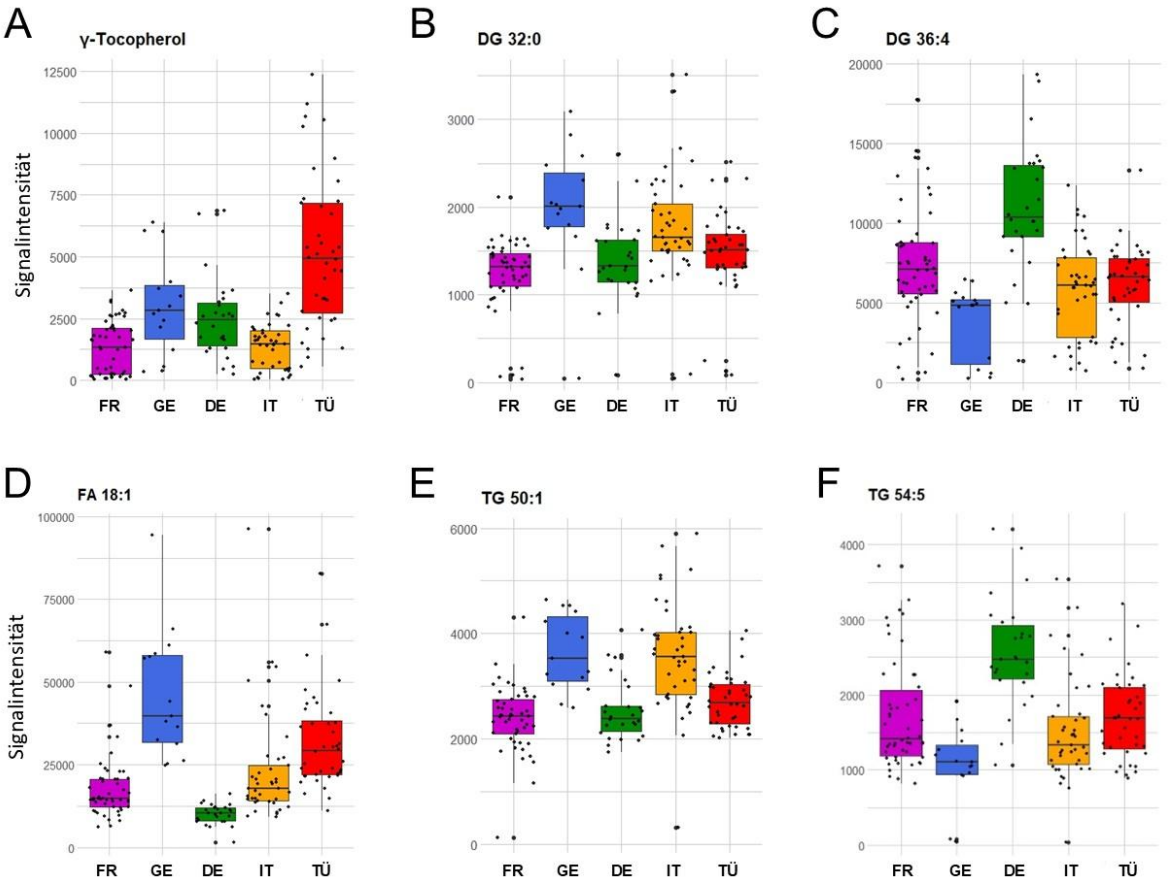


Abbildung 09: Boxplots der Signalintensitäten von sechs identifizierten Markerverbindungen (A: γ-Tocopherol; B: Diacylglycerid 32:0, C: Diacylglycerid 36:4; D Fettsäure 18:1; E: Triacylglycerid 50:1; F Triacylglycerid 54:5) zur Herkunftsbestimmung von Haselnüssen der fünf Länder Frankreich (FR), Georgien (GE), Deutschland (DE), Italien (IT) und der Türkei (TÜ).

Bei der Betrachtung der Signalintensitäten für γ-Tocopherol zeigten sich deutlich höhere Signalintensitäten in türkischen und teilweise auch in den georgischen Proben. Die hohen Signalintensitäten von γ-Tocopherol lassen vermuten, dass die höheren Temperaturen in den Ländern mit einer höheren Bildung von Antioxidantien assoziiert werden können, welche die Pflanze vor Lipidoxidation schützen (Górnaś et al. 2022). Die Signalintensitäten der Di- und Triacylglyceride wiesen einen weiteren Trend auf: Gesättigte oder einfach ungesättigte Verbindungen, wie beispielsweise Diacylglycerid 32:0 und Triacylglycerid 50:1, zeigten hohe Signalintensitäten in den Proben aus Georgien, Italien und teilweise der Türkei. Im Gegensatz dazu zeigten mehrfach ungesättigte Verbindungen, wie Diacylglycerid 36:4 und Triacylglycerid 54:5 hohe Intensitäten in deutschen Proben und teilweise auch in französischen Proben. Diese Intensitätsunterschiede können genutzt werden, um die zentraleuropäischen von den mediterranen Proben abzugrenzen und beruhen möglicherweise auf Anpassungen der Haselnusspflanzen an die heißeren Klimata im Mittel- und Schwarzmeerraum. Die höheren Signalintensitäten an gesättigten Fettsäuren in wärmeren Ländern können mit der Reaktion der Pflanzen auf Hitze- und Trockenstress zusammenhängen. So trugen beispielsweise

mehrfach ungesättigte Fettsäuren zur Toleranz bei niedrigeren Temperaturen bei, während niedrige Gehalte an ungesättigten Fettsäuren die Widerstandsfähigkeit gegenüber hohen Temperaturen erhöhten (Upchurch 2008). Diese Unterschiede konnten bereits in weiteren Arbeiten zur Authentifizierung von Pistazien und Mandeln festgestellt werden (Boukid et al. 2019; Lösel et al. 2024). Des Weiteren lässt sich dieser Trend auch bei der Ölsäure beobachten (Abbildung 09D), die als einfach ungesättigte Fettsäure hohe Intensitäten in den Proben aus mediterranen Ländern aufweist. So konnten die mittels DART-MS/MS Messungen identifizierten Marker der Tocopherole, freien Fettsäuren, Di- und Triacylglyceride, in Verbindung mit den biologischen Prozessen und den Klimata der Anbauregionen gebracht werden. Die sehr guten Ergebnisse der Klassifizierung sowie die aufgedeckten Vernetzungen zwischen Anbaugebiet und den Markern belegten die Möglichkeit der Herkunftsbestimmung von Haselnüssen mittels DART-MS-basierter Metabolomanalyse.

7.4 DART-MS zur Herkunftsunterscheidung

Die zu Beginn entwickelte Fragestellung zur Eignung von DART-MS-basierter Analytik für die Herkunftsauthentifizierung wurde durch die Entwicklung eines Arbeitsablaufes und Erprobung mittels unterschiedlicher Lebensmittel innerhalb dieser Arbeit untersucht. Die hierfür gewählten Beispiele Erdbeeren, Körnermais und Haselnüsse stellen jeweils Vertreter eines wasserreichen, kohlenhydratreichen und fettreichen Lebensmittels dar.

Die unpolare Extraktion der Proben und die Vermessung mittels eines DART-qToF-MS-basierten *non-targeted* Ansatzes führte in den drei untersuchten Datensätzen (Erdbeeren, Körnermais, Haselnüsse) zu einer Anzahl von 610, 647 und 799 *features*. Diese Zahlen der *features* zeigen, dass es möglich ist, zeitgleich mit einem *non-targeted* Ansatz eine hohe dreistellige Anzahl von *features* ohne chromatographische Verfahren zu erfassen. Die hier gewählte Kopplung der DART-Ionenquelle mit einem qToF-MS eignete sich für die Schnellanalytik, da qToF-MS Geräte hohe Scanraten ermöglichen und auf diese Weise eine ausreichend zeitliche Auflösung gewährleisten (Roemmelt et al. 2014). Darüber hinaus muss beachtet werden, dass die Datenverarbeitung der Rohdaten einen großen Einfluss auf das resultierende Ergebnis hat (Karaman 2017). Der hier entwickelte *Open-Source* Workflow gewährleistet durch die Verwendung gleicher Programme und Abläufe bei der Datenanalyse, dass die Datenvorbehandlung nur geringen Einfluss auf die ermittelten Unterschiede hat. Somit können durch die Datenprozessierung hervorgerufene Unterschiede minimiert und eine gute Vergleichbarkeit der Datensätze gewährleistet werden. Des Weiteren trägt die Verwendung ähnlich großer Probenanzahlen der drei Datensätzen von 207, 200 und 172 Proben zu einer höheren Vergleichbarkeit der Ergebnisse bei.

Für eine erste Betrachtung der Messergebnisse stellt die Hauptkomponentenanalyse ein geeignetes Mittel dar. Dieses unüberwachte Verfahren bietet die Möglichkeit multivariate

Datensätze zweidimensional darzustellen und so erste Trends zur Erkennung von herkunftsspezifischen Unterschieden zu ermöglichen. Eine Clusterung nach der Herkunft ist für die Erdbeerproben in den *Score Plots* (Abbildung 04) nur in geringem Umfang vorhanden und es zeigen sich bei Betrachtung, insbesondere der zweiten und dritten Hauptkomponenten, Cluster der verschiedenen Länder, jedoch überlappen diese sehr stark und ergeben keine klare Trennung. Dagegen ergab die Hauptkomponentenanalyse der Körnermais- und Haselnussproben (Abbildung 06 und Abbildung 08) bereits zwischen der ersten und zweiten Hauptkomponente, deutlich durch die Herkunft der Proben definierte, sichtbare Cluster. Weitere Betrachtungen insbesondere der vierten Hauptkomponente zeigten für beide Datensätze eine Auftrennung der Proben nach ihrer Herkunft mit geringen Überlappungen zwischen den Probenbereichen aus unterschiedlichen Ländern. Somit eignet sich bereits ein ungerichtetes Verfahren, um Herkunftsunterschiede in fettreichen und kohlenhydratreichen Lebensmitteln zu erkennen, während in der Erdbeere als wasserreiche Frucht nur ein marginaler Trend zu beobachten war. Damit zeigt sich, dass ungerichtete Verfahren zur Identifizierung von Trends in den mittels DART-MS aufgenommenen Datensätzen eingesetzt werden können. Jedoch kann in keinem der drei Datensätze eine vollständige Trennung der Proben nach der Herkunft erreicht werden, weswegen der Einsatz von gerichteten Verfahren für komplexe Datensätze nötig ist.

Die gerichtete Klassifizierung der Datensätze ergab für die Erdbeerproben eine Richtigkeit von 63,8%, für die Körnermaisproben 84,4% und für die Haselnussproben 93,2%. Die Werte zeigen, dass die Erdbeere am schwierigsten zu klassifizieren war, was die in Abschnitt 6.1 diskutierten Komplexitäten des Erdbeeranbaus widerspiegelt. Innerhalb jeder Klassifizierung zeigen sich Fehlklassifizierungen zwischen Ländern mit ähnlichen Klimata. Beispielsweise existieren bei der Herkunftsbestimmung von Erdbeeren Ähnlichkeiten zwischen griechischen und spanischen sowie zwischen deutschen und polnischen Proben. Dies war auch in geringerem Umfang bei den Körnermaisproben der Fall, indem sich zwischen Frankreich und Ungarn Überlappungen zeigten, welche auf ähnliche Klimata zurückgeführt werden konnten. Diese Limitation für unter sehr ähnlichen Bedingungen oder in räumlicher Nähe gewachsene Proben stellen eine generelle Herausforderung eines *Metabolomics*-basierten Ansatzes dar (Cubero-Leon et al. 2014).

Weiterhin ist zu beachten, dass die unterschiedliche Anzahl an Klassen die *a priori*-Wahrscheinlichkeit einer richtigen Klassifizierung beeinflusst und somit bei einem Vergleich der Gesamtrichtigkeiten berücksichtig werden muss. Die Studie zur Herkunftsbestimmung von Haselnüssen wies die geringste Anzahl an Klassen auf, da nur Proben aus fünf Ländern verfügbar waren. Im Gegensatz dazu konnten bei der Klassifizierung von Körnermais, Proben aus sieben Ländern berücksichtigt werden. Die fünf Klassen der Haselnussproben führen zu einer *a priori*-Wahrscheinlichkeit von 20,0% für eine korrekte Zuordnung, während die sieben

Herkunftsländer beim Körnermais eine geringere *a priori*-Wahrscheinlichkeit von 14,3% ergeben. Diese unterschiedlichen Ausgangswahrscheinlichkeiten erklären zumindest teilweise, warum die Klassifizierung der Körnermaisproben mit 84,4% eine geringere Richtigkeit aufwies als die der Haselnussproben mit 93,2%. Somit ergeben sich sowohl für das fettreiche als auch für das kohlenhydratreiche Lebensmittel eine gute Klassifikationsrate und damit ist die Möglichkeit der Herkunftsbestimmung gegeben.

Im Gegensatz dazu weist die Herkunftsbestimmung von Erdbeeren mit sechs Klassen eine reduzierte Richtigkeit von 63,8% auf. Allerdings müsste die Richtigkeit aufgrund der verschiedenen Klassenanzahlen zwischen der, der anderen beiden Datensätze und damit zwischen 84,4–93,2% liegen. Dies ist mit einer Richtigkeit von 63,8% aber nicht der Fall. Folglich muss es andere Gründe für die reduzierte Richtigkeit geben. Zum einen ist es möglich, dass eine Herkunftsklassifizierung von Erdbeeren, aufgrund der in Abschnitt 6.1 diskutierten Gründe, besonders schwierig mittels einer Metabolomanalyse durchzuführen ist. Hierauf deuten auch eine mittels LC-MS durchgeführte Metabolomanalyse zur Herkunft von Erdbeeren hin, in der dieselben Schwierigkeiten zu beobachten waren (Brockelt et al. 2025). Des Weiteren könnten wasserreiche Lebensmittel insgesamt schwieriger mittels einer DART-MS-basierten Metabolomanalyse zu klassifizieren sein. Dagegen sprechen die bereits erfolgreich mittels DART- oder LC-MS durchgeführten Bestimmungen der Herkünfte von wasserreichen Lebensmitteln wie Spargel, Engelwurz und Orangen (Creydt et al. 2018b; Kim et al. 2015; Díaz et al. 2014). So ist es wahrscheinlich, dass die Erdbeeren einen Sonderfall unter den Lebensmitteln darstellen und dass daher eine Klassifizierung nur mit reduzierter Richtigkeit möglich ist.

Da DART-Quellen bislang nur in vergleichsweise wenigen Publikationen zum Einsatz kamen, liegen derzeit auch nur begrenzte Kenntnisse vor, wie sich ihr Einsatz auf das Fragmentierungsmuster von verschiedenen Substanzklassen auswirkt. Daher wurden, wie in den Abschnitten 7.2 und 7.3 vorgestellt, zunächst die Fragmentierungsmuster von Referenzstandards für wichtige Lipidklassen aufgeklärt. Insgesamt konnte gezeigt werden, dass einzelne Fettsäuren sich direkt ionisieren lassen und Phospholipide wie Phosphatidylenositolamine und Phosphocholine bei der Ionisierung unter den gewählten Bedingungen *in source* fragmentiert werden. Mono-, Di- und Triacylglyceride zeigen unter Verwendung von ESI-Ionenquellen ähnliche Ionisations- und Fragmentierungseigenschaften im Vergleich zu der DART-Quelle, da diese Verbindungen typischerweise an der Esterbindung fragmentieren (Byrdwell und Neff 2002). Im Gegensatz zu ESI-Ionenquellen konnten keine Natriumaddukte identifiziert werden, da Natriumionen aus den Glasgeräten in das Laufmittel übergehen oder sich während der Chromatographie aus der Probe lösen (Ahern et al. 2019; Neto et al. 2021). Anders verhält es sich mit der Bildung von Ammoniumaddukten. Ihre Bildung lässt sich darauf zurückführen, dass in der Atmosphäre von Laboren üblicherweise Ammoniak

in geringen Konzentrationen vorliegt. Insbesondere Di- und Triacylglyceride neigen dazu mit dem atmosphärischen Ammoniak die entsprechenden Addukte zu bilden, die auch in allen drei Datensätzen nachgewiesen werden konnten (Vaclavik et al. 2009). Insgesamt können mittels DART-Ionisierung, abgesehen von den Phospholipiden, die gleichen Stoffklassen wie mittels ESI-Ionisierung erfasst werden. Zudem stellte sich heraus, dass die DART-Ionisierung tendenziell eine härtere Ionisierungsmethode ist als ESI-Ionisierung (Wang et al. 2018).

Anschließend wurden die aufgeklärten Ionisierungsmuster auf die DART-MS/MS-Spektren möglicher Markerverbindungen übertragen. Für den Datensatz der Erdbeeren wurden fünf Marker und 15 weitere Lipide identifiziert, in den Körnermaisproben konnten sieben und in Haselnussproben 16 Markverbindungen in ihrer Struktur aufgeklärt werden. Diese Ergebnisse zeigen, dass es möglich war in jedem Datensatz mittels DART-MS/MS-Messungen herkunftsrelevante Marker zu identifizieren und Strukturaufklärung mittels DART-MS zu betreiben. Die Betrachtung der Signalintensitäten der einzelnen Marker zeigte bei allen drei Datensätzen, dass es keinen perfekten Marker gab, der alle Herkünfte trennte, sondern es immer mehrerer kombinierter Marker bedarf, um die Herkunft zu bestimmen. Die Substanzklassen der identifizierten Marker aller Datensätze waren hauptsächlich Di- und Triacylglyceride. Seltener wurden Tocopherole, einzelne Fettsäuren oder Diacylglyceryl-O-4′-(N,N,N-trimethyl)-homoserine (DGTS) als Marker identifiziert. Diese eher unpolaren Markersubstanzen sind kongruent mit der unpolaren Extraktion nach Bligh and Dyer (Bligh und Dyer 1959).

Insgesamt zeigten die Haselnüsse als fettreiches Lebensmittel die größte Vielfalt an Stoffklassen der detektierten Marker und auch das beste Klassifikationsergebnis. Ebenfalls zeigten die Körnermaisproben, als kohlenhydratreiches Lebensmittel, eine gute Klassifikationsrate mit großen Unterschieden in den Signalintensitäten der identifizierten Marker. Aus diesen beiden Datensätzen kann geschlossen werden, dass die DART-MS-basierte Analytik zur Bestimmung von komplizierten Fragestellungen geeignet ist, solange Markerverbindungen mit deutlichen Unterschieden in der Signalintensität vorhanden sind. Ebenfalls müssen diese Marker in mittleren oder hohen Konzentrationen vorhanden sein, um nicht von Ionensuppressionen überdeckt zu werden und die Marker dürfen keiner *in source* Fragmentierung unterliegen. Eine mögliche Kombination dieser Effekte findet sich bei der Klassifikation der Erdbeeren als wasserreiches Lebensmittel. Aufgrund dessen eignet sich DART-MS-basierte Analytik gut für die Bearbeitung einfacherer Fragestellungen, wie die Differenzierung von ausländischen und deutschen Erdbeeren, aber es zeigen sich Limitationen bei der Unterscheidung von geographisch ähnlichen Regionen.

7.5 Vergleich der Authentifizierung mittels DART-MS und FT-NIR

Um die Herkunftsauthentifizierung von lichtspektroskopischen und MS-basierten Verfahren zu vergleichen, wurde die Herkunftsauthentifizierung von Erdbeeren, neben der DART-MS-basierten Analyse zusätzlich um eine FT-NIR-basierte Analytik ergänzt.

Hierfür wurden 198 Erdbeerproben aus Deutschland, Ägypten, Griechenland und Spanien homogenisiert, gefriergetrocknet und mittels FT-NIR Spektroskopie vermessen. Zum Einsatz kamen zwei NIR-Spektrometer: (i) Zum einen ein laborbasiertes *benchtop* Gerät mit einem Wellenlängenbereich von 11550–3950 cm^{-1} und (ii) zum anderes ein mobiles *handheld* Gerät, welches den Wellenlängenbereich von 8846–5938 cm^{-1} abdeckt. Die Messungen mittels des *benchtop* Gerätes führten aufgrund der Auflösung von 4 cm^{-1} zu 3720 Variablen und die Bauweise des *handheld* Gerätes resultierte in 16 Messpunkten.

Eine Klassifizierung der Proben mittels LDA zeigte für die mit dem *benchtop* Gerät aufgenommenen Daten eine Richtigkeit von 80,5%. Dagegen wurden die Daten des handheld Gerätes mit einer Richtigkeit von 71,8% klassifiziert. Die genauen Ergebnisse sind in Tabelle 04 aufgelistet.

Tabelle 04: Konfusionsmatrizes mit den prozentualen Richtigkeiten der Herkunftsbestimmung von 198 Erdbeerproben aus vier Ländern mit einem *benchtop* NIR-Gerät (oben) und einem mobilen *handheld* Gerät (unten). Die Klassifikation wurde mittels LDA durchgeführt und es ist die Anzahl der vermessenen Proben (#) angegeben.

			Vorhergesagte Klasse [%]				
			Ägypten	Deutschland	Griechenland	Spanien	#
Benchtop NIR-Gerät	Wahre Klasse [%]	Ägypten	79,2	16,4	0,0	4,4	22
		Deutschland	4,9	93,3	0,4	1,5	110
		Griechenland	0,5	7,9	42,4	49,2	21
		Spanien	1,9	4,8	24,2	69,1	45
Handheld NIR-Gerät	Wahre Klasse [%]	Ägypten	52,7	36,1	0,0	11,2	22
		Deutschland	10,3	82,7	0,5	6,5	110
		Griechenland	0,0	0,6	55,3	44,1	21
		Spanien	5,5	16,0	21,3	57,1	45

Die durch das *benchtop* Gerät vermessenen deutschen Proben wurden mit 93,3% und die Proben aus Ägypten mit 79,2% richtig klassifiziert. Diese Richtigkeiten zeigen, dass das Modell in der Lage ist, diese Herkünfte zu erkennen. Dagegen werden die Proben aus Griechenland und Spanien mit 42,4% und 69,1% weniger gut klassifiziert und es existieren mehrere falsche Zuordnungen zwischen diesen beiden Klassen. Die Ähnlichkeiten zwischen griechischen und spanischen Erdbeeren sowie die Limitierungen der Klassifikation durch den Erdbeeranbau wurden bereits in Abschnitt 6.1 ausführlich diskutiert.

Das *handheld* Gerät lieferte im Gegensatz zum *benchtop* Gerät reduzierte Richtigkeiten. So wurden nur 52,7% der ägyptischen, 55,3% der griechischen Proben und 57,1% der spanischen Proben richtig zugeordnet. Lediglich die deutschen Proben wurden, wahrscheinlich aufgrund der hohen Probenanzahl von 110 Proben mit 82,7% ausreichend richtig klassifiziert. Auf diese Weise verdeutlichen die erhaltenen Ergebnisse, dass mittels eines einfacheren mobilen NIR-Spektrometers eine Herkunftsklassifizierung möglich ist, aber nur mit geringen Richtigkeiten und dementsprechend einer großen Anzahl an Fehlklassifikationen. Aufgrund der reduzierten Richtigkeiten wurden für den Vergleich mit der DART-MS-basierten Herkunftsklassifizierung die Ergebnisse des *benchtop* Gerätes herangezogen. Zudem stellt die Herkunftsauthentifizierung mittels laborbasierter NIR-Analytik das etabliertere Verfahren dar (Wang und Yu 2015; Arndt et al. 2020a).

Die Vorbereitungen für die DART-MS- und die NIR-Messung unterscheiden sich nicht in Bezug zur Probenakquise, der Homogenisierung und der Gefriertrocknung. Messungen von authentischen Referenzproben sind für den Aufbau eines jeden Klassifizierungsmodells notwendig. Genauso ist eine Homogenisierung einer Probe aus mehreren Früchten angebracht, da sich hierdurch Abweichungen innerhalb einer Probe ausgleichen lassen. Mittels NIR-Spektroskopie ist eine Vermessung von ganzen Lebensmitteln denkbar, was den Vorteil einer schnelleren Messung ohne Probenvorbereitung hätte, jedoch führte dieses Vorgehen bisher zu reduzierten Richtigkeiten in der Klassifikation (Arndt et al. 2020b). Ebenfalls wäre für die DART-MS Analyse eine Direktionisation eines Lebensmittels denkbar, doch bisher wird dieses Vorgehen nur in der Holz-, Drogen und Sprengstoffanalytik mit halbquantitativen Ansätzen angewendet (An et al. 2019; Guo et al. 2020; Sisco und Forbes 2021).

Die darauffolgende Gefriertrocknung stellt mit einem Zeitaufwand von 1–4 Tagen den längsten Schritt des Authentifizierungsansatzes dar. Bei beiden Messverfahren gibt es Bestrebungen, diesen Schritt zu verkürzen oder zu eliminieren (Lösel et al. 2022). Jedoch ist die Wasserbande im NIR Spektrum bei nicht gefriergetrockneten Lebensmitteln meistens so groß, dass sie relevante Signale überlagert (Arndt et al. 2020b). Ebenfalls wäre es denkbar, Extrakte von nicht gefriergetrockneten Lebensmitteln für DART-MS-basierte Messungen anzufertigen. Da Erdbeeren einen Wassergehalt von circa 90,6% aufweisen ergibt sich eine Verdünnung der Metabolite um fast das Zehnfache. Somit besteht die Gefahr, dass durch die starke Verdünnung, weniger Metabolite erfasst werden und die Extrakte größeren Schwankungen unterliegen, weswegen sich auch für DART-MS-basierte Ansätze der Einsatz von gefriergetrocknetem Probenmaterial etabliert hat (Zacometti et al. 2024; Novotná et al. 2012; Massaro et al. 2021).

Die NIR Messung ist mit circa einer Minute pro Probe genauso schnell wie eine DART-MS Messung, jedoch wird bei NIR Messungen meist eine fünffache Bestimmung durchgeführt, um Streulichteffekte zu minimieren, was zu längeren Messzeiten von circa 5 Minuten pro Probe führt (Arndt et al. 2020b). Die DART-Messung könnte noch auf etwa 10 Sekunden pro Probe reduziert werden, denn die einminütige Messung besteht aus 50 Sekunden Aufheizzeit und 5–10 Sekunden Datenerfassung. In einer Routineanwendung ist die Aufheizzeit nur einmal vor den direkt aufeinanderfolgenden Messungen nötig, was die Messung beschleunigt und einen höheren Probendurchsatz von bis zu 100 Proben pro Stunde, im Gegensatz zu 12 Proben pro Stunde mittels NIR-basierten Authentifizierung, gewährleistet (Jastrzembski und Sacks 2016).

Die Richtigkeit der Klassifizierung lag bei dem NIR-basiertem Ansatz mit vier Ländern bei 80,5% und die des DART-MS-basierten Ansatzes mit sechs Ländern bei 63,8%. Die höhere Richtigkeit der NIR-basierten Klassifizierung kann durch die geringere Anzahl an untersuchten Klassen erklärt werden, da die *a priori* Wahrscheinlichkeit einer richtigen Zuordnung von 16,7% bei sechs Klassen auf 25,0% bei vier Klassen steigt. Bei beiden Klassifizierungen zeigen sich ähnliche Fehlklassifizierungen zwischen Griechenland und Spanien, welche bereits in den Abschnitten 7.1.2 diskutiert wurden. Das diese Überschneidung mittels zweier unabhängiger *Metabolomics*-basierter Ansätze besteht, untermauert die Annahme aus Abschnitt 7.1.2, dass die Ähnlichkeiten im Metabolitprofil der Proben vorhanden sind und nicht aus den DART-MS Messung resultieren.

Gut vergleichbar ist hingegen die Unterscheidung der deutschen und importierten Erdbeeren. da hierfür annähernd gleichgroße Datensätze verwendet wurden. Der mittels NIR untersuchte Datensatz weist 110 deutsche und 88 importierte Proben auf, während der mittels DART-MS untersuchte Datensatz 105 deutsche und 95 importierte Proben beinhaltet. Durch diese gleiche Verteilung der Proben in den Datensätzen können Effekte, die aus unterschiedlich verteilten Datensätzen kommen, ausgeschlossen werden. Die Klassifizierung des mittels NIR-Spektroskopie vermessenen Datensatzes ergab eine Richtigkeit von 91,9% und der DART-MS-basierte Ansatz führte zu 87,0%. Ebenso wie die Untersuchung der einzelnen Länder resultierten beide Ansätze in ähnlichen Richtigkeiten mit einer leicht besseren Richtigkeit des NIR-basierten Ansatzes.

Ein großer Vorteil der DART-MS Analytik im Vergleich zu anderen Schnellmethoden ist die Möglichkeit zur Markeridentifizierung. Mittels maschinellen Lernens und NIR-Spektroskopie können wichtige Banden des NIR-Spektrums zwar erkannt und diese mit funktionellen Gruppen relevanter Stoffklassen in Verbindung gebracht werden, jedoch ist es nicht möglich spezifische Marker beispielsweise der Diacylglyceride, einzeln zu betrachten und spezifisch zu identifizieren. Eine Identifizierung der Marker ist nur möglich, wenn weitere Methoden wie LC-MS- oder NMR-basierte Ansätze hinzugezogen werden, was allerdings mit einem erhöhten

Arbeitsaufwand einhergeht (Shakiba et al. 2022; Lösel et al. 2023). Im Gegensatz dazu ist es mittels DART-MS-basierter Analytik, durch die Kopplung mit einem qToF-MS, relativ einfach Markerverbindungen zu identifizieren. Dies wurde bereits in Abschnitt 7.1.2 diskutiert und anhand herkunftsspezifischer Marker von Erdbeeren erläutert. Auf diese Weise können einzelne Marker identifiziert und die Signalintensitäten untersucht werden, was die größere Selektivität der DART-MS Methode und die Möglichkeit zur biologischen Interpretation hervorhebt.

Fazit

Die Entwicklung von umweltfreundlicheren Methoden stellt eine neue Herausforderung für analytische Verfahren dar. Hierbei kann mittels der DART-MS-basierten Analytik circa 90% der erforderlichen Lösungsmittel im Vergleich zur LC-MS-Analyse eingespart werden, da nur noch geringe Volumina an Lösungsmitteln für die Extraktion benötigt werden. Noch besser schneidet die NIR-basierte Analytik ab, da hier gar keine Chemikalien benötigt werden und es sich somit um ein gänzlich umweltfreundliches Verfahren handelt (Escobedo et al. 2016). Zusammenfassend sind beide Ansätze sehr umweltfreundlich, wobei die NIR-Analytik durch den kompletten Verzicht auf Chemikalien noch vorteilhafter ist.

Insgesamt zeigt sich, dass es bei einem Vergleich zwischen NIR und DART-MS-basierter Analytik kein bevorzugtes Verfahren gibt, sondern die Wahl des Verfahrens vom Einsatzzweck abhängt. Für die Hochdurchsatzanalytik sind beide Verfahren geeignet und auch sind beide Verfahren umweltfreundlich. Die NIR-Spektroskopie zeigte bei der Bestimmung der Herkunft von Erdbeeren geringfügig höhere Richtigkeiten, wobei dieser Vergleich je nach Fragestellung unterschiedlich ausfallen kann (Losso et al. 2022). Sobald eine Identifizierung und detaillierte Betrachtung von Markern nötig sind, stellt der DART-MS-basierte Ansatz das überlegenere Verfahren dar. Hinsichtlich der Kosten der beiden Verfahren stellt die NIR-Spektroskopie das Verfahren mit den geringeren Anschaffungskosten dar, da NIR-Spektrometer kosteneffizienter in der Herstellung sind (Cipriano et al. 2007; Beć et al. 2021). In der Routineanalytik könnten sich DART-MS-basierte Verfahren, aufgrund des höheren Probendurchsatzes, als preisgünstigeres Verfahren etablieren (Schmauder et al. 2024).

7.6 Limitationen und Ausblick

In der vorliegenden Arbeit konnte gezeigt werden, dass sich der entwickelte Arbeitsablauf für die Vermessung von unterschiedlichen Lebensmitteln zur Herkunftsbestimmung mittels eines DART-MS-basierten *Metabolomics*-Ansatzes eignet. Hierbei wurde mit dreistelligen Probenzahlen gearbeitet und zudem eine Markeridentifizierung mittels MS/MS-Messungen durchgeführt, was bisher nur in einer begrenzten Anzahl von Studien der Fall war.

Aus den Ergebnissen lassen sich die in Abschnitt 7.4 beschriebenen Vorteile der schnellen, kostengünstigen und umweltfreundlichen Methode und die Anwendungsmöglichkeiten der DART-MS-basierten Analytik ableiten. Jedoch wurden im Rahmen dieser Dissertation auch Limitationen der DART-MS Analytik deutlich sichtbar. Diese Limitationen teilen sich zum einen in konstruktive Herausforderungen der DART-MS-Kopplung und zum anderen in anwendungsspezifische Hürden, welche bis zu einem Routineeinsatz zu überwinden sind.

Viele der konstruktiven Herausforderungen sind darauf zurückzuführen, dass während der Durchführung dieser Dissertation keine kommerziell erhältliche DART-MS-Kopplung mit einem hochauflösenden MS zu erwerben war, welche sowohl Ionenquelle als auch MS in einem Gerät vereinte. Daher gab es ebenfalls keine Softwarelösungen für eine einfache und schnelle Datenauswertung, welche direkt vom Endnutzer eingesetzt werden kann. Stattdessen musste zunächst der in Abschnitt 4.6.1 vorgestellte *Open-Source* Workflow entwickelt werden, der allerdings auf der Nutzung verschiedener Softwareprogramme beruhte und daher recht umständlich ist. Für einen Routineeinsatz der eingesetzten Gerätekopplung ist es folglich unabdingbar, eine einfache Softwarelösungen für die Endnutzer zu entwickeln. Zudem musste bei dem verwendeten DART-MS-Gerät vor jeder Messserie eine Neujustierung der DART-Ionenquelle vorgenommen werden, was die Vergleichbarkeit zwischen den Messdaten, die zu unterschiedlichen Zeitpunkten aufgenommen erschwerte.

Einige dieser Herausforderungen wurden vor Kurzem seitens der Firma Bruker Daltonics angegangen, welche seit 2024 kommerziell ein DART-MS-Kopplung mit einem *Tripple Quadrupol*-MS (QQQ-MS) vertreibt. Dieses Gerät kann aufgrund des verbauten QQQ-Analysators zwar nicht für *non-targeted*-Experimente eingesetzt werden, eignet sich aber durchaus für schnelle *targeted*-Experimente, wie anhand einiger kürzlich erschienen Publikationen deutlich wird (Bruker Daltonics 2023; Losso et al. 2022; Hsu et al. 2024). Durch die Verwendung eines QQQ-Analysators anstatt eines qToF-Analysators kann zudem der Anschaffungspreis deutlich gesenkt werden.

Nachteilig ist bei dem Einsatz von DART-Ionenquellen, wie bei allen Direktionisationsmethoden, das Auftreten von Ionensuppressionen. Bei Verwendung einer vorgelagerten Chromatographie werden nur einige Metabolite gleichzeitig ionisiert und Ionensuppressionen können so minimiert werden (Perkons et al. 2021; Wang und Liu 2024).

Abhängig von der Fragestellung können bei DART-Ionenquellen mehrere Parameter optimiert werden, um bei einer *targeted* Analytik die Ionensuppression zu minimieren und die Ionisierung der gesuchten Metabolite zu erleichtern. Ergänzend wäre eine Aufreinigung durch spezifische Festphasenextraktionen (SPEs, engl. *solid phase extractions*) möglich, um störende Matrixbestandteile abzutrennen (Badawy et al. 2022). Alternativ wäre der Einsatz einer LC-DART-MS-Kopplung möglich, um die Analyten vorher durch eine Chromatographie zu trennen, dann mittels DART-Ionenquelle zu Ionisieren und so die Ionensuppressionen zu minimieren (Wang und Liu 2024). So ist je nach Fragestellung eine Abwägung zu treffen, ob eine kürzere Analysenzeit oder eine geringere Ionensuppression mit höherer Metabolitabdeckung bevorzugt wird.

Ebenfalls ergeben sich in der Auswertung der mittels DART-MS aufgenommenen Daten einige Herausforderungen. Beispielsweise sind bisher nur eine geringe Anzahl an Referenz- und Fragmentspektren, die mittels einer DART-Quelle aufgenommen wurden, publiziert. So ist bei vielen Stoffklassen nicht bekannt, wie diese während der DART-Ionisierung reagieren und ob eine *in source* Fragmentierung möglich ist. Diese Arbeit konnte einen Beitrag zur Aufklärung der Ionisierung von verschiedenen Lipiden, Tocopherolen und Fettsäuren leisten, jedoch müssen noch weitere Metabolitklassen vermessen werden, um mehr über die Ionisierung lernen zu können. Weitere Forschung auf dem Gebiet der DART-MS-basierten Metabolomanalyse wird dazu führen, dass ein größerer Anteil der Metabolite identifiziert werden kann und ein breiteres Verständnis dieser Ionisationstechnik erlangt wird. Ergänzend zur Aufklärung weiter Fragmentierungsmuster, wäre ein Datenbankaufbau sinnvoll, um zukünftig eine teilautomatisierte Auswertung von Spektren zu ermöglichen.

Für die weitere Etablierung von DART-Ionenquellen, sollten zukünftig zusätzliche Datensätze mittels DART-MS untersucht werden, um die breiten Einsatzzwecke der DART-Ionisierung zu demonstrieren. Insbesondere wäre die Untersuchung von weiteren wasserreichen Lebensmitteln sinnvoll, um die Eignung von DART-MS-basierter Analytik für die Authentifizierung wasserreicher Lebensmittel besser einschätzen zu können. Zudem könnten Fragestellungen untersucht werden, welche über das Gebiet der Herkunftsauthentifizierung hinausgehen. Beispiele hierfür wären die Verfälschung von Arten verschiedener Pflanzen oder Pilze wie Trüffel, die Untersuchung der Streckung von Lebensmitteln mit günstigeren Materialien wie die Mischung von Oregano mit Majoran oder Kakao und Kakaoschale.

Über die Entwicklung der *non-targeted* Ansätze hinaus, ist die Entwicklung von routinefähigen Methoden für den breiteren Einsatz von Direktionisationsmethoden notwendig, da *non-targeted* Analytik meist in der Forschung eingesetzt wird und für den Routineeinsatz häufig zu kostspielig ist (Creydt et al. 2022). Hierfür müssen bekannte, oder über *non-targeted* Experimente identifizierte Marker in eine *targeted* Methode überführt werden. In Frage

kommen die in dieser Arbeit untersuchte Datensätze mit den identifizierten Markern oder weitere bereits mittels *non-targeted* LC-MS Messungen untersuchten Fragestellungen. Für diese Übertragung bieten sich, wie oben bereits diskutiert, insbesondere routinefähige DART-QQQ Geräte an, welche einfacher gebaut sind als die verwendete DART-qToF-MS Kopplung und dadurch kosteneffizienter sind. So könnten insbesondere einfache und mittelkomplexe Fragestellungen in eine schnelle und effiziente Routineanwendung überführt werden. Die Vorteile einer Schnellmethode können insbesondere in Bereichen, welche zügige Entscheidungen erfordern, von Vorteil sein. Dies ist beispielsweise bei Wareneingangskontrollen der Fall, bei denen Entscheidungen innerhalb von Minuten getroffen werden müssen, da die Ware des Lieferanten direkt angenommen oder abgelehnt werden muss. Ein ebenfalls interessanter Einsatzbereich wäre die Prozesskontrolle von chemischen Synthesen, da auch hier der Einsatz von Chromatographie-basierten Systemen durch lange Analysenzeiten limitiert ist. In Routinelaboren bereits vorhandene Methoden, welche nur wenige *targets* untersuchen, könnten ebenfalls durch DART-basierte Methoden ersetzt werden, um die Chromatographie zu eliminieren, die Analysenzeiten signifikant zu kürzen und somit die Durchlaufzeit einer Probe zu verringern. Dieses Vorgehen bietet sich insbesondere für den klinischen und pharmazeutischen Einsatzzweck an, da hier ein hoher Probendurchsatz besteht und ein großer Trend zur Automatisierung sowie Zeitgewinn besteht (Jannetto und Fitzgerald 2016; Vogeser und Kirchhoff 2011).

Weitere Forschungsarbeiten an den Grundlagen der DART-MS-basierten Metabolomanalyse können in Kombination mit der Übertragung bestehender Methoden und der Entwicklung robuster DART-QQQ-MS Geräte für die Routineanalytik zu einem breiten Einsatz von DART-MS Geräten beitragen. Dies wird die zukünftige Anwendung von DART-Ionenquellen maßgeblich fördern und die Lebensmittelauthentifizierung mit dieser schnellen und umweltfreundlichen Analysenmethode disruptiv voranbringen.

8 Literaturverzeichnis

Adak, Nafiye; Heybeli, Nursel; Ertekin, Can (2017): Infrared drying of strawberry. In: *Food chemistry* 219, S. 109–116.

Agusta, Zahra Putri (2019): Modified balanced random forest for improving imbalanced data prediction. In: *International Journal of Advances in Intelligent Informatics* 5 (1), S. 58–65.

Ahern, Katelyn W.; Serbulea, Vlad; Wingrove, Catherine L.; Palas, Zachary T.; Leitinger, Norbert; Harris, Thurl E. (2019): Regioisomer-independent quantification of fatty acid oxidation products by HPLC-ESI-MS/MS analysis of sodium adducts. In: *Scientific Reports* 9 (1), S. 11197.

Akhatou, Ikram; González-Domínguez, Raúl; Fernández-Recamales, Ángeles (2016): Investigation of the effect of genotype and agronomic conditions on metabolomic profiles of selected strawberry cultivars with different sensitivity to environmental stress. In: *Plant Physiology and Biochemistry* 101, S. 14–22.

Amr Kadah (2021): In dieser Saison erwarten wir ein Wachstum der ägyptischen Erdbeerexporte von etwa 20 bis 30%. Hg. v. Fresh Plaza. Online verfügbar unter https://www.freshplaza.de/article/9284781/in-dieser-saison-erwarten-wir-ein-wachstum-der-agyptischen-erdbeerexporte-von-etwa-20-bis-30/, zuletzt aktualisiert am 19.01.2021, zuletzt geprüft am 24.01.2025.

An, ShuQi; Liu, Shuai; Cao, Jie; Lu, ShiFang (2019): Nitrogen-activated oxidation in nitrogen direct analysis in real time mass spectrometry (DART-MS) and rapid detection of explosives using thermal desorption DART-MS. In: *Journal of the American Society for Mass Spectrometry* 30 (10), S. 2092–2100.

Andrade, Francisco J.; Shelley, Jacob T.; Wetzel, William C.; Webb, Michael R.; Gamez, Gerardo; Ray, Steven J.; Hieftje, Gary M. (2008): Atmospheric pressure chemical ionization source. 1. Ionization of compounds in the gas phase. In: *Analytical chemistry* 80 (8), S. 2646–2653.

Arndt, Maike; Drees, Alissa; Ahlers, Christian; Fischer, Markus (2020a): Determination of the geographical origin of walnuts (Juglans regia L.) using near-infrared spectroscopy and chemometrics. In: *Foods* 9 (12), S. 1860.

Arndt, Maike; Rurik, Marc; Drees, Alissa; Ahlers, Christian; Feldmann, Simon; Kohlbacher, Oliver; Fischer, Markus (2021): Food authentication: Determination of the geographical origin of almonds (Prunus dulcis MILL.) via near-infrared spectroscopy. In: *Microchemical Journal* 160, S. 105702.

Arndt, Maike; Rurik, Marc; Drees, Alissa; Bigdowski, Katharina; Kohlbacher, Oliver; Fischer, Markus (2020b): Comparison of different sample preparation techniques for NIR screening and their influence on the geographical origin determination of almonds (Prunus dulcis MILL.). In: *Food Control* 115, S. 107302.

Avula, Bharathi; Smillie, Troy J.; Wang, Yan-Hong; Zweigenbaum, Jerry; Khan, Ikhlas A. (2015): Authentication of true cinnamon (Cinnamon verum) utilising direct analysis in real time (DART)-QToF-MS. In: *Food Additives & Contaminants: Part A* 32 (1), S. 1–8.

Awad, Hanan; Khamis, Mona M.; El-Aneed, Anas (2015): Mass spectrometry, review of the basics: ionization. In: *Applied spectroscopy reviews* 50 (2), S. 158–175.

Badawy, Mohamed E. I.; El-Nouby, Mahmoud am; Kimani, Paul K.; Lim, Lee W.; Rabea, Entsar I. (2022): A review of the modern principles and applications of solid-phase extraction techniques in chromatographic analysis. In: *Analytical Sciences* 38 (12), S. 1457–1487.

Ballabio, Davide; Skov, Thomas; Leardi, Riccardo; Bro, Rasmus (2008): Classification of GC-MS measurements of wines by combining data dimension reduction and variable selection techniques. In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 22 (8), S. 457–463.

Beć, Krzysztof B.; Grabska, Justyna; Huck, Christian W. (2021): NIR spectroscopy of natural medicines supported by novel instrumentation and methods for data analysis and interpretation. In: *Journal of Pharmaceutical and Biomedical Analysis* 193, S. 113686.

Bertoldi, Daniela; Barbero, Alice; Camin, Federica; Caligiani, Augusta; Larcher, ROBERTO (2016): Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. In: *Food Control* 65, S. 46–53.

Bligh, E. Graham; Dyer, W. Justin (1959): A rapid method of total lipid extraction and purification. In: *Canadian journal of biochemistry and physiology* 37 (8), S. 911–917.

Blueberries - Magazine and Consulting (2024): Spain: Strawberry production barely reaches 9% of what was expected in the campaign, half of what is usual. Online verfügbar unter https://blueberriesconsulting.com/en/produccion-de-fresa-apenas-alcanza-un-9-de-la-prevista-en-campana-la-mitad-de-lo-habitual/, zuletzt aktualisiert am 26.02.2024, zuletzt geprüft am 24.01.2025.

Boccard, Julien; Veuthey, Jean-Luc; Rudaz, Serge (2010): Knowledge discovery in metabolomics: an overview of MS data handling. In: *Journal of separation science* 33 (3), S. 290–304.

Bontempo, Luana; Camin, Federica; Larcher, ROBERTO; Nicolini, GIORGIO; Perini, MATTEO; Rossmann, A. (2009): Coast and year effect on H, O and C stable isotope

ratios of Tyrrhenian and Adriatic italian olive oils. In: *Rapid communications in mass spectrometry* 23 (7), S. 1043–1048.

Bortz, Jürgen; Schuster, Christof (2011): Statistik für human-und sozialwissenschaftler: limitierte sonderausgabe: Springer-Verlag.

Boukid, Fatma; Abbattangelo, Simone; Carini, Eleonora; Marseglia, Angela; Caligiani, Augusta; Vittadini, Elena (2019): Geographical origin discrimination of Pistachio (Pistacia vera L.) through combined analysis of physical and chemical features. In: *European Food Research and Technology* 245 (1), S. 143–150.

Breiman, Leo (2001): Random forests. In: *Machine learning* 45, S. 5–32.

Brockelt, Johannes; Schmauder, Felix; Creydt, Marina; Fischer, Markus (2025): Lipid profiling: proving the geographical origin of strawberries (Fragaria× ananassa) using a non-targeted LC-IM-MS approach. In: *Molecular Omics*.

Bruker Daltonics (2023): Bruker Introduces Novel EVOQ™ DART-TQ+ for Chromatography-Free Mass Spectrometry Aimed at the Applied Market Segments. Hg. v. Bruker Daltonics. Online verfügbar unter https://ir.bruker.com/press-releases/press-release-details/2023/Bruker-Introduces-Novel-EVOQ-DART-TQ-for-Chromatography-Free-Mass-Spectrometry-Aimed-at-the-Applied-Market-Segments/default.aspx.

Bundesanstalt für Landwirtschaft und Ernährung (2024): Verbrauch von Obst nach Arten in den Jahren 2010/11-2023/24. Bundesanstalt für Landwirtschaft und Ernährung. Online verfügbar unter https://www.bmel-statistik.de/fileadmin/daten/4040800-0000.xlsx, zuletzt aktualisiert am 10.04.2024, zuletzt geprüft am 24.01.2025.

Bundeszentrum für Ernährung (2023): Erdbeeren - Erzeugung. Die meisten hier angebotenen Erdbeeren stammen aus Deutschland, wegen der großen Nachfrage werden aber auch viele Früchte, vor allem aus Spanien, importiert. Bundeszentrum für Ernährung. Online verfügbar unter https://www.bzfe.de/lebensmittel/vom-acker-bis-zum-teller/erdbeeren/erdbeeren-erzeugung/, zuletzt aktualisiert am 23.11.2023, zuletzt geprüft am 24.01.2025.

BVL (2024): Waldheidelbeeren häufig durch günstigere Arten ersetzt. Bundesweite OPSON-Operation deckt falsche Deklarierung von Produkten auf. Bundesamt für Verbraucherschutz und Lebensmittelsicherheit. Online verfügbar unter https://www.bvl.bund.de/SharedDocs/Pressemitteilungen/01_lebensmittel/2024/OPSON-XIII.html?nn=19752764, zuletzt aktualisiert am 23.10.2024, zuletzt geprüft am 07.08.2025.

Byrdwell, Wm Craig; Neff, William E. (2002): Dual parallel electrospray ionization and atmospheric pressure chemical ionization mass spectrometry (MS), MS/MS and

MS/MS/MS for the analysis of triacylglycerols and triacylglycerol oxidation products. In: *Rapid communications in mass spectrometry* 16 (4), S. 300–319.

Caboi, R.; Cristini, A.; Frau, Franco (2021): Metal concentrations in rainwater (Sardinia, Western Mediterranean Sea). In: Water-Rock Interaction: Routledge, S. 337–340.

Camin, Federica; Boner, Markus; Bontempo, Luana; Fauhl-Hassek, Carsten; Kelly, Simon D.; Riedl, Janet; Rossmann, Andreas (2017): Stable isotope techniques for verifying the declared geographical origin of food in legal cases. In: *Trends in Food Science & Technology* 61, S. 176–187.

Capocasa, Franco; Bordi, M.; Mezzetti, Bruno: Comparing frigo and fresh plants in non-fumigated and heavy soil: The response of 10 strawberry genotypes. In: VI International Strawberry Symposium 842, S. 129–134.

Capuano, Edoardo; Boerrigter-Eenling, Rita; van der Veer, Grishja; van Ruth, Saskia M. (2013): Analytical authentication of organic products: an overview of markers. In: *Journal of the Science of Food and Agriculture* 93 (1), S. 12–28.

Cervantes, Lucía; Ariza, María Teresa; Miranda, Luis; Lozano, David; Medina, Juan Jesús; Soria, Carmen; Martínez-Ferri, Elsa (2020): Stability of fruit quality traits of different strawberry varieties under variable environmental conditions. In: *Agronomy* 10 (9), S. 1242.

Cipriano, Lauren E.; Rupar, C. Anthony; Zaric, Gregory S. (2007): The cost-effectiveness of expanding newborn screening for up to 21 inherited metabolic disorders using tandem mass spectrometry: results from a decision-analytic model. In: *Value in Health* 10 (2), S. 83–97.

Cody, Robert B.; Laramée, James A.; Durst, H. Dupont (2005): Versatile new ion source for the analysis of materials in open air under ambient conditions. In: *Analytical chemistry* 77 (8), S. 2297–2302.

Corain, Giorgio (2023): EU Customs Procedure During The Russo-Ukrainian War: the prohibitions on the export of dual-use goods and the sanctions on Russia.

Cornish, Timothy J.; Cotter, Robert J. (1997): High-order kinetic energy focusing in an end cap reflectron time-of-flight mass spectrometer. In: *Analytical chemistry* 69 (22), S. 4615–4618.

Creydt, Marina (2019): Food profiling : massenspektrometrische Strategien zur geographischen Herkunftsbestimmung von weißem Spargel (Asparagus officinalis). Dissertation.

Creydt, Marina; Arndt, Maike; Hudzik, Daria; Fischer, Markus (2018a): Plant metabolomics: evaluation of different extraction parameters for nontargeted UPLC-ESI-QTOF-mass

spectrometry at the example of white Asparagus officinalis. In: *Journal of Agricultural and Food Chemistry* 66 (48), S. 12876–12887.

Creydt, Marina; Fischer, Markus (2020): Mass-spectrometry-based food metabolomics in routine applications: A basic standardization approach using housekeeping metabolites for the authentication of asparagus. In: *Journal of Agricultural and Food Chemistry* 68 (49), S. 14343–14352.

Creydt, Marina; Flügge, Friedemann; Dammann, Robin; Schütze, Burkhard; Günther, Ulrich L.; Fischer, Markus (2023): Food Fingerprinting: LC-ESI-IM-QTOF-Based Identification of Blumeatin as a New Marker Metabolite for the Detection of Origanum majorana Admixtures to O. onites/vulgare. In: *Metabolites* 13 (5), S. 673.

Creydt, Marina; Hudzik, Daria; Rurik, Marc; Kohlbacher, Oliver; Fischer, Markus (2018b): Food authentication: Small-molecule profiling as a tool for the geographic discrimination of German white asparagus. In: *Journal of Agricultural and Food Chemistry* 66 (50), S. 13328–13339.

Creydt, Marina; Wegner, Benjamin; Gnauck, Alina; Hörner, Roy; Hummert, Christian; Fischer, Markus (2022): Food authentication in the routine laboratory: Determination of the geographical origin of white asparagus using a simple targeted LC-ESI-QqQ-MS/MS approach. In: *Food Control* 135, S. 108690.

Cubero-Leon, Elena; Peñalver, Rosa; Maquet, Alain (2014): Review on metabolomics for food authentication. In: *Food Research International* 60, S. 95–107.

Da Silva, Ricardo R.; Dorrestein, Pieter C.; Quinn, Robert A. (2015): Illuminating the dark matter in metabolomics. In: *Proceedings of the National Academy of Sciences* 112 (41), S. 12549–12550.

DalCorso, Giovanni; Manara, Anna; Piasentin, Silvia; Furini, Antonella (2014): Nutrient metal elements in plants. In: *Metallomics* 6 (10), S. 1770–1788.

Degenhardt, Frauke; Seifert, Stephan; Szymczak, Silke (2019): Evaluation of variable selection methods for random forests and omics data sets. In: *Briefings in bioinformatics* 20 (2), S. 492–503.

Delegierte Verordnung (EU) 2023/2429: Delegierte Verordnung (EU) 2023/2429 der Kommission vom 17. August 2023 zur Ergänzung der Verordnung (EU) Nr. 1308/2013 des Europäischen Parlaments und des Rates hinsichtlich der Vermarktungsnormen für den Sektor Obst und Gemüse, bestimmte Verarbeitungserzeugnisse aus Obst und Gemüse und den Bananensektor, und zur Aufhebung der Verordnung (EG) Nr. 1666/1999 der Kommission und der Durchführungsverordnungen (EU) Nr. 543/2011 und (EU) Nr. 1333/2011 der Kommission.

Di Guida, Riccardo; Engel, Jasper; Allwood, J. William; Weber, Ralf J. M.; Jones, Martin R.; Sommer, Ulf et al. (2016): Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. In: *Metabolomics* 12, S. 1–14.

Díaz, Ramon; Pozo, Oscar J.; Sancho, Juan V.; Hernández, Félix (2014): Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. In: *Food chemistry* 157, S. 84–93.

Do, Kieu Trinh; Wahl, Simone; Raffler, Johannes; Molnos, Sophie; Laimighofer, Michael; Adamski, Jerzy et al. (2018): Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. In: *Metabolomics* 14, S. 1–18.

Drees, Alissa; Brockelt, Johannes; Cvancar, Lina; Fischer, Markus (2023): Rapid determination of the shell content in cocoa products using FT-NIR spectroscopy and chemometrics. In: *Talanta* 256, S. 124310.

Drexler, Dieter M.; Reily, Michael D.; Shipkova, Petia A. (2011): Advances in mass spectrometry applied to pharmaceutical metabolomics. In: *Analytical and bioanalytical chemistry* 399, S. 2645–2653.

Emery, Justin; Pohl, Benjamin; Crétat, Julien; Richard, Yves; Pergaud, Julien; Rega, Mario et al. (2021): How local climate zones influence urban air temperature: Measurements by bicycle in Dijon, France. In: *Urban Climate* 40, S. 101017.

Erenstein, Olaf; Jaleta, Moti; Sonder, Kai; Mottaleb, Khondoker; Prasanna, Boddupalli M. (2022): Global maize production, consumption and trade: trends and R&D implications. In: *Food security* 14 (5), S. 1295–1319.

Escobedo, René; Miranda, René; Martínez, Joel (2016): Infrared irradiation: Toward green chemistry, a review. In: *International Journal of Molecular Sciences* 17 (4), S. 453.

EUROPOL (2024): EUR 91 million worth of counterfeit and substandard food seized in Europe-wide operation. Partners involved in operation OPSON XIII took 22 000 tonnes of food and 850 000 litres of beverages off the market. EUROPOL. Online verfügbar unter https://www.europol.europa.eu/media-press/newsroom/news/15-000-tonnes-of-illegal-food-and-beverages-market, zuletzt aktualisiert am 23.10.2024, zuletzt geprüft am 07.08.2025.

Eurostat (2025): Erntemenge von Erdbeeren in der Europäischen Uniun nach Ländern in den Jahren 2021 bis 2023. Eurostat. Online verfügbar unter https://ec.europa.eu/eurostat/databrowser/view/apro_cpsh1__custom_12081755/book

mark/table?lang=de&bookmarkId=31f30abc-f4d1-4c96-b393-7abb959f0c30, zuletzt aktualisiert am 22.01.2025, zuletzt geprüft am 24.01.2025.

Fahad, Shah; Bajwa, Ali A.; Nazir, Usman; Anjum, Shakeel A.; Farooq, Ayesha; Zohaib, Ali et al. (2017): Crop production under drought and heat stress: plant responses and management options. In: *Frontiers in plant science* 8, S. 1147.

Fisher, Christine M.; Peter, Katherine T.; Newton, Seth R.; Schaub, Andrew J.; Sobus, Jon R. (2022): Approaches for assessing performance of high-resolution mass spectrometry–based non-targeted analysis methods. In: *Analytical and bioanalytical chemistry* 414 (22), S. 6455–6471.

Food and Agriculture Organization of the United Nations (2024): Production Quantity of Hazelnuts 2013-2023. Online verfügbar unter https://www.fao.org/faostat/en/#data/QCL, zuletzt aktualisiert am 20.12.2024, zuletzt geprüft am 23.01.2025.

Fung, Fred; Wang, Huei-Shyong; Menon, Suresh (2018): Food safety in the 21st century. In: *Biomedical journal* 41 (2), S. 88–95.

Furey, Ambrose; Moriarty, Merisa; Bane, Vaishali; Kinsella, Brian; Lehane, Mary (2013): Ion suppression; a critical review on causes, evaluation, prevention and applications. In: *Talanta* 115, S. 104–122.

Ge, Zhuang (2024): Overview of High-Performance Timing and Position-Sensitive MCP Detectors Utilizing Secondary Electron Emission for Mass Measurements of Exotic Nuclei at Nuclear Physics Facilities. In: *Sensors* 24 (22), S. 7261.

Ghosh, Chinmoy; Shinde, Chandrakant P.; Chakraborty, Bhaswat S. (2012): Influence of ionization source design on matrix effects during LC–ESI-MS/MS analysis. In: *Journal of Chromatography B* 893, S. 193–200.

Ghosson, Hikmat; Guitton, Yann; Ben Jrad, Amani; Patil, Chandrashekhar; Raviglione, Delphine; Salvia, Marie-Virginie; Bertrand, Cédric (2021): Electrospray ionization and heterogeneous matrix effects in liquid chromatography/mass spectrometry based meta-metabolomics: a biomarker or a suppressed ion? In: *Rapid communications in mass spectrometry* 35 (2), e8977.

Giera, Martin; Aisporna, Aries; Uritboonthai, Winnie; Siuzdak, Gary (2024): The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. In: *Nature metabolism* 6 (9), S. 1647–1648.

Gilbert, E. R.; Wong, E. A.; Webb Jr, K. E. (2008): Board-invited review: peptide absorption and utilization: implications for animal nutrition and health. In: *Journal of animal science* 86 (9), S. 2135–2155.

Gönenc, Sertac; Tanrıvermis, Harun; Bülbül, Mehmet (2006): Economic assessment of hazelnut production and the importance of supply management approaches in Turkey. In: *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)* 107 (1), S. 19–32.

Górnaś, Paweł; Baškirovs, Georgijs; Siger, Aleksander (2022): Free and esterified tocopherols, tocotrienols and other extractable and non-extractable tocochromanol-related molecules: Compendium of knowledge, future perspectives and recommendations for chromatographic techniques, tools, and approaches used for tocochromanol determination. In: *Molecules* 27 (19), S. 6560.

Gross, Jürgen H. (2014): Direct analysis in real time—a critical review on DART-MS. In: *Analytical and bioanalytical chemistry* 406, S. 63–80.

Gross, Jürgen H. (2019): Massenspektrometrie. Berlin, Heidelberg: Springer. Online verfügbar unter https://doi.org/10.1007/978-3-662-58635-8.

Guilhaus, Michael (1995): Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. In: *Journal of mass spectrometry* 30 (11), S. 1519–1532.

Guo, Juan; Zhang, Maomao; Liu, Jian'an; Luo, Rupeng; Yan, Tingting; Yang, Tao et al. (2020): Evaluation of the deterioration state of archaeological wooden artifacts: a nondestructive protocol based on direct analysis in real time–mass spectrometry (DART-MS) coupled to chemometrics. In: *Analytical chemistry* 92 (14), S. 9908–9915.

Hansen, Jule; Kunert, Christof; Raezke, Kurt-Peter; Seifert, Stephan (2024): Detection of Sugar Syrups in Honey Using Untargeted Liquid Chromatography–Mass Spectrometry and Chemometrics. In: *Metabolites* 14 (11), S. 633.

Hayashi, Yuji; Komatsu, Takanori; Iwashita, Kazuhiro; Fukusaki, Eiichiro (2021): 1H-NMR metabolomics-based classification of Japanese sake and comparative metabolome analysis by gas chromatography–mass spectrometry. In: *Journal of bioscience and bioengineering* 131 (5), S. 557–564.

Hernández-Martínez, Nelda R.; Blanchard, Caroline; Wells, Daniel; Salazar-Gutiérrez, Melba R. (2023): Current state and future perspectives of commercial strawberry production: A review. In: *Scientia Horticulturae* 312, S. 111893.

Hrydziuszko, Olga; Viant, Mark R. (2012): Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. In: *Metabolomics* 8, S. 161–174.

Hsu, Wei-Hsin; Cheng, Kai-Wen; Feng, Tzu-Hsuan; Chen, Ju-Yu; Chen, Guan-Yuan; Chen, Lian-Yu et al. (2024): Rapid screening of new psychoactive substances using pDART-

QqQ-MS. In: *Journal of the American Society for Mass Spectrometry* 35 (6), S. 1370–1376.

Ioanoviciu, Damaschin (1995): Ion-Optical solutions in time-of-flight mass spectrometry. In: *Rapid communications in mass spectrometry* 9 (11), S. 985–997.

Isaksson, Tomas; Næs, Tormod (1988): The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. In: *Applied Spectroscopy* 42 (7), S. 1273–1284.

Izsák, Beatrix; Szentimrey, Tamás; Lakatos, Mónika; Pongrácz, Rita; Szentes, Olivér (2022): Creation of a representative climatological database for Hungary from 1870 to 2020. In: *Időjárás* 126 (1).

Jannetto, Paul J.; Fitzgerald, Robert L. (2016): Effective use of mass spectrometry in the clinical laboratory. In: *Clinical chemistry* 62 (1), S. 92–98.

Jastrzembski, Jillian A.; Sacks, Gavin L. (2016): Solid phase mesh enhanced sorption from headspace (SPMESH) coupled to DART-MS for rapid quantification of trace-level volatiles. In: *Analytical chemistry* 88 (17), S. 8617–8623.

Johnson, Renée (2014): Food fraud and economically motivated adulteration of food and food ingredients: Congressional research service Washington, DC.

Karahalil, Bensu (2016): Overview of systems biology and omics technologies. In: *Current medicinal chemistry* 23 (37), S. 4221–4230.

Karaman, Ibrahim (2017): Preprocessing and pretreatment of metabolomics data for statistical analysis. In: *Metabolomics: From fundamentals to clinical applications*, S. 145–161.

Katerinopoulou, Katerina; Kontogeorgos, Achilleas; Salmas, Constantinos E.; Patakas, Angelos; Ladavos, Athanasios (2020): Geographical origin authentication of agri-food products: A review. In: *Foods* 9 (4), S. 489.

Kelly, Simon; Heaton, Karl; Hoogewerff, Jurian (2005): Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. In: *Trends in Food Science & Technology* 16 (12), S. 555–567.

Khodaei, Diako; Javanmardi, Fardin; Khaneghah, Amin Mousavi (2021): The global overview of the occurrence of mycotoxins in cereals: A three-year survey. In: *Current Opinion in Food Science* 39, S. 36–42.

Kim, Hye Jin; Seo, Yong Taek; Park, Sang-il; Jeong, Se Hee; Kim, Min Kyoung; Jang, Young Pyo (2015): DART–TOF–MS based metabolomics study for the discrimination analysis of geographical origin of Angelica gigas roots collected from Korea and China. In: *Metabolomics* 11, S. 64–70.

Kirpich, Alexander; Ainsworth, Elizabeth A.; Wedow, Jessica M.; Newman, Jeremy R. B.; Michailidis, George; McIntyre, Lauren M. (2018): Variable selection in omics data: A practical evaluation of small sample sizes. In: *PloS one* 13 (6), e0197910.

Kishimoto, Naoki; Furuhashi, Masayuki; Ohno, Koichi (1998): Two-dimensional penning ionization electron spectrum of N2 by collision with He*(23S) metastable atoms. In: *Journal of electron spectroscopy and related phenomena* 88, S. 143–147.

Klampfl, Christian W. (2019): Metabolization of pharmaceuticals by plants after uptake from water and soil: a review. In: *TrAC Trends in Analytical Chemistry* 111, S. 13–26.

Klockmann, Sven; Reiner, Eva; Bachmann, René; Hackl, Thomas; Fischer, Markus (2016): Food fingerprinting: metabolomic approaches for geographical origin discrimination of hazelnuts (Corylus avellana) by UPLC-QTOF-MS. In: *Journal of Agricultural and Food Chemistry* 64 (48), S. 9253–9262.

Kobi, H. B.; Martins, M. C.; Silva, P. I.; Souza, J. L.; Carneiro, J. C.S.; Heleno, F. F. et al. (2018): Organic and conventional strawberries: Nutritional quality, antioxidant characteristics and pesticide residues. In: *Fruits* 73 (1), S. 39–47.

Kravchenko, Ekaterina; Minkina, Tatiana; Mandzhieva, Saglara; Bauer, Tatiana; Lacynnik, Elizaveta; Wong, Ming Hung; Nazarenko, Olga (2025): Ecological and health risk assessments of heavy metal contamination in soils surrounding a coal power plant. In: *Journal of Hazardous Materials* 484, S. 136751.

Kue Foka, Idrice Carther; Ketehouli, Toi; Zhou, Yonggang; Li, Xiao-Wei; Wang, Fa-Wei; Li, Haiyan (2020): The emerging roles of diacylglycerol kinase (DGK) in plant stress tolerance, growth, and development. In: *Agronomy* 10 (9), S. 1375.

Kursa, Miron B.; Rudnicki, Witold R. (2010): Feature selection with the Boruta package. In: *Journal of statistical software* 36, S. 1–13.

Labory, Justine; Njomgue-Fotso, Evariste; Bottini, Silvia (2024): Benchmarking feature selection and feature extraction methods to improve the performances of machine-learning algorithms for patient classification using metabolomics biomedical data. In: *Computational and Structural Biotechnology Journal* 23, S. 1274–1287.

Li, Bin; Friedman, J.; Olshen, R.; Stone, C. (1984): Classification and regression trees (CART). In: *Biometrics* 40 (3), S. 358–361.

Li, Chang; Chu, Shiying; Tan, Siyuan; Yin, Xinchi; Jiang, You; Dai, Xinhua et al. (2021): Towards higher sensitivity of mass spectrometry: A perspective from the mass analyzers. In: *Frontiers in chemistry* 9, S. 813359.

Li, Zhucui; Lu, Yan; Guo, Yufeng; Cao, Haijie; Wang, Qinhong; Shui, Wenqing (2018): Comprehensive evaluation of untargeted metabolomics data processing software in

feature detection, quantification and discriminating marker selection. In: *Analytica Chimica Acta* 1029, S. 50–57.

Looser, Nadja; Kostelac, Drazen; Scherbaum, Ellen; Anastassiades, Michelangelo; Zipper, Hubert (2006): Pesticide residues in strawberries sampled from the market of the Federal State of Baden-Württemberg in the period between 2002 and 2005. In: *Journal für Verbraucherschutz und Lebensmittelsicherheit* 1 (2), S. 135–141.

Lösel, Henri; Arndt, Maike; Wenck, Soeren; Hansen, Lasse; Oberpottkamp, Marie; Seifert, Stephan; Fischer, Markus (2024): Exploring the potential of high-resolution LC-MS in combination with ion mobility separation and surrogate minimal depth for enhanced almond origin authentication. In: *Talanta* 271, S. 125598.

Lösel, Henri; Brockelt, Johannes; Gärber, Florian; Teipel, Jan; Kuballa, Thomas; Seifert, Stephan; Fischer, Markus (2023): Comparative Analysis of LC-ESI-IM-qToF-MS and FT-NIR Spectroscopy Approaches for the Authentication of Organic and Conventional Eggs. In: *Metabolites* 13 (8), S. 882.

Lösel, Henri; Shakiba, Navid; Wenck, Soeren; Le Tan, Phat; Arndt, Maike; Seifert, Stephan et al. (2022): Impact of Freeze-Drying on the determination of the geographical origin of almonds (Prunus dulcis Mill.) by near-infrared (NIR) spectroscopy. In: *Food Analytical Methods* 15 (10), S. 2847–2857.

Losso, Klemens; Bec, Krzysztof B.; Mayr, Sophia; Grabska, Justyna; Stuppner, Stefan; Jones, Michael et al. (2022): Rapid discrimination of Curcuma longa and Curcuma xanthorrhiza using direct analysis in real time mass spectrometry and near infrared spectroscopy. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 265, S. 120347.

Losso, Klemens; Wörz, Hannah; Kappacher, Christoph; Huber, Susanne; Jakschitz, Thomas; Rainer, Matthias; Bonn, Günther K. (2023): Rapid quality control of black truffles using direct analysis in real time mass spectrometry and hydrophilic interaction liquid chromatography mass spectrometry. In: *Food chemistry* 403, S. 134418.

Mamyrin, B. A. (1994): Laser assisted reflectron time-of-flight mass spectrometry. In: *International Journal of Mass Spectrometry and Ion Processes* 131, S. 1–19.

Manly, Bryan F. J. (2018): Randomization, bootstrap and Monte Carlo methods in biology: chapman and hall/CRC.

Mann, M. (2011): John Bennett Fenn, 1917–2010: ACS Publications (22). In: *Journal of the American Society for Mass Spectrometry* (3).

Manning, Louise (2016): Food fraud: Policy and food chain. In: *Current Opinion in Food Science* 10, S. 16–21.

Massaro, Andrea; Stella, Roberto; Negro, Alessandro; Bragolusi, Marco; Miano, Brunella; Arcangeli, Giuseppe et al. (2021): New strategies for the differentiation of fresh and frozen/thawed fish: A rapid and accurate non-targeted method by ambient mass spectrometry and data fusion (part A). In: *Food Control* 130, S. 108364.

McMillen, Teresa S.; Leslie, Andrew; Chisholm, Kenneth; Penny, Susanne; Gallant, Jeffrey; Cohen, Alejandro et al. (2023): A large-scale, targeted metabolomics method for the analysis and quantification of metabolites in human plasma via liquid chromatography-mass spectrometry. In: *Analytica Chimica Acta* 1279, S. 341791.

Mehlenbacher, Shawn A.; Molnar, Thomas J. (2021): Hazelnut breeding. In: *Plant breeding reviews* 45, S. 9–141.

Michelini, Elisa; Cevenini, Luca; Mezzanotte, Laura; Simoni, Patrizia; Baraldini, Mario; Laude, Luca de; Roda, Aldo (2007): One-step triplex-polymerase chain reaction assay for the authentication of yellowfin (Thunnus albacares), bigeye (Thunnus obesus), and skipjack (Katsuwonus pelamis) tuna DNA from fresh, frozen, and canned tuna samples. In: *Journal of Agricultural and Food Chemistry* 55 (19), S. 7638–7647.

Moros, Georgios; Chatziioannou, Anastasia Chrysovalantou; Gika, Helen G.; Raikos, Nikolaos; Theodoridis, Georgios (2017): Investigation of the derivatization conditions for GC–MS metabolomics of biological samples. In: *Bioanalysis* 9 (1), S. 53–65.

Müller-Maatsch, Judith; van Ruth, Saskia M. (2021): Handheld devices for food authentication and their applications: A review. In: *Foods* 10 (12), S. 2901.

Nagy, Maii M.; Wang, Shengpeng; Farag, Mohamed A. (2022): Quality analysis and authentication of nutraceuticals using near IR (NIR) spectroscopy: A comprehensive review of novel trends and applications. In: *Trends in Food Science & Technology* 123, S. 290–309.

Nassar, Lobna; Okwuchi, Ifeanyi Emmanuel; Saad, Muhammad; Karray, Fakhri; Ponnambalam, Kumaraswamy; Agrawal, Prarabdha (2020): Prediction of strawberry yield and farm price utilizing deep learning. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, S. 1–7. Online verfügbar unter https://doi.org/10.1109/IJCNN48605.2020.9206998.

Neto, Fausto Carnevale; Pascua, Vadim; Raftery, Daniel (2021): Formation of sodium cluster ions Complicates LC-MS metabolomics analyses. In: *Rapid communications in mass spectrometry: RCM* 35 (20), e9175.

Nezhadahmadi, Arash; Faruq, Golam; Rashid, Kamaludin (2015): The impact of drought stress on morphological and physiological parameters of three strawberry varieties in different growing conditions. In: *Pakistan Journal of Agricultural Sciences* 52 (1).

Novotná, H.; Kmiecik, O.; Gałązka, M.; Krtková, V.; Hurajová, A.; Schulzová, V. et al. (2012): Metabolomic fingerprinting employing DART-TOFMS for authentication of tomatoes and peppers from organic and conventional farming. In: *Food Additives & Contaminants: Part A* 29 (9), S. 1335–1346.

Oğuz, İlbilge; Oğuz, Halil İbrahim; Kafkas, Nesibe Ebru (2022): Strawberry cultivation techniques. In: Recent Studies on Strawberries: IntechOpen.

Pandey, Aditi; Tiwari, Nidhi; Sahu, Amrita; Baishya, Bikash (2025): HSQC/F1-PSYCHE TOCSY NOAH Supersequence for High-Resolution NMR Analysis of Urine Metabolites. In: *Magnetic Resonance in Chemistry*.

Pearson, Karl (1901): LIII. On lines and planes of closest fit to systems of points in space. In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2 (11), S. 559–572.

Perkons, Ingus; Rusko, Janis; Zacs, Dzintars; Bartkevics, Vadims (2021): Rapid determination of pharmaceuticals in wastewater by direct infusion HRMS using target and suspect screening analysis. In: *Science of The Total Environment* 755, S. 142688.

Pichersky, Eran; Lewinsohn, Efraim (2011): Convergent evolution in plant specialized metabolism. In: *Annual review of plant biology* 62 (1), S. 549–566.

Pluskal, Tomáš; Castillo, Sandra; Villar-Briones, Alejandro; Orešič, Matej (2010): MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. In: *BMC bioinformatics* 11, S. 1–11.

Podwojski, Katharina; Fritsch, Arno; Chamrad, Daniel C.; Paul, Wolfgang; Sitek, Barbara; Stühler, Kai et al. (2009): Retention time alignment algorithms for LC/MS data must consider non-linear shifts. In: *Bioinformatics* 25 (6), S. 758–764.

Powers, Robert; Andersson, Erik R.; Bayless, Amanda L.; Brua, Robert B.; Chang, Mario C.; Cheng, Leo L. et al. (2024): Best practices in NMR metabolomics: Current state. In: *TrAC Trends in Analytical Chemistry* 171, S. 117478.

Prabhu, Gurpur Rakesh D.; Williams, Evan R.; Wilm, Matthias; Urban, Pawel L. (2023): Mass spectrometry using electrospray ionization. In: *Nature Reviews Methods Primers* 3 (1), S. 23.

Quarshi, Huma Qureshi; Ahmed, Waseem; Azmant, Rafia; Chendouh-Brahmi, Nabila; Quyyum, Abdul; Abbas, Asad (2023): Post-harvest problems of strawberry and their solutions. In: Recent studies on strawberries: IntechOpen.

Reeve, Jennifer R.; Smith, Jeffrey L.; Carpenter-Boggs, Lynne; Reganold, John P. (2008): Soil-based cycling and differential uptake of amino acids by three species of strawberry (Fragaria spp.) plants. In: *Soil Biology and Biochemistry* 40 (10), S. 2547–2552.

Richter, Bernadette; Gurk, Stephanie; Wagner, Deniz; Bockmayr, Michael; Fischer, Markus (2019): Food authentication: Multi-elemental analysis of white asparagus for provenance discrimination. In: *Food chemistry* 286, S. 475–482.

Ringnér, Markus (2008): What is principal component analysis? In: *Nature biotechnology* 26 (3), S. 303–304.

Roca, Marta; Alcoriza, Maria Isabel; Garcia-Cañaveras, Juan Carlos; Lahoz, Agustín (2021): Reviewing the metabolome coverage provided by LC-MS: Focus on sample preparation and chromatography-A tutorial. In: *Analytica Chimica Acta* 1147, S. 38–55.

Roemmelt, Andreas T.; Steuer, Andrea E.; Poetzsch, Michael; Kraemer, Thomas (2014): Liquid chromatography, in combination with a quadrupole time-of-flight instrument (LC QTOF), with sequential window acquisition of all theoretical fragment-ion spectra (SWATH) acquisition: systematic studies on its use for screenings in clinical and forensic toxicology and comparison with information-dependent acquisition (IDA). In: *Analytical chemistry* 86 (23), S. 11742–11749.

Rouf Shah, Tajamul; Prasad, Kamlesh; Kumar, Pradyuman (2016): Maize—A potential source of human nutrition and health: A review. In: *Cogent Food & Agriculture* 2 (1), S. 1166995.

Rubingh, Carina M.; Bijlsma, Sabina; Derks, Eduard PPA; Bobeldijk, Ivana; Verheij, Elwin R.; Kochhar, Sunil; Smilde, Age K. (2006): Assessing the performance of statistical validation tools for megavariate metabolomics data. In: *Metabolomics* 2, S. 53–61.

Rukundo, Isaac R.; Danao, Mary-Grace C.; Mitchell, Robert B.; Masterson, Steven D.; Weller, Curtis L. (2021): Comparing the use of handheld and benchtop NIR spectrometers in predicting nutritional value of forage. In: *Applied Engineering in Agriculture* 37 (1), S. 171–181.

Saang'onyo, Daudi S.; Smith, Darrin L. (2012): Optimization of direct analysis in real time (DART) linear ion trap parameters for the detection and quantitation of glucose. In: *Rapid communications in mass spectrometry* 26 (3), S. 385–391.

Safi, J. M.; Abou-Foul, N. S.; El-Nahhal, Y. Z.; El-Sebae, A. H. (2002): Monitoring of pesticide residues on cucumber, tomatoes and strawberries in Gaza Governorates, Palestine. In: *Food/Nahrung* 46 (1), S. 34–39.

Sammarco, Giuseppe; Bardin, Daniele; Quaini, Federica; Dall'Asta, Chiara; Christmann, Joscha; Weller, Philipp; Suman, Michele (2023a): A geographical origin assessment of Italian hazelnuts: Gas chromatography-ion mobility spectrometry coupled with multivariate statistical analysis and data fusion approach. In: *Food Research International* 171, S. 113085.

Sammarco, Giuseppe; Dall'Asta, Chiara; Suman, Michele (2023b): Near infrared spectroscopy and multivariate statistical analysis as rapid tools for the geographical origin assessment of Italian hazelnuts. In: *Vibrational Spectroscopy* 126, S. 103531.

Schmauder, Felix; Schütz, David; Creydt, Marina; Riedl, Janet; Fischer, Markus (2024): Food authentication goes green: Determination of the geographical origin of grain maize (Zea mays L.) using direct analysis in real time mass spectrometry (DART-MS). In: *Food Control* 163, S. 110489.

Schütz, David; Achten, Elisabeth; Creydt, Marina; Riedl, Janet; Fischer, Markus (2021): Non-targeted LC-MS metabolomics approach towards an authentication of the geographical origin of grain maize (Zea mays L.) samples. In: *Foods* 10 (9), S. 2160.

Shakiba, Navid; Gerdes, Annika; Holz, Nathalie; Wenck, Soeren; Bachmann, René; Schneider, Tobias et al. (2022): Determination of the geographical origin of hazelnuts (Corylus avellana L.) by Near-Infrared spectroscopy (NIR) and a Low-Level Fusion with nuclear magnetic resonance (NMR). In: *Microchemical Journal* 174, S. 107066.

Shen, Shigang; Xia, Liya; Xiong, Na; Liu, Zhenghao; Sun, Hanwen (2013): Determination of the geographic origin of rice by element fingerprints and correlation analyses with the soil of origin. In: *Analytical Methods* 5 (21), S. 6177–6185.

Sisco, Edward; Forbes, Thomas P. (2021): Forensic applications of DART-MS: A review of recent literature. In: *Forensic Chemistry* 22, S. 100294.

Smith, Lincoln G. (1951): Magnetic electron multipliers for detection of positive ions. In: *Review of Scientific Instruments* 22 (3), S. 166–170.

Song, Liguo; Dykstra, Andrew B.; Yao, Huifang; Bartmess, John E. (2011): Ionization mechanism of negative ion-direct analysis in real time: a comparative study with negative ion-atmospheric pressure photoionization. In: *Journal of the American Society for Mass Spectrometry* 20 (1), S. 42–50.

Spargel & Erdbeerprofi (2024): Bilanz zur Erdbeersaison 2024 in Süddeutschland. Online verfügbar unter https://www.spargel-erdbeerprofi.de/nachricht-spargelerdbeer/bilanz-zur-erdbeersaison-2024-in-sueddeutschland, zuletzt aktualisiert am 25.06.2024, zuletzt geprüft am 24.01.2025.

Speiser, Jaime Lynn; Miller, Michael E.; Tooze, Janet; Ip, Edward (2019): A comparison of random forest variable selection methods for classification prediction modeling. In: *Expert systems with applications* 134, S. 93–101.

Spink, John; Moyer, Douglas C. (2011): Defining the public health threat of food fraud. In: *Journal of food science* 76 (9), R157-R163.

Stanton, Henry E.; Chupka, William A.; Inghram, Mark G. (1956): Electron multipliers in mass spectrometry; effect of molecular structure. In: *Review of Scientific Instruments* 27 (2), S. 109.

Statistisches Bundesamt (2024): Import- und Exportmenge von Erdbeeren in Deutschland in den Jahren 2006 bis 2023 (in Tonnen). Statistisches Bundesamt. Online verfügbar unter https://www-genesis.destatis.de/datenbank/online/statistic/51000/details/search/s/ZXJkYmVlcmVu, zuletzt aktualisiert am 14.02.2024, zuletzt geprüft am 24.01.2025.

Statistisches Bundesamt (2025a): Index der Einfurpreise für Haselnüsse, Deutschland. Online verfügbar unter https://www-genesis.destatis.de/datenbank/online/statistic/61411/details/search/s/aGFzZWxuJUMzJUJDc3Nl, zuletzt aktualisiert am 08.01.2025, zuletzt geprüft am 23.01.2025.

Statistisches Bundesamt (2025b): Betriebe mit Erdbeeranbau, Anbaufläche, Erntemenge: Deutschland, Jahre, Anbauform. Statistisches Bundesamt. Online verfügbar unter https://www-genesis.destatis.de/datenbank/online/statistic/41215/table/41215-0007/search/s/ZXJkYmVlcmVuJTIwZXJudGVtZW5nZQ==, zuletzt aktualisiert am 09.01.2025, zuletzt geprüft am 24.01.2025.

Sud, Manish; Fahy, Eoin; Cotter, Dawn; Brown, Alex; Dennis, Edward A.; Glass, Christopher K. et al. (2007): Lmsd: Lipid maps structure database. In: *Nucleic acids research* 35 (suppl_1), D527-D532.

United States Department of Agriculture (2025): Grain: World Markets and Trade. United States Department of Agriculture. Online verfügbar unter https://apps.fas.usda.gov/psdonline/circulars/grain.pdf, zuletzt aktualisiert am 01.01.2025, zuletzt geprüft am 24.01.2025.

Upchurch, Robert G. (2008): Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. In: *Biotechnology letters* 30, S. 967–977.

Vaclavik, Lukas; Cajka, Tomas; Hrbek, Vojtech; Hajslova, Jana (2009): Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. In: *Analytica Chimica Acta* 645 (1-2), S. 56–63.

van Berkel, Gary J.; Kertesz, Vilmos (2007): Using the electrochemistry of the electrospray ion source. In: *Analytical chemistry* 79 (15).

Vásquez, Héctor V.; Vigo, Carmen N.; Saravia, David; Valqui, Leandro; Bobadilla, Leidy G. (2024): Agronomic performance of forage corn for cattle feeding in Amazonas, Peru. In: *Heliyon* 10 (9).

Verordnung (EG) 1308/2013: Verordnung (EG) 1308/2013 des Europäischen Parlaments und des Rates vom 17. Dezember 2013 über eine gemeinsame Marktorganisation für

landwirtschaftliche Erzeugnisse und zur Aufhebung der Verordnungen (EWG) Nr. 922/72, (EWG) Nr. 234/79, (EG) Nr. 1037/2001 und (EG) Nr. 1234/2007.

Verordnung (EG) Nr. 396/2005: VERORDNUNG (EG) Nr. 396/2005 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 23. Februar 2005 über Höchstgehalte an Pestizidrückständen in oder auf Lebens- und Futtermitteln pflanzlichen und tierischen Ursprungs und zur Änderung der Richtlinie 91/414/EWG des Rates.

Verordnung (EG) Nr. 510/2006: Verordnung (EG) Nr. 510/2006 des Rates vom 20. März 2006 zum Schutz von geografischen Angaben und Ursprungsbezeichnungen für Agrarerzeugnisse und Lebensmittel, zuletzt geändert durch Verordnung (EG) Nr. 417/2008 der Kommission vom 8. Mai 2008 200.

Vogeser, Michael; Kirchhoff, Fabian (2011): Progress in automation of LC-MS in laboratory medicine. In: *Clinical biochemistry* 44 (1), S. 4–13.

Wang, Fuyun; Lin, Hao; Xu, Peiting; Bi, Xiakun; Sun, Li (2021): Egg freshness evaluation using transmission and reflection of NIR spectroscopy coupled multivariate analysis. In: *Foods* 10 (9), S. 2176.

Wang, Lu; Sun, Da-Wen; Pu, Hongbin; Cheng, Jun-Hu (2017): Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. In: *Critical reviews in food science and nutrition* 57 (7), S. 1524–1538.

Wang, Pei; Yu, Zhiguo (2015): Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. In: *Journal of pharmaceutical analysis* 5 (5), S. 277–284.

Wang, Yan; Sun, Jianghui; Qiao, Jinping; Ouyang, Jin; Na, N. A. (2018): A “soft” and “hard” ionization method for comprehensive studies of molecules. In: *Analytical chemistry* 90 (24), S. 14095–14099.

Wang, Yang; Liu, Shuying (2024): Recent application of direct analysis in real time mass spectrometry in plant materials analysis with emphasis on traditional Chinese herbal medicine. In: *Mass Spectrometry Reviews* 43 (5), S. 1150–1171.

Wani, Idrees Ahmed; Ayoub, Aneesa; Bhat, Naseer Ahmad; Dar, Aamir Hussain; Gull, Amir (2020): Hazelnut. In: *Antioxidants in Vegetables and Nuts-Properties and Health Benefits*, S. 559–572.

Weidlein, Johann; Müller, Ulrich; Dehnicke, Kurt (1982): Schwingungsspektroskopie: Eine Einführung: Thieme.

Williams, Brianna; Hewage, Shamika PWR; Alexander, Denzel; Fernando, Harshica (2025): 1H-NMR Lipidomics, Comparing Fatty Acids and Lipids in Cow, Goat, Almond,

Cashew, Soy, and Coconut Milk Using NMR and Mass Spectrometry. In: *Metabolites* 15 (2), S. 110.

Wishart, David S.; Guo, AnChi; Oler, Eponine; Wang, Fei; Anjum, Afia; Peters, Harrison et al. (2022): HMDB 5.0: the human metabolome database for 2022. In: *Nucleic acids research* 50 (D1), D622-D631.

Wolff, M. M.; Stephens, W. E. (1953): A pulsed mass spectrometer with time dispersion. In: *Review of Scientific Instruments* 24 (8), S. 616–617.

Worley, Bradley; Powers, Robert (2013): Multivariate analysis in metabolomics. In: *Current metabolomics* 1 (1), S. 92–107.

Wu, Tian; Zou, Rui; Pu, Dian; Lan, Zengquan; Zhao, Bingyu (2021): Non-targeted and targeted metabolomics profiling of tea plants (Camellia sinensis) in response to its intercropping with Chinese chestnut. In: *BMC plant biology* 21 (1), S. 55.

Yamashita, Masamichi; Fenn, John B. (1984): Electrospray ion source. Another variation on the free-jet theme. In: *The Journal of physical chemistry* 88 (20), S. 4451–4459.

Yeni, Filiz; Yavaş, Sinem; Alpas, HAMI; Soyer, YESIM (2016): Most common foodborne pathogens and mycotoxins on fresh produce: a review of recent outbreaks. In: *Critical reviews in food science and nutrition* 56 (9), S. 1532–1544.

Zacometti, Carmela; Lante, Anna; Cisneros, Miluska; Massaro, Andrea; Mihaylova, Dasha; Chalova, Vesela et al. (2024): Rapid Assessment of Metabolomic Fingerprinting of Recycled Sunflower By-Products via DART-HRMS. In: *Molecules* 29 (17), S. 4092.

Zhao, Shuping; Zhang, Bob; Yang, Jian; Zhou, Jianhang; Xu, Yong (2024): Linear discriminant analysis. In: *Nature Reviews Methods Primers* 4 (1), S. 70.

Zhong, Liang; Huang, Ruiqi; Gao, Lele; Yue, Jianan; Zhao, Bing; Nie, Lei et al. (2023): A Novel Variable Selection Method Based on Binning-Normalized Mutual Information for Multivariate Calibration. In: *Molecules* 28 (15), S. 5672.

9 Anhang

9.1 Supporting Information zu: DART-MS approaches for metabolomics-based authentication: Challenges in determining the geographical origin of strawberries (*Fragaria* × *ananassa*)

Electronic Supplementary Information

DART-MS Approaches for Metabolomics-Based Authentication: Challenges in Determining the Geographical Origin of Strawberries (*Fragaria* × *ananassa*)

Felix Schmauder, Johannes Brockelt, Kim Brettschneider, Stephan Seifert, Marina Creydt Markus Fischer*

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

*Corresponding author:

Markus Fischer, Hamburg School of Food Science - Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

Tel.: +49 40 42838-4359, Fax: +49 40 42838-4342

Email: markus.fischer@uni-hamburg.de

Table S1: Dataset 1 consisting of 62 samples with information on origin, variety and year of cultivation, if available.

Number	Origin	State	Year	Variety
1	Greece	Westgriechenland	2022	Not specified
2	Spain	Huelva	2022	Not specified
3	Spain	Huelva	2022	Fortuna
4	Spain	Valencia	2022	Fortuna
5	Spain	Huelva	2022	Not specified
6	Spain	Huelva	2022	Not specified
7	Germany	Baden-Württemberg	2022	Asia
8	Germany	Baden-Württemberg	2022	Not specified
9	Germany	Baden-Württemberg	2022	Not specified
10	Germany	Baden-Württemberg	2022	Clery
11	Germany	Baden-Württemberg	2022	Sibilla
12	Germany	Niedersachsen	2022	Faith
13	Germany	Niedersachsen	2022	Magnus
14	Germany	Niedersachsen	2022	Malwina
15	Germany	Niedersachsen	2022	Faith
16	Germany	Niedersachsen	2022	Sonata
17	Greece	Westgriechenland	2022	Not specified
18	Greece	Not specified	2022	Not specified
19	Greece	Westgriechenland	2023	Not specified
20	Spain	Huelva	2023	Not specified
21	Spain	Andalusien	2023	Fortuna
22	Spain	Sevilla	2023	Not specified
23	Spain	Huelva	2023	Not specified
24	Spain	Huelva	2023	Not specified
25	Spain	Huelva	2023	Not specified
26	Greece	Westgriechenland	2023	Not specified
27	Spain	Huelva	2023	Not specified
28	Spain	Huelva	2023	Not specified
29	Spain	Andalusien	2023	Not specified
30	Egypt	Not specified	2023	Not specified
31	Spain	Huelva	2023	Felicity
32	Greece	Westgriechenland	2023	Not specified
33	Egypt	Not specified	2023	Not specified
34	Egypt	Not specified	2023	Not specified
35	Greece	Not specified	2023	Not specified
36	Greece	Westgriechenland	2023	Not specified
37	Egypt	Not specified	2023	Not specified
38	Egypt	Not specified	2023	Not specified
39	Egypt	Not specified	2023	Fortuna
40	Greece	Westgriechenland	2023	Not specified
41	Egypt	Not specified	2023	Fortuna
42	Greece	Westgriechenland	2023	Not specified
43	Greece	Westgriechenland	2023	Not specified
44	Greece	Zentralmakkedonien	2023	Victory
45	Greece	Westgriechenland	2023	Calderon
46	Greece	Westgriechenland	2023	Not specified
47	Greece	Westgriechenland	2023	Not specified
48	Egypt	Not specified	2023	Not specified
49	Egypt	Not specified	2023	Not specified
50	Egypt	Not specified	2023	Not specified
51	Egypt	Not specified	2023	Not specified
52	Egypt	Not specified	2023	Not specified

Number	Origin	State	Year	Variety
53	Germany	Not specified	2023	Not specified
54	Germany	Not specified	2023	Not specified
55	Germany	Not specified	2023	Not specified
56	Germany	Not specified	2023	Not specified
57	Germany	Not specified	2023	Not specified
58	Germany	Niedersachsen	2023	Lady Emma
59	Germany	Niedersachsen	2023	Sonsation
60	Germany	Niedersachsen	2023	Not specified
61	Germany	Niedersachsen	2023	Not specified
62	Germany	Niedersachsen	2023	Not specified

Table S2: Dataset 2 consisting of 200 samples with with information on origin, variety and year of cultivation, if available.

Number	Origin	State	Year	Variety
1	Germany	Niedersachsen	2021	Malwina
2	Germany	Niedersachsen	2021	Not specified
3	Germany	Schleswig-Holstein	2021	Malwina
4	Germany	Niedersachsen	2021	Not specified
5	Germany	Niedersachsen	2021	Malwina
6	Germany	Niedersachsen	2021	Not specified
7	Germany	Baden-Württemberg	2021	Not specified
8	Germany	Baden-Württemberg	2021	Not specified
9	Germany	Baden-Württemberg	2021	Malwina
10	Germany	Baden-Württemberg	2021	Magnus
11	Germany	Baden-Württemberg	2021	Not specified
12	Germany	Schleswig-Holstein	2021	Not specified
13	Germany	Niedersachsen	2021	Not specified
14	Spain	Huelva	2022	Not specified
15	Spain	Huelva	2022	Not specified
16	Spain	Huelva	2022	Not specified
17	Spain	Huelva	2022	Not specified
18	Spain	Huelva	2022	Not specified
19	Spain	Valencia	2022	Not specified
20	Spain	Valencia	2022	Not specified
21	Greece	Westgriechenland	2022	Not specified
22	Spain	Huelva	2022	Not specified
23	Spain	Huelva	2022	Fortuna
24	Spain	Valencia	2022	Fortuna
25	Spain	Huelva	2022	Not specified
26	Spain	Huelva	2022	Not specified
27	Spain	Valencia	2022	Not specified
28	Spain	Huelva	2022	Not specified
29	Netherlands	Not specified	2022	Not specified
30	Netherlands	Holland	2022	Not specified
31	Spain	Huelva	2022	Not specified
32	Spain	Not specified	2022	Not specified
33	Spain	Huelva	2022	Not specified
34	Netherlands	Not specified	2022	Not specified
36	Germany	Rheinland-Pfalz	2022	Clery
37	Germany	Rheinland-Pfalz	2022	Clery
38	Germany	Rheinland-Pfalz	2022	Malling Centenary
42	Germany	Rheinland-Pfalz	2022	Clery
49	Germany	Hessen	2022	Clery

Number	Origin	State	Year	Variety
50	Germany	Hessen	2022	Flair
51	Germany	Hessen	2022	Malling Centenary
52	Germany	Hessen	2022	Clery
53	Germany	Hessen	2022	Lady Emma
54	Germany	Hessen	2022	Clery
55	Germany	Hessen	2022	Not specified
56	Germany	Nordrhein-Westfalen	2022	Magnum
57	Germany	Nordrhein-Westfalen	2022	Flair
58	Germany	Nordrhein-Westfalen	2022	Alegro
61	Germany	Nordrhein-Westfalen	2022	Lambada
63	Germany	Nordrhein-Westfalen	2022	Malling Centenary
65	Germany	Nordrhein-Westfalen	2022	Clery
67	Germany	Nordrhein-Westfalen	2022	Glorielle
69	Germany	Nordrhein-Westfalen	2022	Hademar
70	Germany	Nordrhein-Westfalen	2022	Florentina
71	Germany	Nordrhein-Westfalen	2022	Magnum
73	Germany	Bayern	2022	Clery
74	Germany	Bayern	2022	Salsa
94	Germany	Baden-Württemberg	2022	Sibilla
95	Germany	Baden-Württemberg	2022	Clery
103	Germany	Bayern	2022	Clery
112	Germany	Bayern	2022	Alegro
116	Germany	Bayern	2022	Asia
120	Germany	Bayern	2022	Not specified
124	Germany	Sachsen	2022	Not specified
125	Germany	Sachsen	2022	Verdi
126	Germany	Sachsen	2022	Asia
127	Germany	Sachsen	2022	Asia
129	Germany	Sachsen	2022	Alegro
130	Germany	Sachsen	2022	Not specified
131	Germany	Brandenburg	2022	Not specified
132	Germany	Brandenburg	2022	Not specified
135	Germany	Sachsen-Anhalt	2022	Florentina
136	Germany	Sachsen-Anhalt	2022	Not specified
143	Germany	Brandenburg	2022	Asia
149	Germany	Schleswig-Holstein	2022	Magnum
150	Germany	Schleswig-Holstein	2022	Sonsation
153	Germany	Schleswig-Holstein	2022	Korona
154	Germany	Schleswig-Holstein	2022	Malling Allure
155	Germany	Schleswig-Holstein	2022	Florence
159	Germany	Schleswig-Holstein	2022	Magnus
160	Germany	Schleswig-Holstein	2022	Polka
163	Germany	Schleswig-Holstein	2022	Korona
164	Germany	Hamburg	2022	Malvina
165	Germany	Hamburg	2022	Not specified

Number	Origin	State	Year	Variety
166	Germany	Niedersachsen	2022	Polka
170	Germany	Niedersachsen	2022	Magnus
172	Germany	Niedersachsen	2022	Faith
177	Germany	Niedersachsen	2022	Malvina
182	Netherlands	Not specified	2022	Not specified
183	Greece	Nea Manolada	2022	Not specified
184	Greece	Not specified	2022	Not specified
185	Greece	Nea Manolada	2023	Not specified
186	Spain	Huelva	2023	Not specified
187	Spain	Andalusien	2023	Fortuna
188	Spain	Sevilla	2023	Not specified
189	Spain	Huelva	2023	Not specified
190	Spain	Huelva	2023	Not specified
191	Spain	Huelva	2023	Not specified
192	Greece	Westgriechenland	2023	Not specified
193	Spain	Huelva	2023	Not specified
194	Spain	Huelva	2023	Not specified
195	Spain	Andalusien	2023	Not specified
196	Egypt	Not specified	2023	Not specified
197	Spain	Huelva	2023	Felicity
198	Greece	Westgriechenland	2023	Not specified
199	Egypt	Not specified	2023	Not specified
200	Spain	Huelva	2023	Not specified
201	Egypt	Not specified	2023	Not specified
202	Greece	Not specified	2023	Not specified
203	Greece	Westgriechenland	2023	Not specified
204	Spain	Andalusien	2023	Not specified
205	Egypt	Not specified	2023	Not specified
206	Spain	Huelva	2023	Not specified
207	Egypt	Not specified	2023	Not specified
208	Spain	Huelva	2023	Not specified
209	Egypt	Not specified	2023	Fortuna
210	Greece	Westgriechenland	2023	Not specified
211	Spain	Huelva	2023	Victory
212	Spain	Not specified	2023	Inspire
213	Marokko	Not specified	2023	Fortuna
214	Marokko	Not specified	2023	Fortuna
215	Marokko	Not specified	2023	Not specified
216	Egypt	Not specified	2023	Fortuna
217	Spain	Huelva	2023	Not specified
218	Greece	Westgriechenland	2023	Not specified
219	Marokko	Not specified	2023	Not specified
220	Italien	Not specified	2023	Victory
221	Marokko	Not specified	2023	Marisol
222	Marokko	Not specified	2023	Fortuna
223	Greece	Westgriechenland	2023	Not specified
224	Spain	Andalusien	2023	Rociera
225	Spain	Andalusien	2023	Calinda
226	Spain	Andalusien	2023	Not specified
227	Spain	Andalusien	2023	Not specified
228	Greece	Zentralmakkedonien	2023	Victory
229	Greece	Westgriechenland	2023	Calderon
230	Greece	Westgriechenland	2023	Not specified
231	Greece	Westgriechenland	2023	Not specified
232	Greece	Westgriechenland	2023	Not specified

Number	Origin	State	Year	Variety
233	Greece	Westgriechenland	2023	Not specified
234	Greece	Westgriechenland	2023	Not specified
235	Marokko	Not specified	2023	Not specified
236	Spain	Andalusien	2023	Not specified
237	Spain	Andalusien	2023	Calinda
238	Greece	Westgriechenland	2023	Not specified
239	Marokko	Not specified	2023	Not specified
240	Spain	Andalusien	2023	Not specified
241	Greece	Westgriechenland	2023	Not specified
242	Netherlands	Gelderland	2023	Malling Centenary
243	Netherlands	Limburg	2023	Elsanta
244	Netherlands	Gelderland	2023	Malling Centenary
245	Netherlands	Not specified	2023	Sonata
246	Italien	Verona	2023	Not specified
247	Spain	Andalusien	2023	Not specified
248	Italien	Verona	2023	Sabrosa
249	Greece	Schleswig-Holstein	2023	Not specified
250	Egypt	Not specified	2023	Not specified
251	Egypt	Not specified	2023	Not specified
252	Egypt	Not specified	2023	Not specified
253	Egypt	Not specified	2023	Not specified
254	Egypt	Not specified	2023	Not specified
255	Netherlands	Limburg	2023	Not specified
257	Germany	Rheinland-Pfalz	2023	Malvina
262	Germany	Rheinland-Pfalz	2023	Clery
265	Germany	Hessen	2023	Not specified
266	Germany	Hessen	2023	Faith
267	Germany	Hessen	2023	Alegro
268	Germany	Hessen	2023	Sonata
269	Germany	Hessen	2023	Verdi
272	Germany	Nordrhein-Westfalen	2023	Sonsation
276	Germany	Nordrhein-Westfalen	2023	Florentina
277	Germany	Nordrhein-Westfalen	2023	Not specified
279	Germany	Nordrhein-Westfalen	2023	Not specified
282	Germany	Thüringen	2023	Kimberly
285	Germany	Thüringen	2023	Not specified
286	Germany	Thüringen	2023	Not specified
287	Germany	Sachsen	2023	Asia
289	Germany	Sachsen	2023	Asia
290	Germany	Sachsen	2023	Rendezvous
291	Germany	Sachsen	2023	Daragal
293	Germany	Brandenburg	2023	Clairy
294	Germany	Brandenburg	2023	Lucia
295	Germany	Brandenburg	2023	Glorielle
296	Germany	Brandenburg	2023	Flair
297	Germany	Brandenburg	2023	Clery
298	Germany	Brandenburg	2023	Jenkka
299	Germany	Brandenburg	2023	Glorielle
301	Germany	Sachsen-Anhalt	2023	Not specified
302	Germany	Sachsen-Anhalt	2023	Asia
305	Germany	Baden-Württemberg	2023	Malvina

Number	Origin	State	Year	Variety
306	Germany	Baden-Württemberg	2023	Clairy
312	Germany	Baden-Württemberg	2023	Verdi
313	Germany	Baden-Württemberg	2023	Thear
314	Germany	Baden-Württemberg	2023	Sonata
319	Germany	Bayern	2023	Verdi
323	Germany	Bayern	2023	Salsa
330	Germany	Niedersachsen	2023	Polka
336	Germany	Hamburg	2023	Florance
338	Germany	Schleswig-Holstein	2023	Sonata
339	Germany	Schleswig-Holstein	2023	Verdi

Table S3: Dataset 3 consisting of 207 samples with information on origin, variety and year of cultivation, if available.

Number	Origin	State	Year	Variety
21	Greece	Westgriechenland	2022	Not specified
22	Spain	Huelva	2022	Not specified
23	Spain	Huelva	2022	Fortuna
24	Spain	Valencia	2022	Fortuna
25	Spain	Huelva	2022	Not specified
26	Spain	Huelva	2022	Not specified
27	Spain	Valencia	2022	Not specified
28	Spain	Huelva	2022	Not specified
29	Netherlands	Not specified	2022	Not specified
30	Netherlands	Holland	2022	Not specified
31	Spain	Huelva	2022	Not specified
32	Spain	Not specified	2022	Not specified
33	Spain	Huelva	2022	Not specified
34	Netherlands	Not specified	2022	Not specified
48	Germany	Hessen	2022	Not specified
49	Germany	Hessen	2022	Clery
53	Germany	Hessen	2022	Lady Emma
54	Germany	Hessen	2022	Clery
55	Germany	Hessen	2022	Not specified
84	Germany	Baden-Württemberg	2022	Aprica
85	Germany	Baden-Württemberg	2022	Aprica
86	Germany	Baden-Württemberg	2022	Clery
87	Germany	Baden-Württemberg	2022	Clery
88	Germany	Baden-Württemberg	2022	Asia
89	Germany	Baden-Württemberg	2022	Clery
90	Germany	Baden-Württemberg	2022	Sonsation
94	Germany	Baden-Württemberg	2022	Sibilla
95	Germany	Baden-Württemberg	2022	Clery
96	Germany	Baden-Württemberg	2022	Asia
97	Germany	Baden-Württemberg	2022	Not specified
98	Germany	Baden-Württemberg	2022	Not specified
112	Germany	Bayern	2022	Alegro
113	Germany	Bayern	2022	Not specified
114	Germany	Bayern	2022	Not specified
115	Germany	Bayern	2022	Rumba
116	Germany	Bayern	2022	Asia
117	Germany	Bayern	2022	Clery
118	Germany	Bayern	2022	Limalexia

Number	Origin	State	Year	Variety
119	Germany	Bayern	2022	Sonata
121	Germany	Thüringen	2022	Rumba
122	Germany	Thüringen	2022	Kimberly
123	Germany	Thüringen	2022	Not specified
171	Germany	Niedersachsen	2022	Malwina
172	Germany	Niedersachsen	2022	Faith
173	Germany	Niedersachsen	2022	Faith
174	Germany	Niedersachsen	2022	Sonata
175	Germany	Niedersachsen	2022	Faith
182	Netherlands	Not specified	2022	Not specified
183	Greece	Nea Manolada	2022	Not specified
184	Greece	Not specified	2022	Not specified
185	Greece	Nea Manolada	2023	Not specified
190	Spain	Huelva	2023	Not specified
191	Spain	Huelva	2023	Not specified
192	Greece	Westgriechenland	2023	Not specified
193	Spain	Huelva	2023	Not specified
194	Spain	Huelva	2023	Not specified
195	Spain	Andalusien	2023	Not specified
196	Egypt	Not specified	2023	Not specified
197	Spain	Huelva	2023	Felicity
198	Greece	Westgriechenland	2023	Not specified
199	Egypt	Not specified	2023	Not specified
200	Spain	Huelva	2023	Not specified
201	Egypt	Not specified	2023	Not specified
203	Greece	Westgriechenland	2023	Not specified
204	Spain	Andalusien	2023	Not specified
205	Egypt	Not specified	2023	Not specified
206	Spain	Huelva	2023	Not specified
207	Egypt	Not specified	2023	Not specified
208	Spain	Huelva	2023	Not specified
209	Egypt	Not specified	2023	Fortuna
210	Greece	Westgriechenland	2023	Not specified
211	Spain	Huelva	2023	Victory
216	Egypt	Not specified	2023	Fortuna
218	Greece	Westgriechenland	2023	Not specified
223	Greece	Westgriechenland	2023	Not specified
224	Spain	Andalusien	2023	Rociera
225	Spain	Andalusien	2023	Calinda
227	Spain	Andalusien	2023	Not specified
228	Greece	ZentralmakkedoNiedersachsenen	2023	Victory
229	Greece	Westgriechenland	2023	Calderon
230	Greece	Westgriechenland	2023	Not specified
231	Greece	Westgriechenland	2023	Not specified
232	Greece	Westgriechenland	2023	Not specified
233	Greece	Westgriechenland	2023	Not specified
234	Greece	Westgriechenland	2023	Not specified
236	Spain	Andalusien	2023	Not specified
237	Spain	Andalusien	2023	Calinda
238	Greece	Westgriechenland	2023	Not specified
240	Spain	Andalusien	2023	Not specified
241	Greece	Westgriechenland	2023	Not specified
242	Netherlands	Gelderland	2023	Malling Centenary
243	Netherlands	Limburg	2023	Elsanta

Number	Origin	State	Year	Variety
244	Netherlands	Gelderland	2023	Malling Centenary
245	Netherlands	Not specified	2023	Sonata
247	Spain	Andalusien	2023	Not specified
249	Greece	Westgriechenland	2023	Not specified
250	Egypt	Not specified	2023	Not specified
251	Egypt	Not specified	2023	Not specified
252	Egypt	Not specified	2023	Not specified
253	Egypt	Not specified	2023	Not specified
254	Egypt	Not specified	2023	Not specified
255	Netherlands	Limburg	2023	Not specified
265	Germany	Hessen	2023	Not specified
266	Germany	Hessen	2023	Faith
269	Germany	Hessen	2023	Not specified
312	Germany	Baden-Württemberg	2023	Verdi
313	Germany	Baden-Württemberg	2023	Thear
314	Germany	Baden-Württemberg	2023	Sonata
315	Germany	Baden-Württemberg	2023	Sonsation
322	Germany	Bayern	2023	Lambada
323	Germany	Bayern	2023	Salsa
333	Germany	Niedersachsen	2023	Not specified
334	Germany	Niedersachsen	2023	Not specified
335	Germany	Niedersachsen	2023	Not specified
341	Netherlands	Limburg	2023	Not specified
342	Netherlands	Gelderland	2023	Elsanta
343	Netherlands	Limburg	2023	Opera
344	Netherlands	Not specified	2023	Not specified
345	Greece	Westgriechenland	2023	Not specified
346	Egypt	Not specified	2023	Not specified
347	Egypt	Not specified	2023	Not specified
348	Egypt	Not specified	2023	Not specified
349	Egypt	Not specified	2023	Not specified
350	Egypt	Not specified	2023	Not specified
351	Egypt	Not specified	2023	Not specified
352	Egypt	Not specified	2023	Not specified
353	Egypt	Not specified	2023	Not specified
354	Egypt	Not specified	2023	Not specified
355	Polen	Not specified	2023	Not specified
356	Polen	Not specified	2023	Not specified
357	Polen	Not specified	2023	Not specified
358	Polen	Not specified	2023	Not specified
359	Polen	Not specified	2023	Not specified
360	Polen	Not specified	2023	Not specified
361	Polen	Not specified	2023	Not specified
362	Spain	Huelva	2024	Not specified
363	Spain	Huelva	2024	Not specified
364	Spain	Huelva	2024	Felicity
365	Spain	Huelva	2024	Not specified
366	Netherlands	Not specified	2024	Sonsation
367	Spain	Huelva	2024	Not specified
368	Greece	Not specified	2024	Fortuna
369	Spain	Huelva	2024	Rabida
370	Netherlands	Not specified	2024	Inspire
372	Egypt	Not specified	2024	Not specified
373	Greece	Westgriechenland	2024	Not specified

Number	Origin	State	Year	Variety
374	Spain	Huelva	2024	Not specified
375	Greece	Westgriechenland	2024	Not specified
377	Egypt	Not specified	2023	Not specified
378	Egypt	Not specified	2023	Not specified
379	Egypt	Not specified	2023	Not specified
380	Egypt	Not specified	2023	Not specified
381	Egypt	Not specified	2023	Not specified
382	Egypt	Not specified	2023	Not specified
383	Egypt	Not specified	2023	Not specified
384	Egypt	Not specified	2023	Not specified
385	Egypt	Not specified	2023	Not specified
386	Egypt	Not specified	2023	Not specified
387	Egypt	Not specified	2023	Not specified
388	Polen	Not specified	2023	Not specified
389	Polen	Not specified	2023	Not specified
390	Polen	Not specified	2023	Not specified
391	Polen	Not specified	2023	Not specified
392	Polen	Not specified	2023	Not specified
393	Greece	Not specified	2024	Not specified
394	Greece	Not specified	2024	Not specified
395	Greece	Westgriechenland	2024	Not specified
397	Netherlands	Limburg	2024	Elsanta
398	Netherlands	Gelderland	2024	Lady Emma
400	Netherlands	Not specified	2024	Not specified
401	Spain	Not specified	2024	Not specified
407	Germany	Baden-Württemberg	2024	Not specified
410	Polen	Not specified	2024	Not specified
411	Polen	Not specified	2024	Not specified
412	Polen	Not specified	2024	Not specified
413	Polen	Not specified	2024	Not specified
414	Polen	Not specified	2024	Not specified
415	Polen	Not specified	2024	Hademar
417	Polen	Not specified	2024	Not specified
419	Polen	Not specified	2024	Hademar
420	Polen	Not specified	2024	Not specified
421	Polen	Not specified	2024	Not specified
422	Polen	Not specified	2024	Hademar
423	Polen	Not specified	2024	Flair
424	Polen	Not specified	2024	Not specified
425	Germany	Hessen	2024	Wima Zanta
426	Germany	Hessen	2024	Sonsation
427	Germany	Hessen	2024	Clery
428	Germany	Baden-Württemberg	2024	Asia
435	Netherlands	Not specified	2024	Not specified
436	Netherlands	Not specified	2024	Not specified
437	Netherlands	Not specified	2024	Not specified
438	Netherlands	Not specified	2024	Not specified
439	Netherlands	Not specified	2024	Not specified
440	Netherlands	Not specified	2024	Not specified
441	Netherlands	Not specified	2024	Not specified
442	Netherlands	Not specified	2024	Not specified
443	Netherlands	Not specified	2024	Not specified
444	Netherlands	Not specified	2024	Not specified
445	Netherlands	Not specified	2024	Not specified
446	Netherlands	Not specified	2024	Not specified

Number	Origin	State	Year	Variety
447	Netherlands	Not specified	2024	Not specified
448	Netherlands	Not specified	2024	Not specified
454	Germany	Niedersachsen	2024	Not specified
455	Polen	Not specified	2024	Not specified
456	Polen	Not specified	2024	Not specified
457	Polen	Not specified	2024	Hademar

Table S4: Percentage accuracies of the random forest classification of dataset 2 for the geographical origin determination of German and non-German strawberry samples using the Quick-strip and Dip-it approaches. In addition, the number of samples (#) used is provided.

			Predicted Class [%]		
			Germany	Other Origin	#
True Class [%]	Quick-Strip approach	Germany	94.3	5.7	105
		Other Origin	22.1	77.9	95
	Dip-it approach	Germany	95.2	4.8	105
		Other Origin	22.1	77.9	95

Table S5: Identified signals to determine their geographical origin of strawberries. The signals were identified using MS/MS spectra and the DART-qToF platform.

Tentative compound	Proposed formula	Adduct	Theoretical *m/z*	Measured *m/z*	Delta *m/z*	Relevant Fragments	Selected by Boruta
TG 54:5	$C_{57}H_{100}O_6$	$[M+NH_4]^+$	898.7858	898.7850	0.0008	599.5043; 601.5196	
TG 54:6	$C_{57}H_{98}O_6$	$[M+NH_4]^+$	896.7702	896.7712	0.0010	597.4894; 599.5053; 601.5197	
TG 54:7	$C_{57}H_{96}O_6$	$[M+NH_4]^+$	894.7545	894.7564	0.0019	597.4895; 599.5050	
TG 54:8	$C_{57}H_{94}O_6$	$[M+NH_4]^+$	892.7389	892.7067	0.0016	595.4739; 597.4895	
TG 52:3	$C_{55}H_{100}O_6$	$[M+NH_4]^+$	874.7858	874.7861	0.0003	575.5042; 577.5183; 601.5196	
TG 52:4	$C_{55}H_{98}O_6$	$[M+NH_4]^+$	872.7702	872.7702	0.0014	575.5046; 599.5042	
TG 44:3	$C_{47}H_{84}O_6$	$[M+NH_4]^+$	762.6606	762.6590	0.0016	201.1496; 465.3945	
TG 44:4	$C_{47}H_{82}O_6$	$[M+NH_4]^+$	760.6450	760.6469	0.0019	201.1484; 465.3950	
TG 44:5	$C_{47}H_{80}O_6$	$[M+NH_4]^+$	758.6293	758.6309	0.0016	201.1473; 465.3804	
TG 44:6	$C_{47}H_{78}O_6$	$[M+NH_4]^+$	756.6137	756.6168	0.0031	201.1501; 461.3654	
TG 42:2	$C_{45}H_{82}O_6$	$[M+NH_4]^+$	736.6450	736.6453	0.0003	201.1477; 439.3804; 463.3804	
DG 36:2	$C_{39}H_{72}O_5$	$[M+NH_4]^+$	638.5718	638.5732	0.0014	337.2758; 339.2911; 341.3069	X
DG 36:3	$C_{39}H_{70}O_5$	$[M+NH_4]^+$	636.5561	636.5563	0.0002	335.2595; 337.2758; 339.2616; 341.3071	
DG 36:4	$C_{39}H_{68}O_5$	$[M+NH_4]^+$	634.5405	634.5427	0.0022	335.2601; 337.2759; 339,2913	X
DG 36:5	$C_{39}H_{66}O_5$	$[M+NH_4]^+$	632.5248	632.5274	0.0026	335.2601; 337.2753	
DG 34:2	$C_{37}H_{68}O_5$	$[M+NH_4]^+$	610.5405	610.5428	0.0023	313.2761; 337.2862	
DG 34:3	$C_{37}H_{70}O_5$	$[M+NH_4]^+$	608.5248	608.5276	0.0028	311.2612; 313,2759; 335.2600; 337.2761	

TG 54:5	$C_{57}H_{100}O_6$	$[M+H]^+$	881.7520	881.7588	0.0068	599.5024; 601.5201	
TG 54:6	$C_{57}H_{98}O_6$	$[M+H]^+$	879.7363	879.7448	0.0085	597.4890; 599.5045; 601.5168	
TG 54:7	$C_{57}H_{96}O_6$	$[M+H]^+$	877.7280	877.7295	0.0015	597.4878; 599.5034	
TG 54:8	$C_{57}H_{94}O_6$	$[M+H]^+$	875.7123	875.7143	0.0020	595.4709; 597.4888	X
TG 52:5	$C_{55}H_{96}O_6$	$[M+H]^+$	853.7280	853.7301	0.0021	571.4701; 575.5050; 599.5042	
TG 52:4	$C_{55}H_{98}O_6$	$[M+H]^+$	855.7439	855.7539	0.0000	577.5165; 599.5041	
TG 44.4	$C_{47}H_{82}O_6$	$[M+H]^+$	743.6184	743.6190	0.0006	201.1492; 465.3955	
TG 44:5	$C_{47}H_{80}O_6$	$[M+H]^+$	741.6028	741.6058	0.0030	201.1150; 465.3938	
DG 36:4	$C_{39}H_{68}O_5$	$[M+H]^+$	617.5139	617.5147	0.0008	335.2597; 337.2765; 339.2912	X
DG 36:5	$C_{39}H_{66}O_5$	$[M+H]^+$	615.4983	615.5007	0.0024	335.2603; 337.2759	
DG 36:3	$C_{39}H_{70}O_5$	$[M+H-H_2O]^+$	601.5190	601.5217	0.0027	335.2586; 337.2773; 339.2916; 341.3090	
DG 34:2	$C_{37}H_{68}O_5$	$[M+H-H_2O]^+$	575.5034	575.5048	0.0014	239.2369; 313.2767	
DG 34:1	$C_{37}H_{70}O_5$	$[M+H-H_2O]^+$	577.5190	577.5208	0.0018	239.2374; 313.2754	X

Abbreviations: FA, fatty acid; MG, monoacylglycerol; DG, diacylglycerol; TG, triacylglycerol; PE, glycerophosphoethanolamine; PC, glycerophosphocholine

Table S6: *m/z* of the features selected by Borota applied to dataset 3 , which consits of 207 strawberry samples from Egypt, Germany, Greece, the Netherlands, Poland and Spain.

m/z	*m/z*	*m/z*	*m/z*
103.0638	256.2737	331.1546	495.2562
109.1027	256.2847	331.1656	501.3802
114.0928	257.1760	339.3139	501.3948
131.0704	257.2506	341.3051	501.4323
136.0224	265.2653	341.3177	503.3368
136.0625	268.1047	341.3301	506.3860
139.0498	270.3166	355.0711	551.3303
149.0452	274.2742	357.1344	577.5363
151.0763	274.2848	357.2984	607.3919
151.1118	275.2020	357.3082	610.5741
158.1547	277.2165	383.3313	617.5140
165.0552	277.2253	393.2416	617.5311
165.0920	280.2859	409.3871	634.5433
169.1338	281.0515	409.4011	638.5903
173.0811	282.2704	409.4138	638.6066
175.1480	282.2809	411.3517	647.4740
177.1274	282.2926	413.3537	647.4914
177.1642	282.3042	413.3674	650.5352
193.1594	282.3158	415.2118	663.4408
195.1022	282.3266	415.2259	663.4583
197.0825	283.2419	415.2400	663.4764
200.2377	285.2073	415.2530	663.4942
201.1645	285.2197	423.3260	682.6386
205.1599	286.9893	423.3991	706.6476
211.0973	287.0569	429.3710	715.5864
214.2549	287.0678	435.3269	875.6906
214.2638	291.1960	437.3571	875.7008
217.1069	293.1035	453.3258	928.8447
223.0645	293.2116	453.3382	
223.0717	293.2211	453.3519	
228.1966	297.2594	453.3656	
228.2700	298.2757	471.3423	
229.1556	298.2855	471.3623	
231.2120	301.0556	471.3742	
235.1714	310.239	473.3994	
235.1814	310.3122	473.4110	
239.2006	310.3243	487.3341	
247.1690	313.1443	489.3454	
250.9995	327.0091	489.3574	
256.2626	327.3048	489.3721	

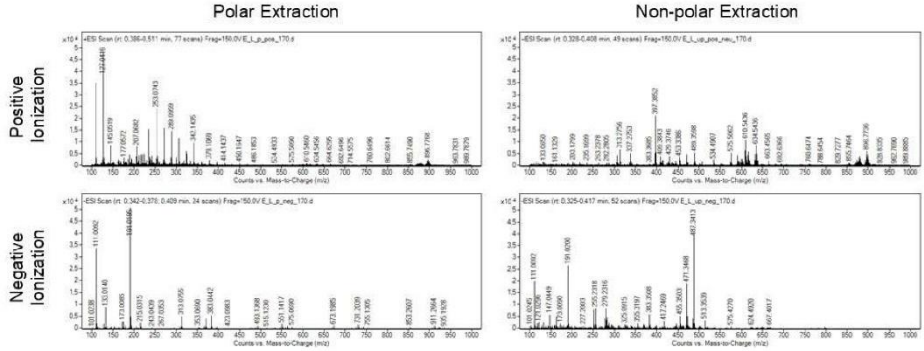


Figure S1: Mass spectra of a German strawberry sample that was homogenized and freeze-dried. The extraction was carried out with both polar and non-polar solvents. The results of DART-qToF measurements are shown in positive and negative ionization mode.

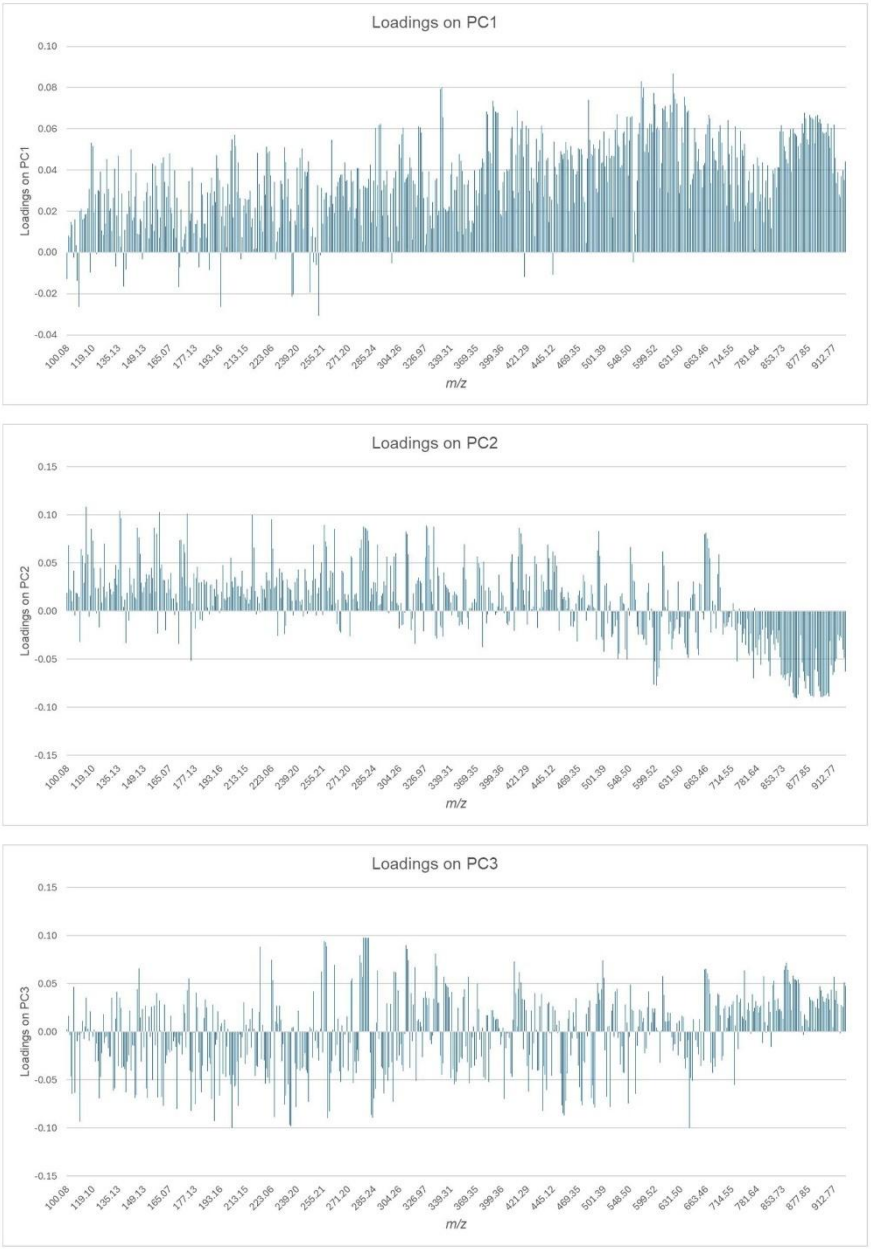


Figure S2: Shown are the loadings of the PCA of dataset 3 on PC 1, PC 2 and PC 3.

9.2 Supporting Information: Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS)

Electronic Supplementary Information

Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS)

Felix Schmauder[1], David Schütz[1], Marina Creydt[1], Janet Riedl[2], Markus Fischer[1,*]

[1]Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany
[2]Bundesinstitut für Risikobewertung, BfR, Max-Dohrn-Straße 8-10, 10589 Berlin, Germany

*Corresponding author:

Markus Fischer, Hamburg School of Food Science - Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

Tel.: +49 40 42838-4359, Fax: +49 40 42838-4342

Email: markus.fischer@uni-hamburg.de

Table A1: List of samples and available metadata.

Number	Origin	State	Variety
1	France	Alsace	not specified
2	France	Alsace	not specified
3	France	Alsace	not specified
4	France	Alsace	not specified
5	France	Alsace	not specified
6	France	Alsace	not specified
7	France	Alsace	not specified
8	France	Alsace	not specified
9	France	Alsace	not specified
10	France	Alsace	not specified
11	France	Alsace	not specified
12	France	Alsace	not specified
13	France	Alsace	not specified
14	France	Alsace	not specified
15	France	Alsace	not specified
16	France	Alsace	not specified
17	France	Alsace	not specified
18	France	Alsace	not specified
19	France	Alsace	not specified
20	France	Alsace	not specified
21	France	Alsace	not specified
22	France	Alsace	not specified
23	France	Alsace	not specified
24	France	Alsace	not specified
25	France	Alsace	not specified
26	France	Alsace	not specified
27	France	Alsace	not specified
28	France	Alsace	not specified
29	France	Alsace	not specified
30	France	Alsace	not specified
31	France	Alsace	not specified
32	France	Alsace	not specified
33	France	Alsace	not specified
34	France	Alsace	not specified
35	France	Alsace	not specified
36	France	Alsace	not specified
37	France	Alsace	not specified
38	Germany	North Rhine-Westphalia	Amagrano
39	Germany	Brandenburg	Amagrano
40	Germany	Bavaria	Amagrano
41	Germany	North Rhine-Westphalia	ES Crossman
42	Germany	Brandenburg	ES Crossman
43	Germany	Bavaria	ES Crossman
44	Germany	North Rhine-Westphalia	SY Talisman
45	Germany	Brandenburg	SY Talisman
46	Germany	Bavaria	SY Talisman
47	Germany	North Rhine-Westphalia	KWS Stabil
48	Germany	Brandenburg	KWS Stabil
49	Germany	North Rhine-Westphalia	Benedictio KWS
50	Germany	Brandenburg	Benedictio KWS
51	Germany	Bavaria	Benedictio KWS
52	Germany	North Rhine-Westphalia	Ridley

Table A1 continued.

Number	Origin	State	Variety
53	Germany	Brandenburg	Ridley
54	Germany	Bavaria	Ridley
55	Germany	North Rhine-Westphalia	ES Hubble
56	Germany	Brandenburg	ES Hubble
57	Germany	Bavaria	ES Hubble
58	Germany	North Rhine-Westphalia	Galactus
59	Germany	Brandenburg	Galactus
60	Germany	Bavaria	Galactus
61	Germany	North Rhine-Westphalia	P 8723
62	Germany	Brandenburg	P 8723
63	Germany	Bavaria	P 8723
64	Hungary	not specified	not specified
65	Hungary	not specified	not specified
66	Hungary	not specified	not specified
67	Hungary	not specified	not specified
68	Hungary	not specified	not specified
69	Hungary	not specified	not specified
70	Hungary	not specified	not specified
71	Hungary	not specified	not specified
72	Hungary	not specified	not specified
73	Hungary	not specified	not specified
74	Hungary	not specified	not specified
75	Hungary	not specified	not specified
76	Hungary	not specified	not specified
77	Hungary	not specified	not specified
78	Hungary	not specified	not specified
79	Hungary	not specified	not specified
80	Hungary	not specified	not specified
81	Hungary	not specified	not specified
82	Hungary	not specified	not specified
83	Hungary	not specified	not specified
84	Hungary	not specified	not specified
85	Hungary	not specified	not specified
86	Hungary	not specified	not specified
87	Hungary	not specified	not specified
88	Hungary	not specified	not specified
89	Hungary	not specified	not specified
90	Hungary	not specified	not specified
91	Hungary	not specified	not specified
92	Hungary	not specified	not specified
93	Hungary	not specified	not specified
94	Hungary	not specified	not specified
95	Hungary	not specified	not specified
96	Hungary	not specified	not specified
97	Hungary	not specified	not specified
98	Hungary	not specified	not specified
99	Hungary	not specified	not specified
100	Hungary	not specified	not specified
101	Hungary	not specified	not specified
102	Hungary	not specified	not specified
103	Hungary	not specified	not specified
104	Hungary	not specified	not specified
105	Hungary	not specified	not specified

Table A1 continued.

Number	Origin	State	Variety
106	Hungary	not specified	not specified
107	Hungary	not specified	not specified
108	Hungary	not specified	not specified
109	Hungary	not specified	not specified
110	Hungary	not specified	not specified
111	Hungary	not specified	not specified
112	Hungary	not specified	not specified
113	Hungary	not specified	not specified
114	Peru	Pátapo	not specified
115	Peru	Pátapo	not specified
116	Peru	Pátapo	not specified
117	Peru	Tumán	not specified
118	Peru	Tumán	not specified
119	Peru	Pomalca	not specified
120	Peru	Pomalca	not specified
121	Peru	Pomalca	not specified
122	Peru	Ferrenafe	not specified
123	Peru	Ferrenafe	not specified
124	Peru	Mochumí	not specified
125	Peru	Mochumí	not specified
126	Peru	Mochumí	not specified
127	Peru	Túcume	not specified
128	Peru	Túcume	not specified
129	Peru	Pacora	not specified
130	Peru	Pacora	not specified
131	Peru	Pacora	not specified
132	Peru	Jayanca	not specified
133	Peru	Jayanca	not specified
134	Peru	Mórrope	not specified
135	Peru	Mórrope	not specified
136	Peru	Mórrope	not specified
137	Peru	Lambayeque	not specified
138	Peru	Lambayeque	not specified
139	Spain	Lleida	68.K
140	Spain	Lleida	CHARLESTON
141	Spain	Lleida	DKC6728
142	Spain	Lleida	DKC6650
143	Spain	Lleida	DKC5830
144	Spain	Lleida	IXABEL
145	Spain	Lleida	KEFIEROS
146	Spain	Lleida	KERIDOS
147	Spain	Lleida	KONTIGOS
148	Spain	Lleida	LG30.600
149	Spain	Lleida	LG30.681
150	Spain	Lleida	MAS 64.P
151	Spain	Lleida	MAS 66.C
152	Spain	Lleida	MAS 75.A
153	Spain	Lleida	MILOXAN
154	Spain	Lleida	NYSTAR
155	Spain	Lleida	P1524
156	Spain	Lleida	P1570
157	Spain	Lleida	P1574
158	Spain	Lleida	P1921

Table A1 continued.

Number	Origin	State	Variety
159	Ukraine	not specified	not specified
160	Ukraine	not specified	not specified
161	Ukraine	not specified	not specified
162	Ukraine	not specified	not specified
163	Ukraine	not specified	not specified
164	Ukraine	not specified	not specified
165	Ukraine	not specified	not specified
166	Ukraine	not specified	not specified
167	Ukraine	not specified	not specified
168	Ukraine	not specified	not specified
169	Ukraine	not specified	not specified
170	Ukraine	not specified	not specified
171	Ukraine	not specified	not specified
172	Ukraine	not specified	not specified
173	Ukraine	not specified	not specified
174	Ukraine	not specified	not specified
175	Ukraine	not specified	not specified
176	Ukraine	not specified	not specified
177	USA	Alabama	not specified
178	USA	Alabama	not specified
179	USA	Alabama	not specified
180	USA	Alabama	not specified
181	USA	Arkansas	not specified
182	USA	Arkansas	not specified
183	USA	Arkansas	not specified
184	USA	Arkansas	not specified
185	USA	Georgia	not specified
186	USA	Georgia	not specified
187	USA	Georgia	not specified
188	USA	Georgia	not specified
189	USA	Lousiana	not specified
190	USA	Lousiana	not specified
191	USA	Lousiana	not specified
192	USA	Lousiana	not specified
193	USA	Mississippi	not specified
194	USA	Mississippi	not specified
195	USA	Mississippi	not specified
196	USA	Mississippi	not specified
197	USA	Texas	not specified
198	USA	Texas	not specified
199	USA	Texas	not specified
200	USA	Texas	not specified

Table A2: The 100 most important variables with their mean decrease in accuracy, calculated using Random Forest. The entire model is broken down by country (Spain (Esp), France (Fra), Germany (Ger), Peru (Per), Ukraine (Ukr), Hungary (Hun), and the USA).

	m/z	Esp	Fra	Ger	Per	Ukr	Hun	USA	Entire model
1	885.8209	0.4	7.3	2.5	2.8	1.4	7.3	-0.3	8.1
2	610.5209	1.5	1.3	2.0	3.3	2.2	8.0	3.5	7.5
3	310.2587	5.2	6.2	2.6	2.3	5.5	4.6	4.6	7.3
4	615.5158	5.2	2.4	4.4	1.9	1.3	2.8	0.9	6.7
5	383.2408	3.7	1.5	2.7	5.9	1.4	3.7	2.1	6.1
6	229.1412	1.5	3.5	3.3	3.6	2.3	5.6	1.4	5.7
7	892.8179	0.7	1.1	0.9	5.4	-1.2	3.4	1.1	5.6
8	311.2408	1.9	4.2	2.9	0.6	1.2	4.0	2.6	5.1
9	415.2257	4.1	1.9	-0.7	0.7	3.4	2.4	5.2	5.1
10	467.3858	2.3	0.1	4.5	0.6	-0.1	4.0	-0.7	4.9
11	650.5320	2.9	3.9	0.5	3.6	0.4	2.6	1.0	4.8
12	618.4990	3.3	0.6	0.5	-0.4	-0.8	3.4	1.4	4.5
13	311.2231	1.5	3.1	1.7	3.2	1.3	3.8	0.4	4.5
14	610.5551	1.9	0.2	1.5	2.0	1.0	2.3	1.0	4.4
15	597.4853	4.2	1.4	-0.2	1.0	2.0	3.1	2.9	4.4
16	631.4921	3.2	-0.5	4.6	0.9	1.8	4.3	1.0	4.3
17	481.3878	-1.0	2.8	3.7	2.3	1.0	2.1	-0.1	4.3
18	577.5221	0.6	2.6	3.7	-0.8	1.1	3.4	1.2	4.3
19	634.5755	0.7	2.3	2.8	2.3	2.9	3.5	2.3	4.3
20	479.3726	1.7	2.3	2.9	1.4	1.6	3.3	2.1	4.2
21	507.4031	-1.2	3.9	3.7	2.5	-2.0	0.7	-0.3	4.2
22	650.5496	-1.1	0.3	0.0	3.4	2.3	1.2	3.1	4.2
23	235.1680	2.9	2.8	1.5	1.5	2.0	2.8	-1.2	4.2
24	846.6815	0.2	1.9	3.1	0.9	2.0	1.6	-1.0	4.1
25	617.4922	-0.5	1.0	1.6	2.1	0.8	3.9	1.8	4.1
26	575.5014	2.2	3.0	1.9	2.2	0.1	3.6	3.0	4.0
27	398.3847	-0.6	4.2	-0.9	1.7	1.1	1.0	1.4	4.0
28	309.2050	1.8	1.5	1.5	1.1	3.3	2.8	-0.2	4.0
29	355.2943	0.5	2.7	1.6	1.3	3.0	3.0	2.5	4.0
30	603.5468	0.6	1.5	2.8	2.8	0.4	2.7	-0.4	3.8
31	664.4540	2.2	2.1	0.1	2.6	1.9	0.5	2.7	3.8
32	505.3862	0.4	1.6	1.9	2.9	0.4	3.7	1.5	3.8
33	633.5067	1.2	4.0	3.1	2.1	-0.8	3.9	-0.4	3.8
34	903.8226	2.8	0.6	-0.1	2.2	2.0	3.0	2.3	3.7
35	918.8242	1.5	1.0	2.2	3.0	-1.1	1.5	1.9	3.7
36	617.5092	1.6	-0.5	1.7	2.4	-0.7	3.4	-0.3	3.7
37	227.1265	3.0	0.5	-0.4	3.2	2.3	2.5	2.9	3.7
38	591.4963	-0.6	1.3	3.4	0.1	0.8	2.3	-0.9	3.7
39	494.3919	1.0	2.1	1.5	2.4	-1.1	1.4	1.9	3.7
40	575.5180	3.7	2.2	0.3	0.0	0.3	-1.1	1.5	3.7
41	313.2343	-0.9	1.1	1.6	1.9	2.7	1.5	-0.6	3.6
42	663.4721	1.4	2.6	1.9	1.5	1.2	1.8	1.9	3.6
43	578.5023	1.0	0.4	1.5	3.7	0.1	2.8	-0.5	3.6
44	297.2423	2.1	0.0	2.5	1.9	1.7	-0.2	2.8	3.6
45	423.3963	-0.5	3.3	0.9	1.9	0.7	0.1	2.4	3.6
46	664.4719	2.0	3.0	0.1	2.6	2.6	1.1	1.4	3.5
47	605.5290	1.1	3.3	1.1	3.5	2.2	2.2	0.4	3.5
48	615.4821	3.6	2.6	0.4	1.7	1.1	3.2	1.2	3.5

Table A2 continued.

	m/z	Esp	Fra	Ger	Per	Ukr	Hun	USA	Entire model
49	229.1522	2.9	0.9	0.0	2.6	1.9	1.6	0.9	3.4
50	601.5333	0.8	1.5	-0.5	3.2	-0.2	1.4	0.6	3.4
51	577.5553	1.2	0.7	2.7	1.3	0.1	2.3	-1.7	3.4
52	211.1435	1.4	1.8	0.9	1.7	2.5	2.8	2.6	3.4
53	282.2884	2.0	2.8	0.3	0.0	1.2	1.6	2.6	3.3
54	599.4831	2.0	0.2	-0.6	-1.0	1.0	2.6	1.0	3.3
55	634.5406	1.1	1.7	-0.3	1.9	2.3	1.6	2.2	3.3
56	265.2615	1.0	1.0	-0.1	1.9	-1.4	2.3	2.2	3.2
57	573.4855	-0.2	1.7	-1.2	0.6	1.2	2.9	2.8	3.2
58	279.2557	2.7	1.4	2.2	1.0	1.7	-1.0	2.1	3.2
59	900.8392	0.9	0.9	2.2	1.4	1.4	1.4	1.0	3.2
60	692.6312	2.1	-1.1	0.1	0.5	3.0	2.0	1.4	3.1
61	415.2116	1.4	1.6	0.9	0.9	-0.6	0.6	3.2	3.1
62	282.3000	2.0	1.2	1.1	0.2	-0.2	0.9	2.7	3.1
63	393.3505	1.1	2.2	-0.5	0.9	0.1	2.1	1.1	3.1
64	295.2248	1.5	0.2	1.2	1.1	0.7	2.0	0.0	3.1
65	605.5630	0.1	2.4	1.4	0.8	0.9	1.8	-0.2	3.0
66	355.3073	1.6	0.6	-0.1	2.2	0.0	-1.0	1.8	3.0
67	634.5930	1.7	0.1	-1.4	2.5	0.3	1.4	1.6	3.0
68	736.6621	1.5	0.5	1.0	1.4	1.5	2.1	0.6	3.0
69	265.2502	2.4	0.6	-0.8	1.8	2.1	1.8	0.5	3.0
70	680.4813	1.5	2.3	-1.6	1.3	-0.3	1.4	3.1	3.0
71	310.2343	1.7	1.7	-1.0	0.1	1.9	1.4	1.7	2.9
72	279.2441	1.4	1.7	-1.1	0.3	2.1	1.4	0.7	2.9
73	743.6175	1.4	0.3	1.4	-0.5	1.8	2.8	-1.0	2.9
74	696.6632	-1.4	0.4	1.4	0.0	3.0	1.0	-0.7	2.9
75	219.1734	1.0	-0.2	1.8	1.0	0.0	2.2	1.0	2.9
76	397.3841	1.7	2.1	1.0	-0.8	1.8	2.4	0.4	2.9
77	413.3868	0.0	0.9	1.4	0.7	1.0	2.4	1.7	2.9
78	738.6568	1.1	-0.9	1.2	3.2	2.3	0.0	-1.4	2.9
79	467.3708	-0.5	0.3	1.0	1.4	1.0	1.9	0.2	2.8
80	493.2931	-0.8	1.7	1.1	2.1	2.1	0.7	1.7	2.8
81	806.7025	2.0	0.3	2.2	1.4	1.3	1.2	1.2	2.8
82	663.4364	2.2	1.7	-1.0	0.8	2.0	-1.0	2.3	2.8
83	694.6643	-1.0	1.5	1.4	1.9	1.9	1.4	-0.4	2.8
84	577.5055	0.5	-0.7	0.2	1.3	0.2	2.6	0.5	2.8
85	654.5637	-0.1	2.3	0.2	2.1	-1.4	0.7	1.1	2.7
86	864.7612	0.8	1.2	2.6	-0.1	1.4	0.6	-1.0	2.7
87	415.3913	0.1	1.6	0.9	0.4	1.0	1.9	1.2	2.7
88	833.7515	1.4	1.0	0.3	0.9	-0.3	3.0	0.9	2.7
89	257.2237	1.6	-0.1	0.0	0.0	1.4	1.4	1.4	2.7
90	618.5506	2.5	0.5	0.7	1.0	1.1	1.4	2.4	2.7
91	575.5512	1.9	-1.4	1.5	1.2	0.5	1.2	0.0	2.7
92	855.7379	0.4	0.7	1.0	1.7	1.4	2.1	0.0	2.7
93	491.3874	1.4	0.5	2.2	0.0	-0.2	2.6	2.1	2.7
94	578.5522	0.2	1.5	1.5	2.0	0.9	1.9	0.2	2.7
95	353.2815	1.8	-1.8	1.7	2.3	1.4	1.5	2.1	2.6
96	874.8141	1.8	0.3	1.1	1.3	2.2	1.6	-0.2	2.6
97	881.7675	1.0	0.6	0.8	-0.1	1.7	2.5	1.4	2.6
98	209.1517	1.0	0.8	0.6	2.1	3.0	-0.6	0.3	2.6
99	887.7996	0.1	0.2	-1.6	3.0	1.9	0.5	1.3	2.6
100	764.6408	1.9	2.1	-0.2	2.9	0.3	0.4	-1.7	2.6

Table A3: Results of the PLS-DA classification for the geographical origin determination of grain maize samples. The samples contributed to the classes of Spain, France, Germany, Hungary, Peru, Ukraine and the USA. Validation was performed using 10-fold cross validation

Number of components	1	2	3	4	5	6	7	8
Accuracy	0.36	0.47	0.65	0.68	0.72	0.82	0.87	0.85
R^2	0.12	0.25	0.39	0.43	0.55	0.64	0.72	0.76
Q^2	0.07	0.15	0.31	0.36	0.42	0.46	0.54	0.62

Table A4: Cost breakdown of all material costs for a DART-MS analysis compared to LC-MS analysis of a sample. Listed are the purchase price of the unit, the number of pieces of the unit and the price of a single analysis added separately for DART-MS and LC-MS. The table refers to the price level of May 2023.

	Price per unit [€]	Pieces per unit	Price DART-MS [€]	Price LC-MS [€]	References
DIP-it Sampler	181.95	96	1.90	/	[1]
Helium 5.0	1190.00	21600	0.06	/	[2]
UHPLC-column	951.72	1500	/	0.63	[3]
Vial	17.80	100	/	0.18	[4]
Cap for vial	26.30	100	/	0.26	[5]
Acetonitrile 1L	184.00	100	/	1.84	[6]
Methanol 1L	50.50	500	0.10	0.10	[7]
Chloroform 2.5L	154.00	5000	0.03	0.03	[8]
3x Reaction tube	38.40	166	0.23	0.23	[9]
2 steelballs	12.19	500	0.02	0.02	[10]
Pipette tips	30.90	250	0.12	0.12	[11]
Sum in €	/	/	2.46	3.41	

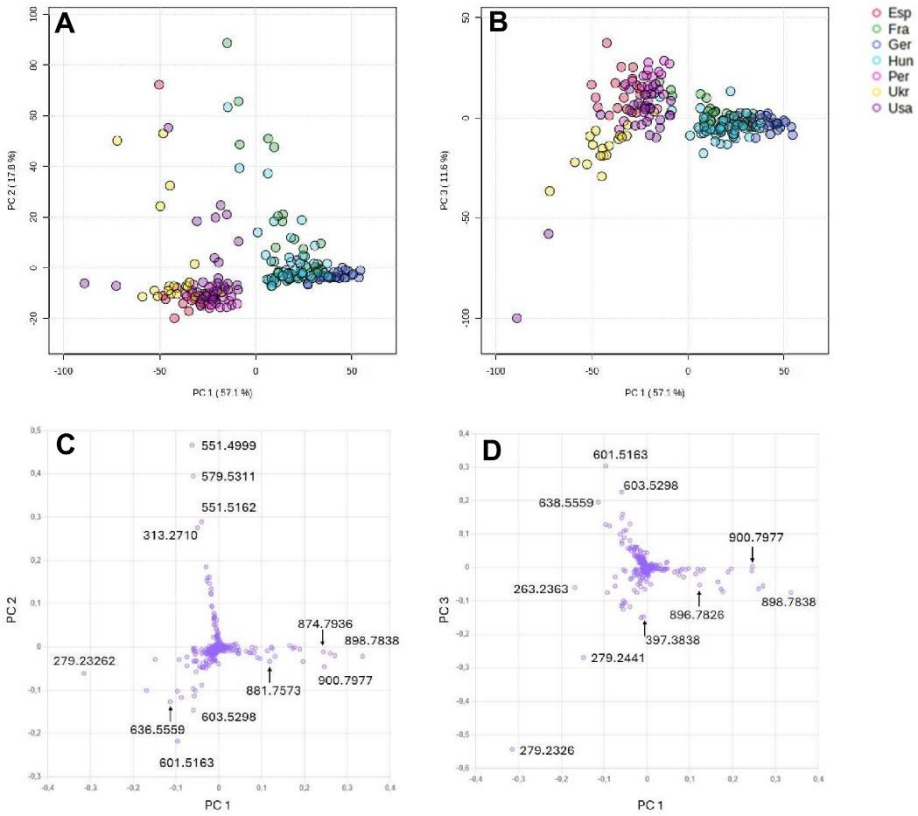


Figure A1: PCA of all samples from Spain, France, Germany, Peru, Ukraine, Hungary and the USA. Shown are the score plot of PC1 vs. PC2 (A), the score plot of PC1 vs.PC3 (B) and the corresponding loading plots of PC1 vs. PC2 (C) and PC1 vs. PC3 (D). The *m/z* ratios are given for important features with high loadings.

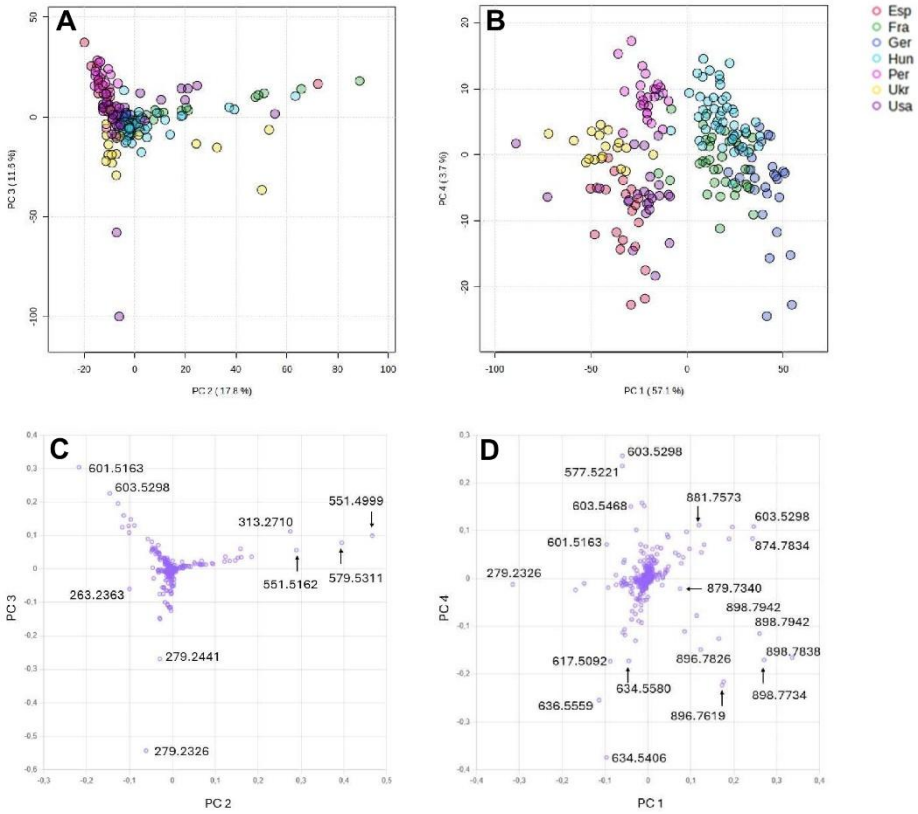


Figure A2: PCA of all samples from Spain, France, Germany, Peru, Ukraine, Hungary and the USA. Shown are the score plot of PC2 vs. PC3 (A), the score plot of PC1 vs.PC4 (B) and the corresponding loading plots of PC2 vs. PC3 (C) and PC1 vs. PC4 (D). The *m/z* ratios are given for important features with high loadings.

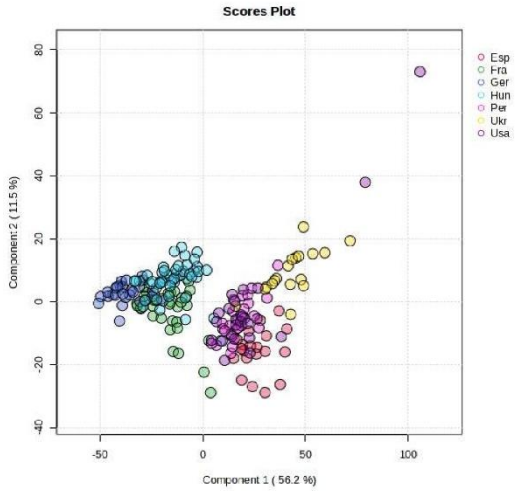


Figure A3: PLS-DA of all samples Spain, France, Germany, Peru, Ukraine, Hungary and the USA. Shown is the score plot of component 1 vs. component 2, corresponding to a variance cover of 67.7%.

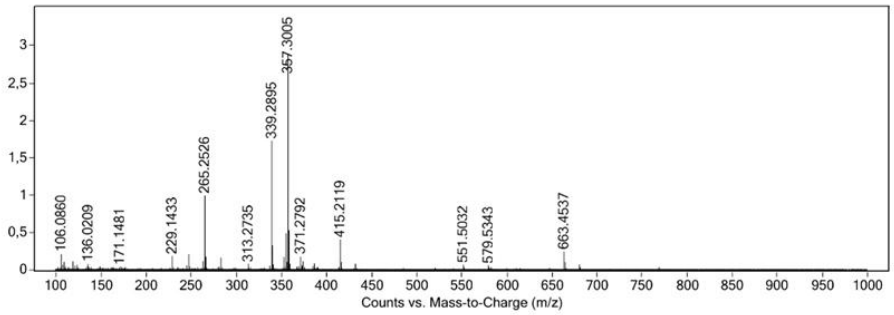


Figure A4: MS[1] spectrum of MG 18:1.

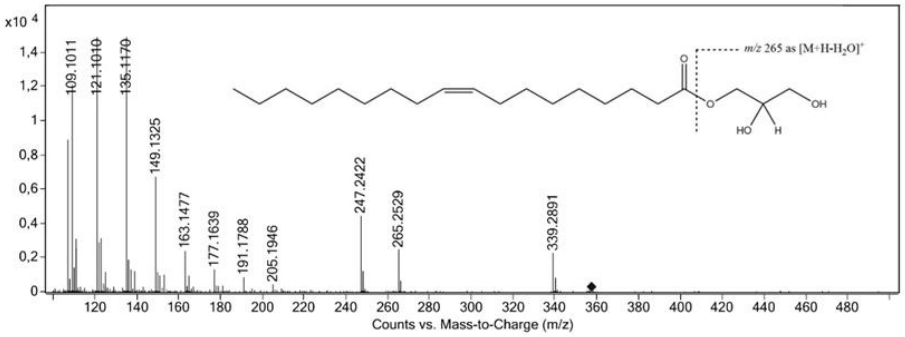


Figure A5: MS[2] spectrum of MG 18:1.

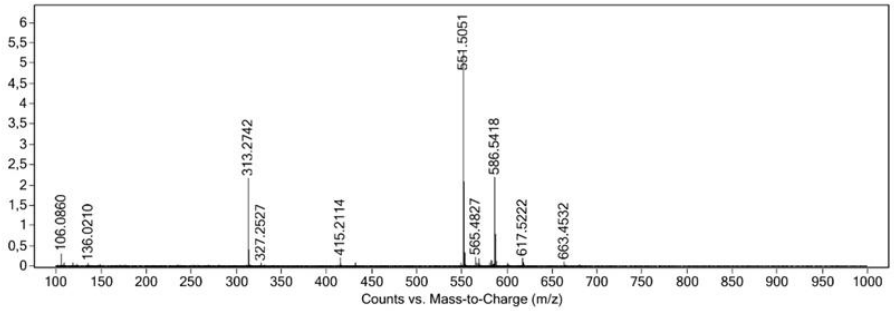


Figure A6: MS[1] spectrum of DG 32:0.

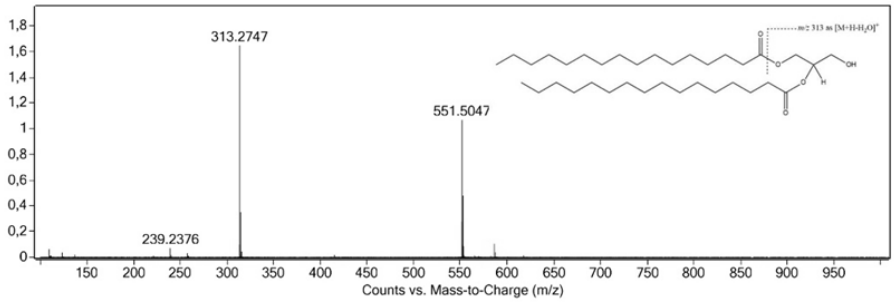


Figure A7: MS^2 spectrum of DG 32:0.

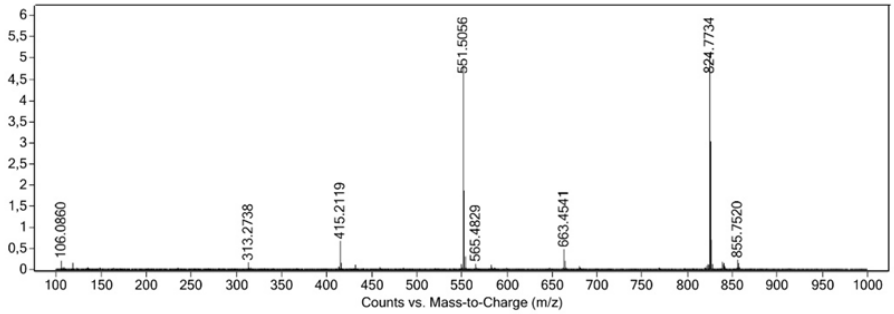


Figure A8: MS^1 spectrum of TG 48:0.

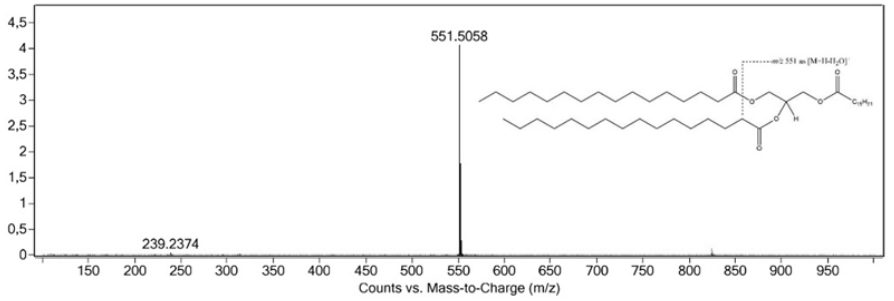


Figure A9: MS^2 spectrum of TG 48:0.

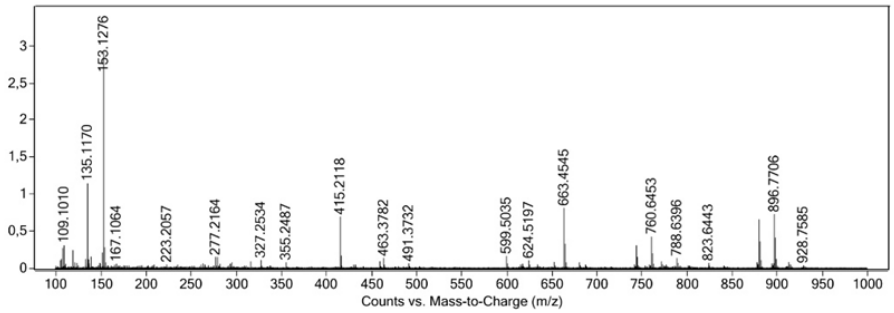


Figure A10: MS^1 spectrum of TG 54:6.

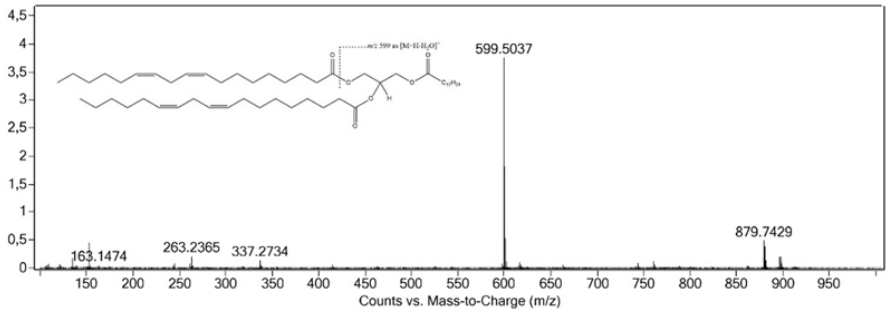


Figure A11: MS^2 spectrum of TG 54:6.

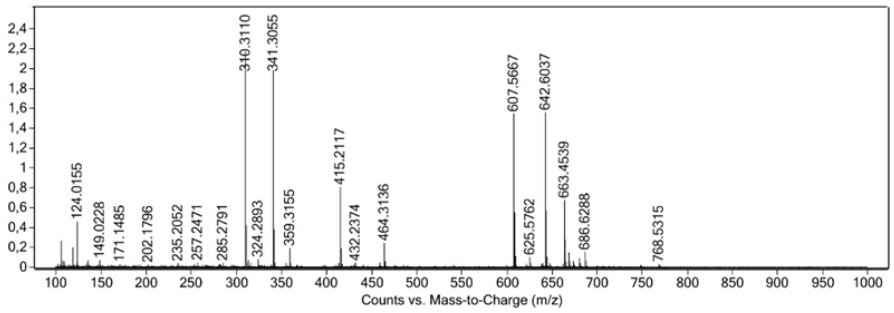


Figure A12: MS^1 spectrum of PE 36:0.

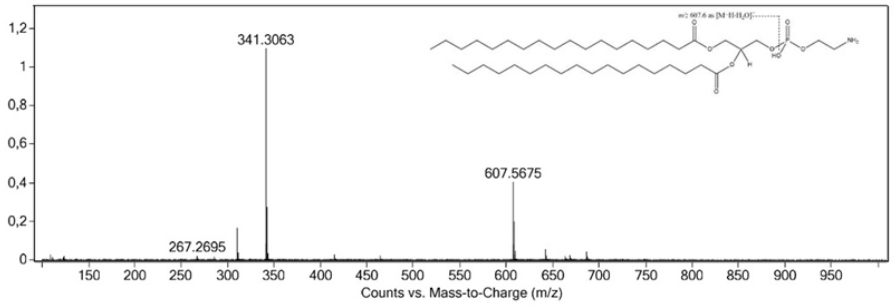


Figure A13: MS^2 spectrum of PE 36:0.

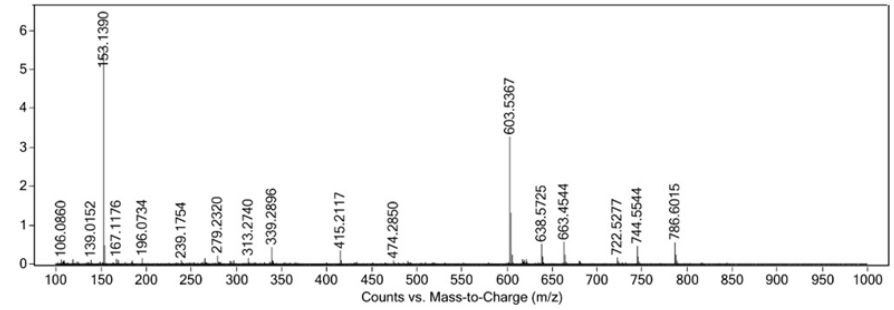


Figure A14: MS[1] spectrum of PC 36:2.

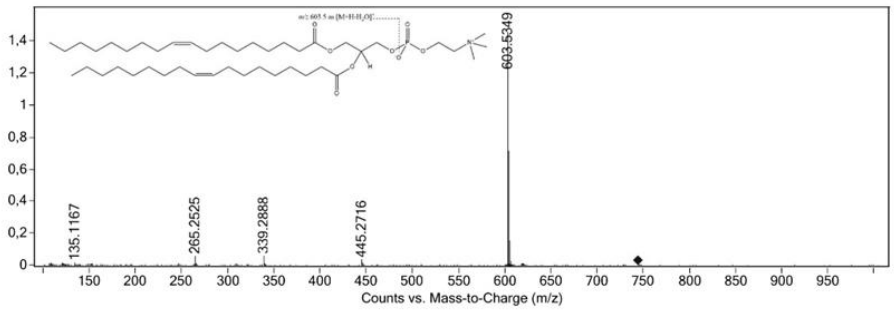


Figure A15: MS[2] spectrum of PC 36:2.

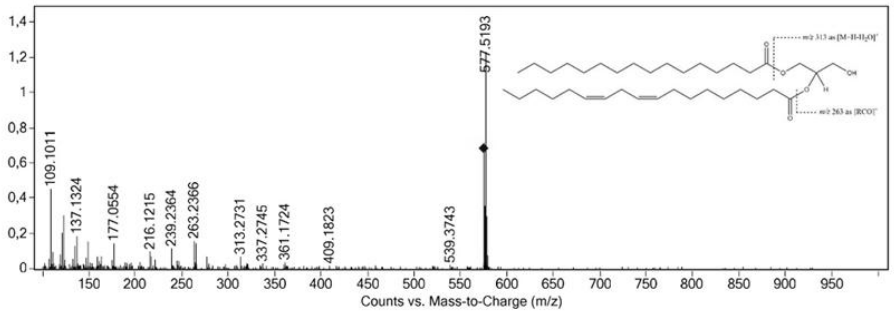


Figure A16: MS[2] spectrum of DG 34:2.

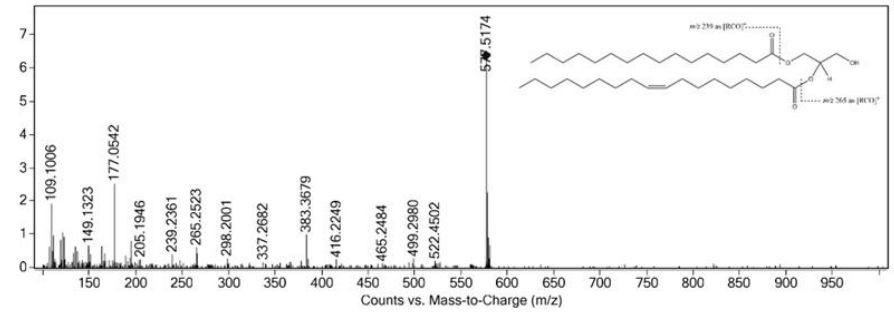


Figure A17: MS[2] spectrum of DG 34:1.

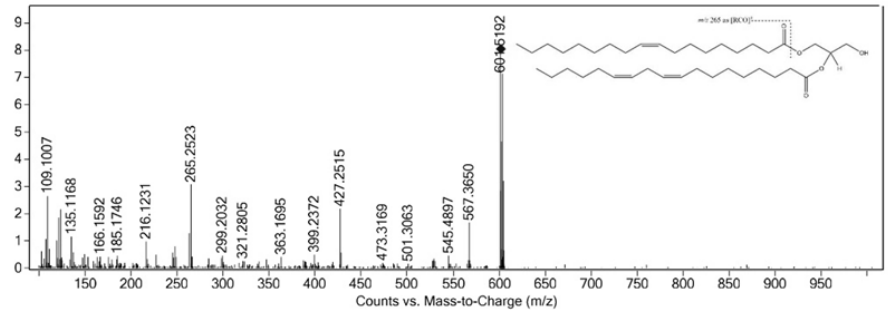


Figure A18: MS^2 spectrum of DG 36:3.

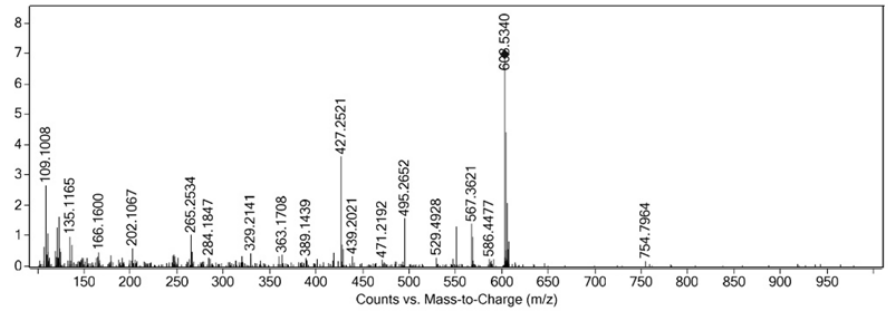


Figure A19: MS^2 spectrum of DG 36:2.

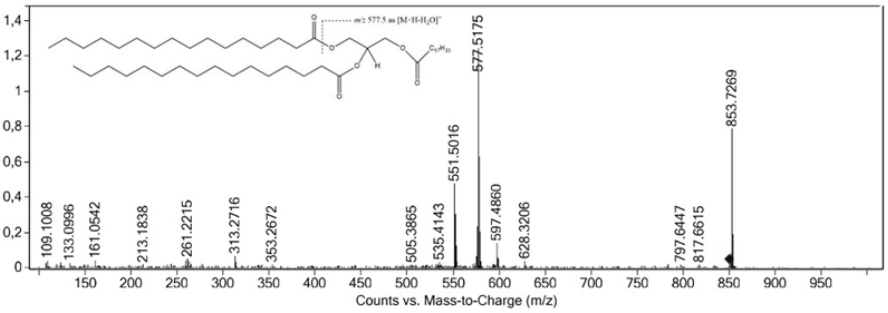


Figure A20: MS^2 spectrum of TG 50:1.

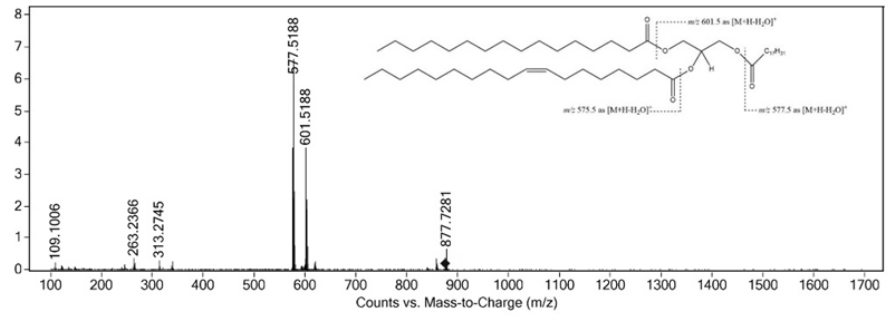


Figure A21: MS^2 spectrum of TG 52:3.

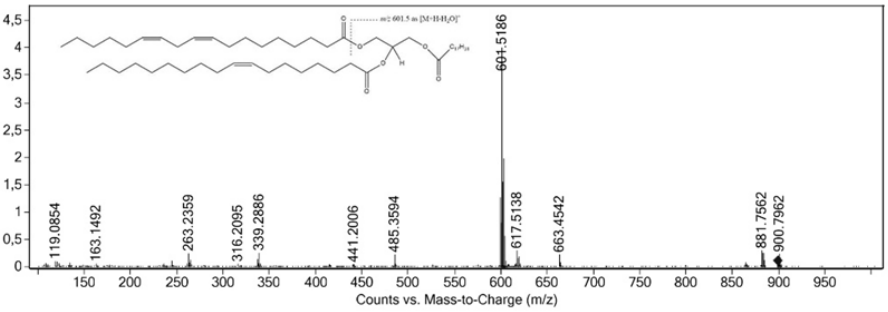


Figure A22: MS^2 spectrum of TG 54:5.

References

1. Bruker. Available online: https://store.bruker.com/en-DE/products/dip-it-sampler (accessed on 25 May 2023).
2. Linde AG. Available online: https://www.linde-gas.de/shop/de/de-ig/helium-50-helium-50 (accessed on 26 May 2023).
3. Thermo Fisher Scientific. Available online: https://www.thermofisher.com/order/catalog/product/de/de/17626-032130.
4. Macherey-Nagel. Available online: https://www.mn-net.com/de/gewindeflasche-n-9-11-6x32-0-mm-1-5-ml-schriftfeld-flacher-boden-braun-702284?c=4540 (accessed on 26 May 2023).
5. Macherey-Nagel. Available online: https://www.mn-net.com/de/schraubverschluss-n-9-pp-schwarz-loch-silikon-weiss/ptfe-rot-1-0-mm-702036?c=4544 (accessed on 26 May 2023).
6. Merck Germany. Acetonitrile hypergrade for LC-MS LiChrosolv®. Available online: https://www.merckmillipore.com/DE/de/product/Acetonitrile,MDA_CHEM-100029 (accessed on 16 March 2023).
7. Merck Germany. Available online: https://www.merckmillipore.com/DE/de/product/Methanol,MDA_CHEM-106035 (accessed on 26 May 2023).
8. Merck Germany. Available online: https://www.merckmillipore.com/DE/de/product/Chloroform,MDA_CHEM-107024 (accessed on 26 May 2023).
9. Carl Roth GmbH + Co. KG. Available online: https://www.carlroth.com/de/de/reaktionsgefaesse/reaktionsgefaesse-2-ml/p/ck06.1 (accessed on 26 May 2023).
10. Technischer Handel Straub. Available online: https://www.ths-straub.de/Kugellager-Keilriemen/Kugellager/Stahlkugeln-kaufen/Stahlkugeln-3-mm-Niro-Edelstahl-VE-1-000-Stk:3243.html (accessed on 26 May 2023).
11. Carl Roth GmbH + Co. KG. Available online: https://www.carlroth.com/de/de/pipettenspitzen/pipettenspitzen-m%CE%BClti-universal-100-1000-%CE%BCl/p/8160.1 (accessed on 26 May 2023).

9.3 Supporting Information zu: Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (*Corylus avellana* L.)

Electronic Supplementary Information

Novel DART-MS approach for rapid and environmentally friendly determination of the geographical origin of hazelnuts (*Corylus avellana* L.)

Felix Schmauder[1], Marina Creydt[1], Markus Fischer[1,*]

[1]Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

*Corresponding author:

Markus Fischer, Hamburg School of Food Science - Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

Tel.: +49 40 42838-4359, Fax: +49 40 42838-4342

Email: markus.fischer@uni-hamburg.de

Table S1: Metadata of the samples used in this study.

Number	Origin	State	Harvest Year	Variety
1	France	Aquitanien	2014	Pauetet
2	France	Aquitanien	2014	Tonda di Giffoni
3	France	Aquitanien	2014	Ennis
4	France	Aquitanien	2014	Butler
5	France	Aquitanien	2014	Corabel
6	France	Centre	2014	Ennis
7	France	Centre	2014	Pauetet
8	France	Midi-Pyrénées	2014	Corabel
9	France	Midi-Pyrénées	2014	Segorbe
10	France	Midi-Pyrénées	2014	Ennis
11	France	Midi-Pyrénées	2014	Pauetet
12	France	Poitou-Charentes	2014	Pauetet
13	France	Poitou-Charentes	2014	Corabel
14	France	Pays de la Loire	2014	Fertile de Coutard
15	France	Pays de la Loire	2014	Ennis
16	France	Midi-Pyrénées	2015	Pauetet
17	France	Aquitanien	2015	Barcelona
18	France	Aquitanien	2015	Segorbe
19	France	Aquitanien	2015	Ennis
20	France	Aquitanien	2015	Lewis
21	France	Aquitanien	2015	Segorbe
22	France	Centre	2015	Corabel
23	France	Centre	2015	Ennis
24	France	Centre	2015	Pauetet
25	France	Centre	2015	Segorbe
26	France	Midi-Pyrénées	2015	Corabel
27	France	Midi-Pyrénées	2015	Segorbe
28	France	Midi-Pyrénées	2015	Ennis
29	France	Poitou-Charentes	2015	Pauetet
30	France	Poitou-Charentes	2015	Corabel
31	France	Poitou-Charentes	2015	Tonda di Giffoni
32	France	Pays de la Loire	2015	Barcelona
33	France	Pays de la Loire	2015	Ennis
34	France	Pays de la Loire	2015	Ennis
35	France	Aquitanien	2016	Tonda di Giffoni
36	France	Aquitanien	2016	Corabel
37	France	Aquitanien	2016	Fertile de Coutard
38	France	Aquitanien	2016	Pauetet
39	France	Centre	2016	Melange
40	France	Centre	2016	Corabel
41	France	Centre	2016	Ennis
42	France	Centre	2016	Segorbe
43	France	Centre	2016	Pautet
44	France	Midi-Pyrénées	2016	Corabel
45	France	Midi-Pyrénées	2016	Melange
46	France	Midi-Pyrénées	2016	Corabel
47	France	Midi-Pyrénées	2016	Fertile de Coutard
48	France	Poitou-Charentes	2016	Ennis
49	France	Poitou-Charentes	2016	Pautet
50	France	Pays de la Loire	2016	Barcelona
51	Georgia	Gurien	2014	Guria

52	Georgia	Imeretien	2014	Mixture
53	Georgia	Imeretien	2014	Mixture
54	Georgia	Imeretien	2014	Vani
55	Georgia	Mingrelien und Oberswanetien	2014	Mixture
56	Georgia	Mingrelien und Oberswanetien	2014	Mixture
57	Georgia	Mingrelien	2014	Migrelien
58	Georgia	Not specified	2014	Mixture
59	Georgia	Not specified	2014	Mixture
60	Georgia	Not specified	2014	Mixture
61	Georgia	Mingrelien	2015	Mixture
62	Georgia	Mingrelien und Oberswanetien	2015	Megrelia
63	Georgia	Not specified	2015	Mixture
64	Georgia	Not specified	2015	Mixture
65	Georgia	Not specified	2015	Mixture
66	Germany	Baden-Wuerttemberg	2014	Mixture
67	Germany	Bavaria	2014	Mixture
68	Germany	Bavaria	2014	Mixture
69	Germany	Bavaria	2014	Mixture
70	Germany	Bavaria	2014	Mixture
71	Germany	Bavaria	2014	Corabel
72	Germany	Bavaria	2014	Tonda Romana
73	Germany	Bavaria	2014	Tonda di Giffoni
74	Germany	Bavaria	2014	Not specified
75	Germany	Bavaria	2014	Daria
76	Germany	Bavaria	2014	Mixture
77	Germany	Bavaria	2014	Mixture
78	Germany	Bavaria	2014	Mixture
79	Germany	Bavaria	2014	Mixture
80	Germany	Saxony	2014	Mixture
81	Germany	Baden-Wuerttemberg	2015	Mixture
82	Germany	Bavaria	2015	Mixture
83	Germany	Bavaria	2015	Mixture
84	Germany	Bavaria	2015	Mixture
85	Germany	Bavaria	2015	Mixture
86	Germany	Bavaria	2015	Mixture
87	Germany	Bavaria	2015	Mixture
88	Germany	Bavaria	2015	Mixture
89	Germany	Bavaria	2015	Mixture
90	Germany	Rhineland-Palatinate	2015	Mixture
91	Germany	Bayern	2015	Mixture
92	Italy	Campania	2014	Mortarella
93	Italy	Lazio	2014	Tonda Gentile Romana
94	Italy	Not specified	2014	Mortarella
95	Italy	Not specified	2014	Romane
96	Italy	Campania	2015	Mortarella
97	Italy	Campania	2015	Lange Neapler
98	Italy	Campania	2015	Camponica
99	Italy	Campania	2015	San Giovanni
100	Italy	Campania	2015	Mortarella
101	Italy	Campania	2015	Tonda di Giffoni

102	Italy	Campania	2015	Tonda Gentile Romana
103	Italy	Lazio	2015	Tonda Gentile Romana
104	Italy	Lazio	2015	Tonda Gentile Romana
105	Italy	Lazio	2015	Tonda Gentile Romana
106	Italy	Not specified	2015	Mixture
107	Italy	Not specified	2015	Mixture
108	Italy	Not specified	2015	Mixture
109	Italy	Not specified	2015	Tonda Gentile Romana
110	Italy	Piemonte	2015	Tonda Gentile Trilobata
111	Italy	Piemonte	2015	Tonda Gentile Trilobata
112	Italy	Piemonte	2015	Tonda Gentile Trilobata
113	Italy	Piemonte	2015	Tonda Gentile Trilobata
114	Italy	Piemonte	2015	Tonda Gentile Trilobata
115	Italy	Campania	2016	Tonda di Giffoni
116	Italy	Campania	2016	Tonda di Giffoni
117	Italy	Campania	2016	Camponica
118	Italy	Campania	2016	Mortarella
119	Italy	Campania	2016	Mortarella
120	Italy	Lazio	2016	Tonda Gentile Romana
121	Italy	Lazio	2016	Tonda Gentile Romana
122	Italy	Lazio	2016	Tonda Gentile Romana
123	Italy	Not specified	2016	Not specified
124	Italy	Not specified	2016	Not specified
125	Italy	Not specified	2016	Not specified
126	Italy	Not specified	2016	Not specified
127	Italy	Piemonte	2016	Tonda Gentile Trilobata
128	Italy	Piemonte	2016	Mortarella
129	Italy	Piemonte	2016	Tonda Gentile Trilobata
130	Italy	Piemonte	2016	Tonda Gentile Trilobata
131	Italy	Piemonte	2016	Tonda Gentile Trilobata
132	Italy	Piemonte	2016	Tonda Gentile Trilobata
133	Turkey	Ordu	2013	Mixture
134	Turkey	Ordu	2013	Mixture
135	Turkey	Ordu	2013	Mixture
136	Turkey	Ordu	2013	Mixture
137	Turkey	Ordu	2013	Mixture
138	Turkey	Düzce	2014	Mixture
139	Turkey	Giresun	2014	Mixture
140	Turkey	Not specified	2014	Mixture
141	Turkey	Not specified	2014	Mixture
142	Turkey	Ordu	2014	Mixture
143	Turkey	Ordu	2014	Mixture
144	Turkey	Ordu	2014	Mixture
145	Turkey	Ordu	2014	Mixture
146	Turkey	Ordu	2014	Tombul
147	Turkey	Ordu	2014	Mixture
148	Turkey	Samsun	2014	Tombul
149	Turkey	Samsun	2014	Mixture
150	Turkey	Samsun	2014	Tombul
151	Turkey	Samsun	2014	Tombul
152	Turkey	Düzce	2015	Çakildak
153	Turkey	Düzce	2015	Mixture
154	Turkey	Düzce	2015	Mixture
155	Turkey	Düzce	2015	Mixture
156	Turkey	Düzce	2015	Mixture
157	Turkey	Düzce	2015	Mixture

158	Turkey	Düzce	2015	Mixture
159	Turkey	Düzce	2015	Mixture
160	Turkey	Düzce	2015	Mixture
161	Turkey	Düzce	2015	Mixture
162	Turkey	Düzce	2015	Mixture
163	Turkey	Düzce	2015	Mixture
164	Turkey	Düzce	2015	Mixture
165	Turkey	Not specified	2015	Mixture
166	Turkey	Not specified	2015	Mixture
167	Turkey	Not specified	2015	Mixture
168	Turkey	Trabzon	2015	Mixture
169	Turkey	Düzce	2016	Tombul
170	Turkey	Düzce	2016	Tombul
171	Turkey	Düzce	2016	Tombul
172	Turkey	Not specified	2016	Not specified

Table S2: Ranking of the 100 most important features for origin classification of hazelnuts by means of the variable importance of Random Forest. Listed is the rank of the variable the m/z of the feature as well as the mean decrease in accuracy.

	m/z	Mean decrease in accuracy
1	431.3522	3.186
2	256.2747	2.547
3	123.0808	2.321
4	465.4090	2.288
5	144.0817	2.034
6	123.0809	1.802
7	417.3861	1.664
8	447.3478	1.440
9	329.3204	1.359
10	316.2122	1.294
11	465.4090	1.246
12	157.1230	1.203
13	638.6223	1.124
14	281.2580	1.089
15	241.2277	1.087
16	329.3330	1.056
17	282.2802	1.024
18	181.1588	1.013
19	329.3331	1.005
20	151.0765	0.975
21	638.5868	0.971
22	417.3991	0.939
23	638.6042	0.936
24	850.7784	0.922
25	638.5695	0.874
26	310.3109	0.834
27	256.2636	0.832
28	417.3699	0.810
29	621.5620	0.806
30	312.2524	0.770
31	543.4776	0.769
32	621.5447	0.744
33	544.4208	0.706
34	282.2918	0.698
35	269.2476	0.682
36	181.1596	0.679
37	225.1857	0.664
38	294.2420	0.647
39	612.5530	0.644
40	175.1126	0.638
41	463.3776	0.597
42	280.2662	0.593
43	175.1119	0.588
44	833.7549	0.581
45	329.3205	0.572
46	283.2636	0.568
47	447.4004	0.567
48	617.5309	0.546
49	265.2535	0.544

50	282.3034	0.527
51	374.3254	0.525
52	282.3150	0.505
53	850.8087	0.494
54	247.2419	0.488
55	612.5715	0.474
56	663.4520	0.463
57	269.2484	0.460
58	357.3268	0.459
59	294.2423	0.441
60	850.8188	0.436
61	417.3620	0.434
62	640.6000	0.422
63	111.1176	0.419
64	850.7986	0.411
65	256.2736	0.410
66	577.5175	0.405
67	185.1540	0.397
68	678.5643	0.396
69	256.2858	0.391
70	181.1731	0.390
71	297.2423	0.389
72	595.5282	0.380
73	612.5873	0.379
74	241.2278	0.378
75	463.3806	0.370
76	780.7088	0.369
77	283.2753	0.366
78	459.3017	0.362
79	417.3764	0.347
80	163.1118	0.345
81	416.3651	0.335
82	111.1172	0.334
83	313.2378	0.330
84	282.3441	0.329
85	312.2769	0.328
86	316.2230	0.322
87	227.1999	0.321
88	680.4793	0.320
89	615.4940	0.320
90	168.0661	0.317
91	167.1579	0.317
92	581.5146	0.303
93	876.8354	0.301
94	193.1237	0.291
95	640.6175	0.289
96	312.2647	0.286
97	850.7885	0.279
98	310.3231	0.276
99	647.4779	0.273
100	900.8118	0.270

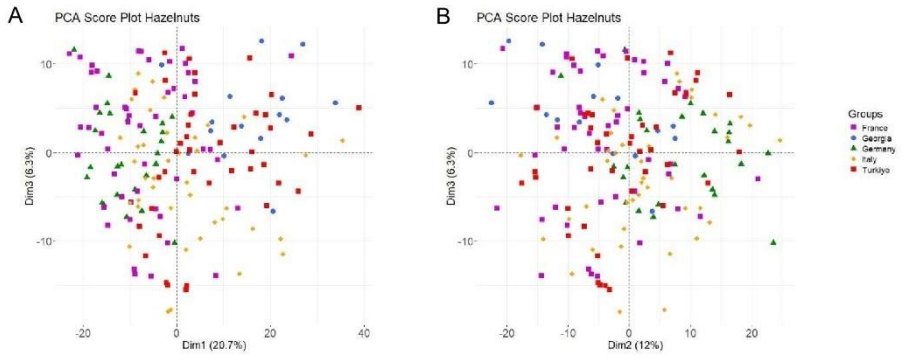


Figure S1: PCA of all hazelnut samples measured with DART-MS. PC1 vs PC3 (A) and PC2 vs. PC3 (B). with the samples colored according to the sample origin from France (purple). Georgia (blue). Germany (green). Italy (orange) and Türkiye (red).

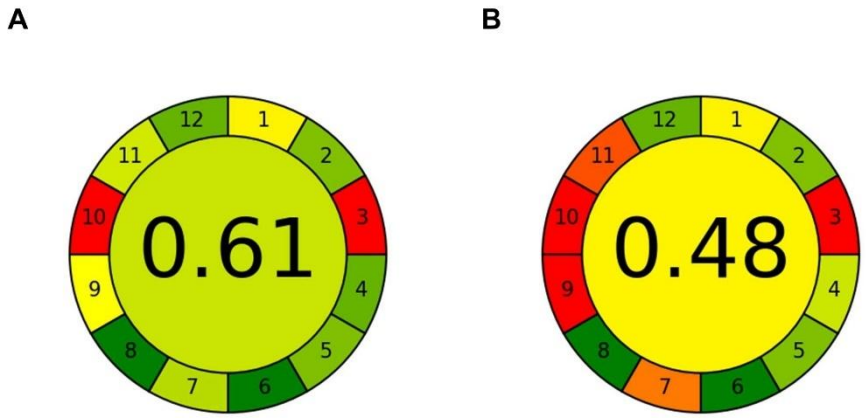


Figure S2: Ranking of the sustainability of a DART-MS measurement (A) compared to an LC-MS measurement (B). using the Analytical GREEnness metric approach and software (AGREE) tool.

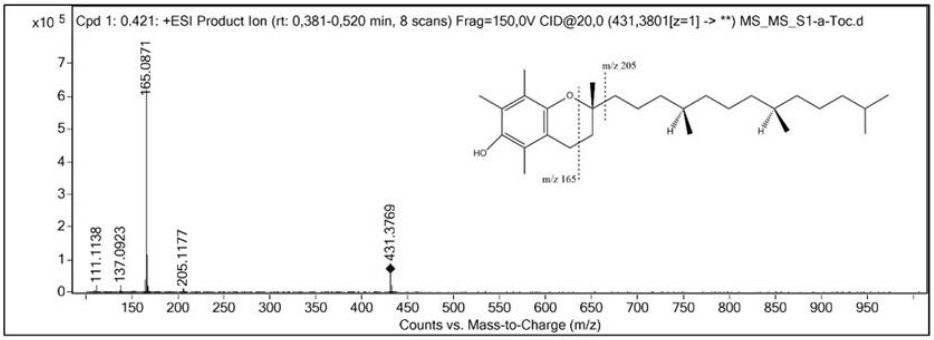


Figure S3: MS^2 spectrum of α-Tocopherol.

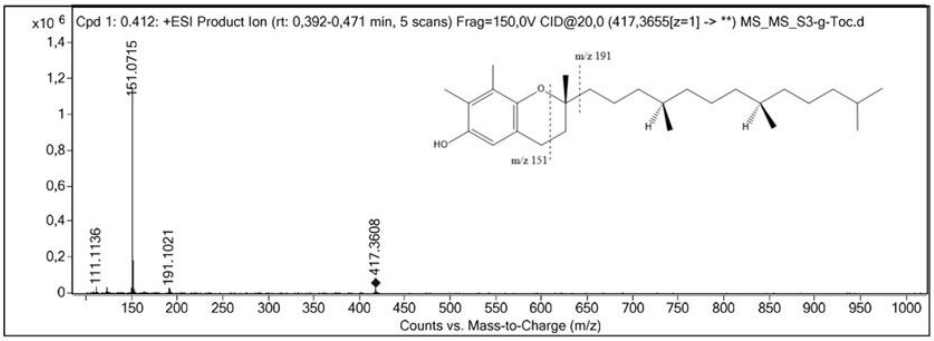


Figure S4: MS^2 spectrum of γ-Tocopherol.

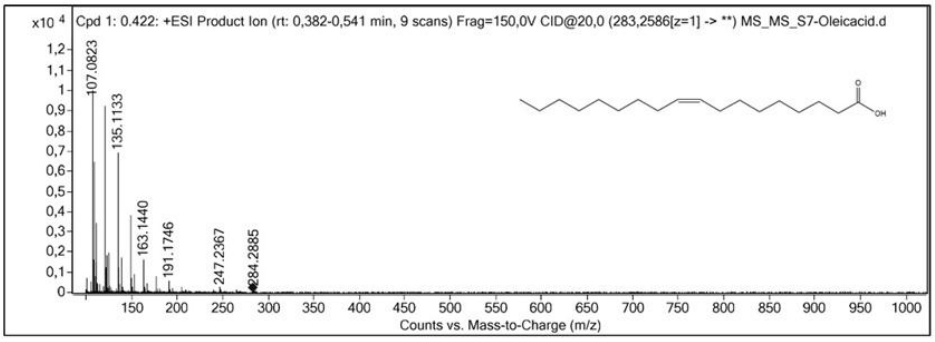


Figure S5: MS^2 spectrum of FA 18:1.

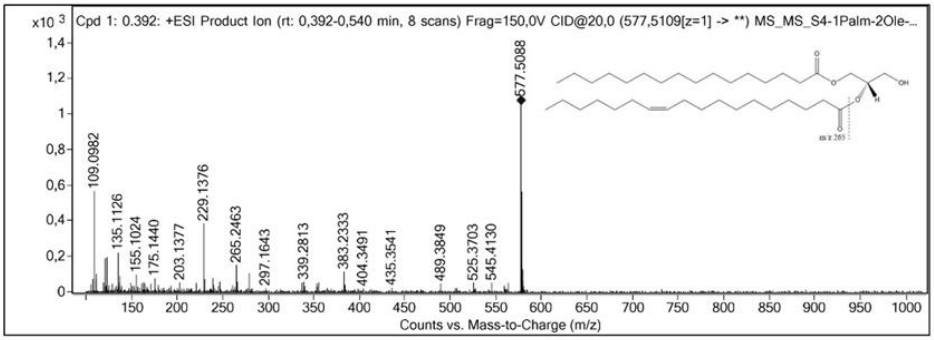


Figure S6: MS^2 spectrum of DG 34:1.

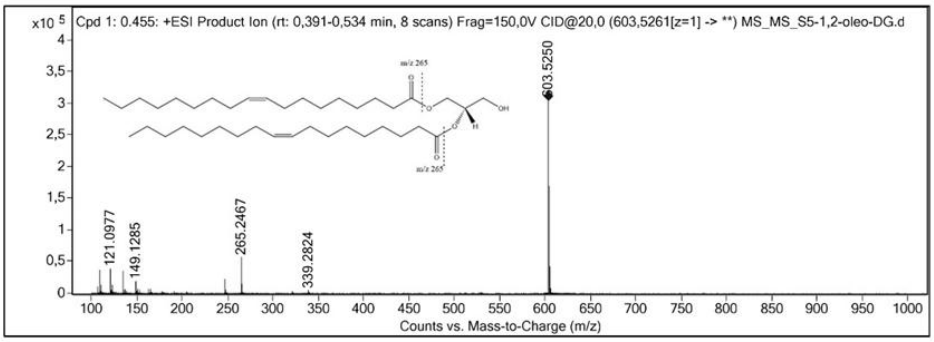


Figure S7: MS^2 spectrum of DG 36:2.

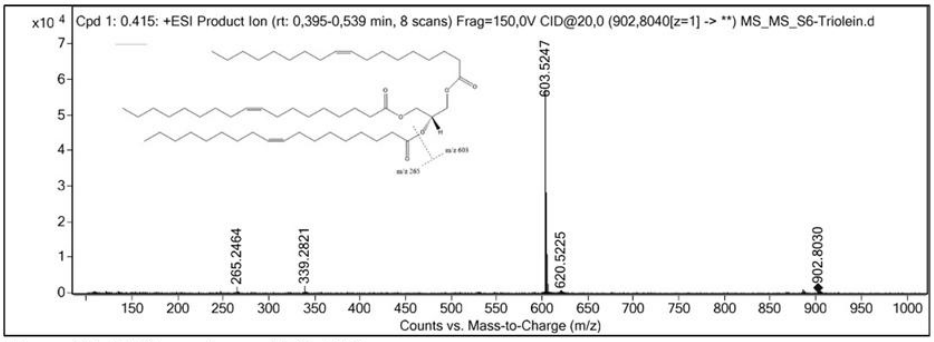


Figure S8: MS^2 spectrum of TG 54:3.

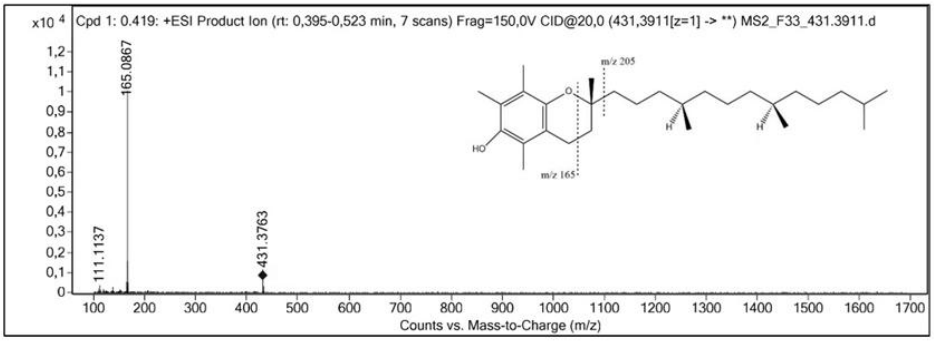


Figure S9: MS^2 spectrum of α-Tocopherol.

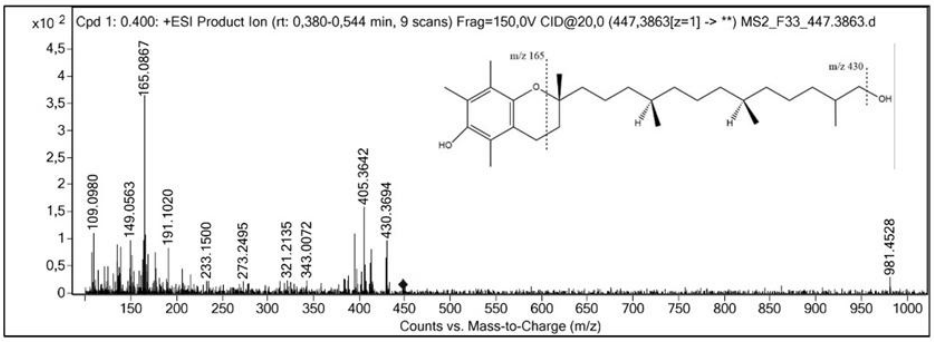


Figure S10: MS^2 spectrum of α-OH-Tocopherol.

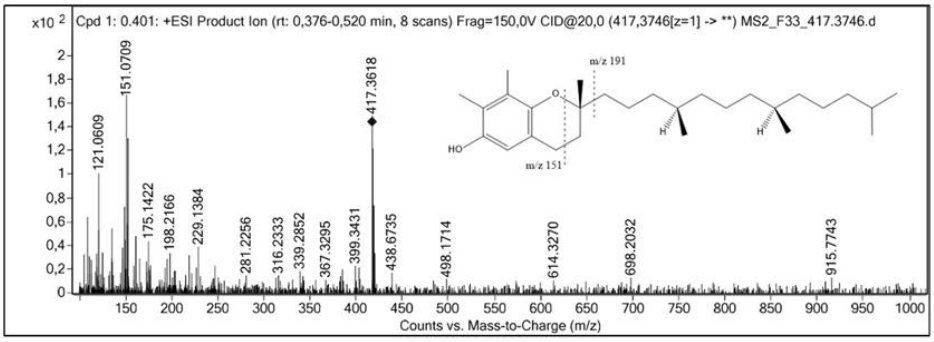


Figure S11: MS^2 spectrum of γ-Tocopherol.

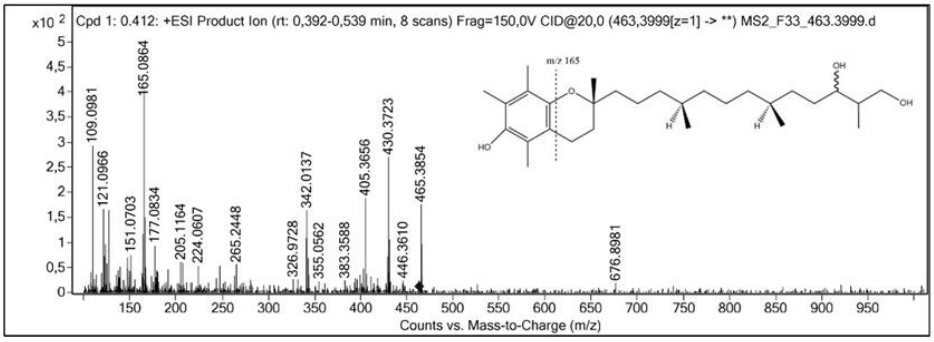


Figure S12: MS^2 spectrum of α-OH-OH-Tocopherol.

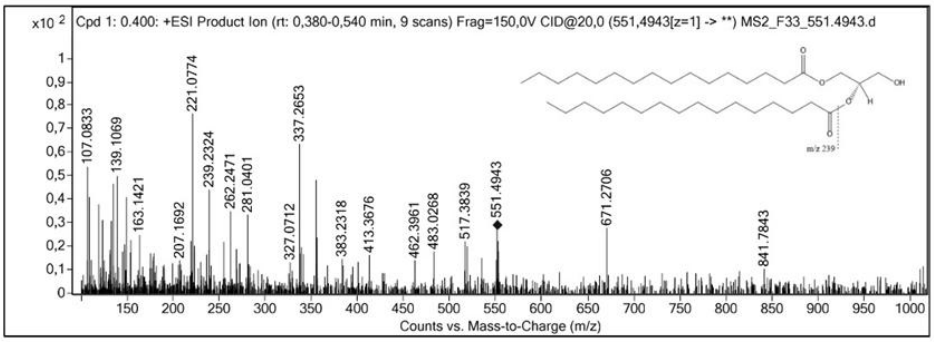


Figure S13: MS^2 spectrum of DG 32:0.

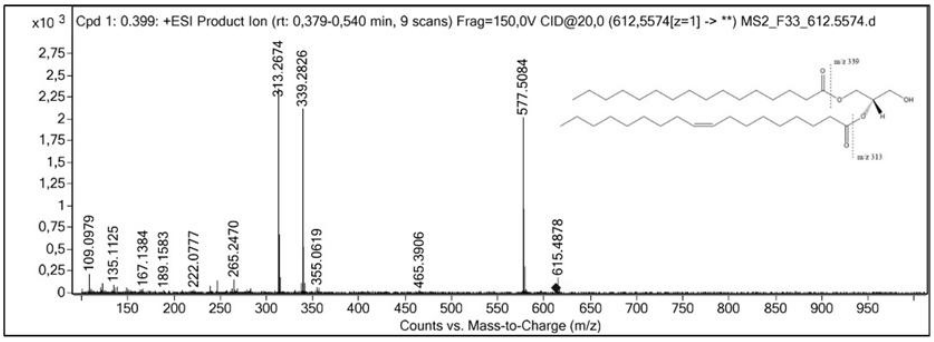


Figure S14: MS^2 spectrum of DG 34:1.

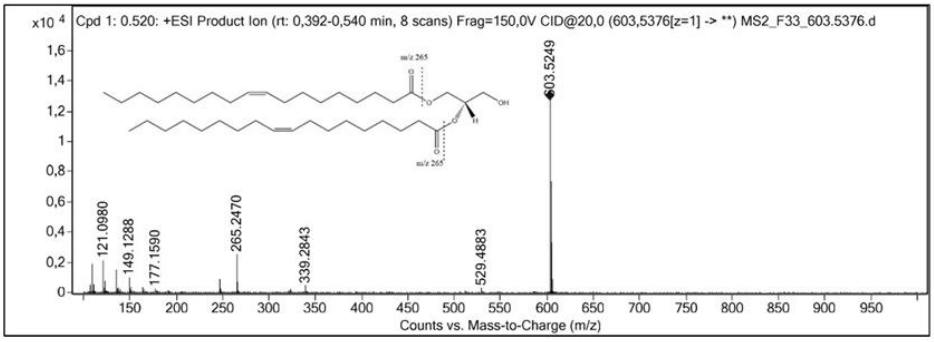


Figure S15: MS^2 spectrum of DG 36:2.

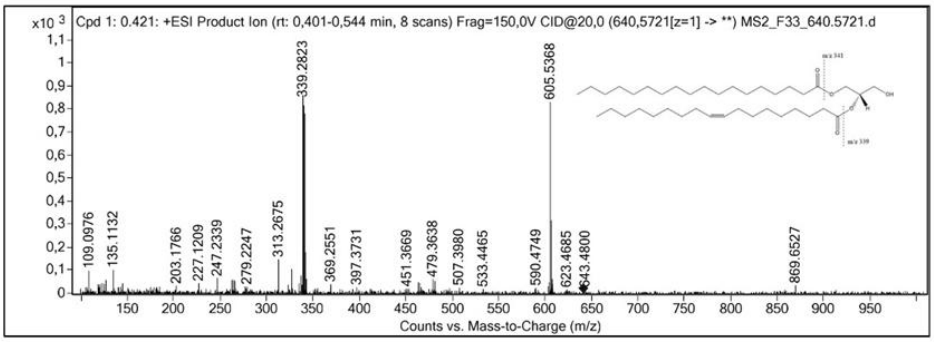


Figure S16: MS^2 spectrum of DG 36:1.

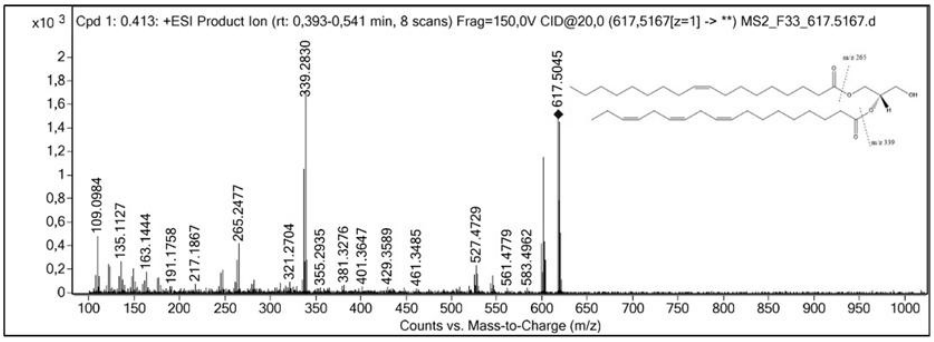


Figure S17: MS^2 spectrum of DG 36:4.

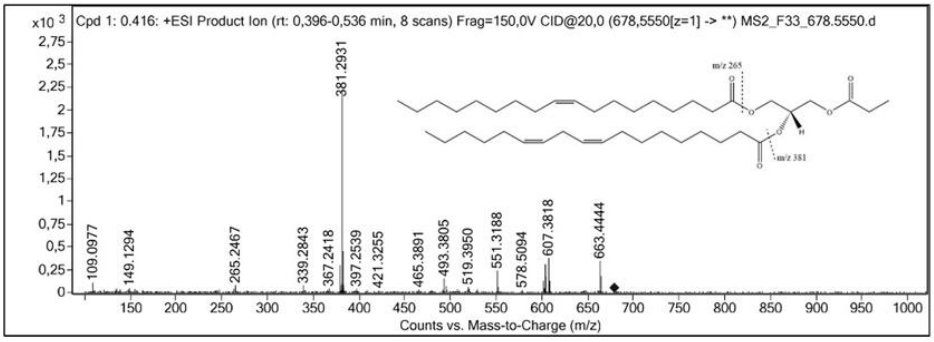


Figure S18: MS^2 spectrum of TG 38:3.

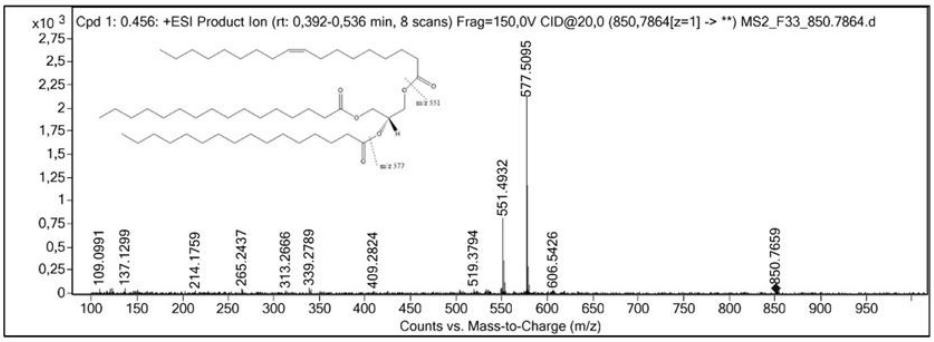


Figure S19: MS^2 spectrum of TG 50:1.

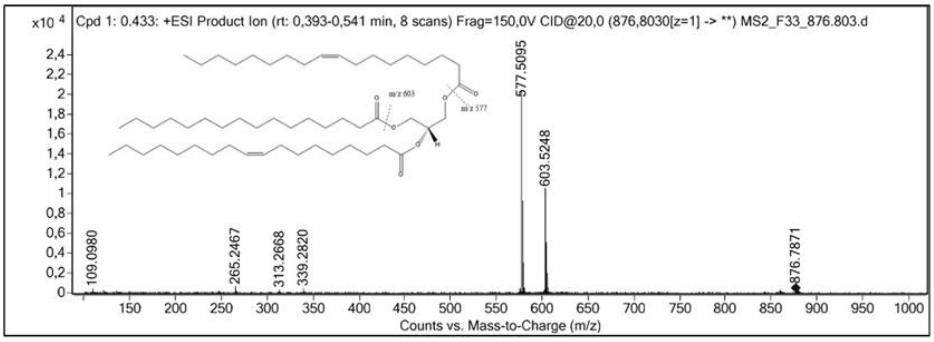


Figure S20: MS^2 spectrum of TG 52:2.

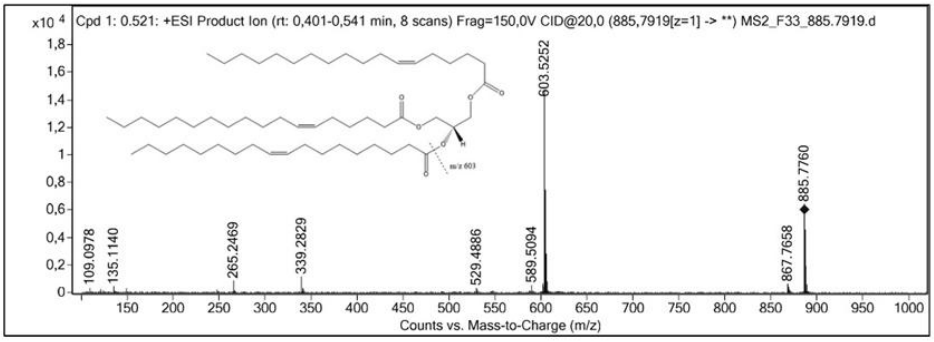


Figure S21: MS^2 spectrum of TG 54:3.

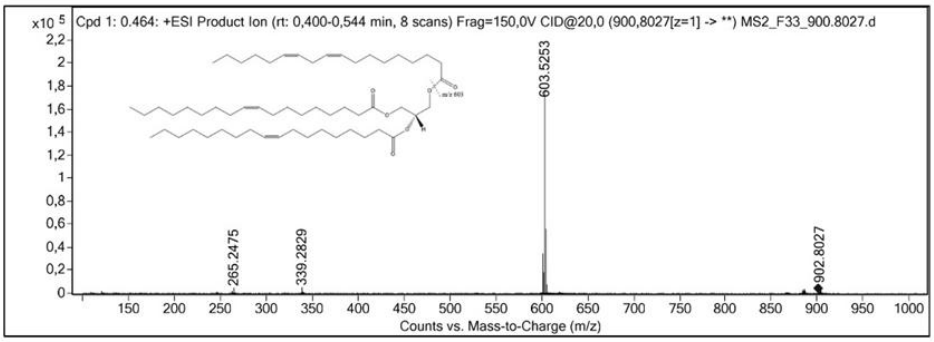


Figure S22: MS^2 spectrum of TG 54:4.

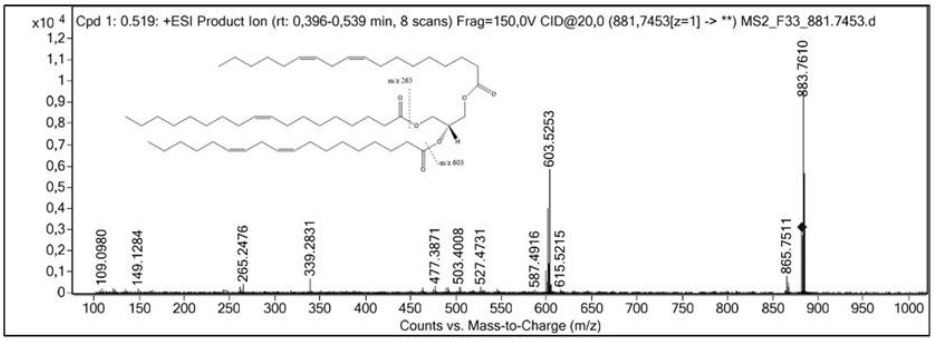


Figure S23: MS^2 spectrum of TG 54:5.

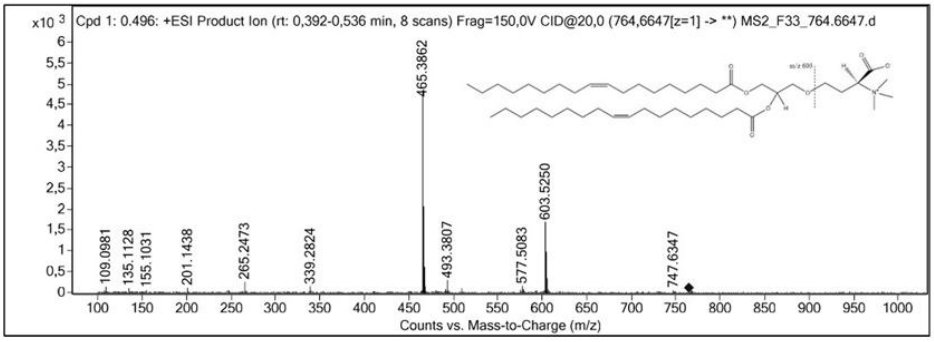


Figure A21: MS^2 spectrum of DGTS 36:2.

9.4 Supporting Information zu: Competing technologies: determining the geographical origin of strawberries (*Fragaria* × *ananassa*) using laboratory based near-infrared spectroscopy compared to a simple portable device

Supplementary Information (SI) for Molecular Omics.
This journal is © The Royal Society of Chemistry 2024

Electronic Supplementary Information

Competing Technologies: Determining the Geographical Origin of Strawberries (*Fragaria* × *ananassa*) using laboratory based Near-Infrared Spectroscopy compared to a simple portable device

Johannes Brockelt†, Felix Schmauder†, Kim Brettschneider, Marina Creydt, Stephan Seifert, Markus Fischer*

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

†Authors contributed equally to this work

*Corresponding author:

Markus Fischer, Hamburg School of Food Science - Institute of Food Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

Tel.: +49 40 42838-4359, Fax: +49 40 42838-4342

Email: markus.fischer@uni-hamburg.de

Table S1: All samples with their origin by country and state as well as their variety and year of cultivation, where available.

#	Origin	State	Region	Cultivation	Harvest Year	Harvest Month	Variety
1	Germany	Niedersachsen	North	Conventional	2021	July	Malwina
2	Germany	Niedersachsen	North	Conventional	2021	July	n.s.
3	Germany	Schleswig-Holstein	North	Conventional	2021	July	Malwina
4	Germany	Niedersachsen	North	Organic	2021	July	n.s.
5	Germany	Niedersachsen	North	Conventional	2021	July	Malwina
6	Germany	Niedersachsen	North	Conventional	2021	July	n.s.
7	Germany	Baden-Württemberg	South	Conventional	2021	July	n.s.
8	Germany	Baden-Württemberg	South	Conventional	2021	July	n.s.
9	Germany	Baden-Württemberg	South	Conventional	2021	July	Malwina
12	Germany	Schleswig-Holstein	North	Conventional	2021	July	n.s.
13	Germany	Niedersachsen	North	Conventional	2021	July	n.s.
15	Spain	Andalusia	n.s.	Conventional	2022	Februray	n.s.
16	Spain	Andalusia	n.s.	Conventional	2022	Februray	n.s.
20	Spain	Valencia	n.s.	Conventional	2022	Februray	n.s.
23	Spain	Andalusia	n.s.	Conventional	2022	Februray	Fortuna
26	Spain	Andalusia	n.s.	Organic	2022	Februray	n.s.
27	Spain	Andalusia	n.s.	Conventional	2022	Februray	n.s.
28	Spain	Andalusia	n.s.	Organic	2022	Februray	n.s.
31	Spain	Andalusia	n.s.	Organic	2022	April	n.s.
36	Germany	Rheinland-Pfalz	South	Conventional	2022	May	Clery
38	Germany	Rheinland-Pfalz	South	Conventional	2022	May	Malling Centenary
42	Germany	Rheinland-Pfalz	South	Organic	2022	May	Clery
49	Germany	Hessen	West	Conventional	2022	May	Clery
50	Germany	Hessen	West	Conventional	2022	May	Flair
51	Germany	Hessen	West	Conventional	2022	May	Malling Centenary
52	Germany	Hessen	West	Conventional	2022	May	Clery
53	Germany	Hessen	West	Conventional	2022	May	Lady Emma
55	Germany	Hessen	West	Conventional	2022	May	n.s.
56	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Magnum
57	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Flair
58	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Alegro
61	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Lambada

63	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Malling Centenary
69	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Hademar
70	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Florentina
71	Germany	Nordrhein-Westfalen	West	Conventional	2022	May	Magnum
94	Germany	Baden-Württemberg	South	Organic	2022	May	Sibilla
95	Germany	Baden-Württemberg	South	Organic	2022	June	Clery
103	Germany	Bayern	South	Organic	2022	June	Clery
124	Germany	Sachsen	East	Conventional	2022	June	n.s.
132	Germany	Brandenburg	East	Conventional	2022	June	n.s.
135	Germany	Sachsen-Anhalt	East	Conventional	2022	June	Florentina
136	Germany	Sachsen-Anhalt	East	Conventional	2022	June	n.s.
143	Germany	Brandenburg	East	Conventional	2022	June	Asia
149	Germany	Schleswig-Holstein	North	Conventional	2022	June	Magnum
150	Germany	Schleswig-Holstein	North	Conventional	2022	June	Sonsation
153	Germany	Schleswig-Holstein	North	Organic	2022	June	Korona
154	Germany	Schleswig-Holstein	North	Organic	2022	June	Malling Allure
155	Germany	Schleswig-Holstein	North	Conventional	2022	June	Florence
159	Germany	Schleswig-Holstein	North	Conventional	2022	June	Magnus
164	Germany	Hamburg	North	Conventional	2022	June	Malvina
166	Germany	Niedersachsen	North	Conventional	2022	June	Polka
177	Germany	Niedersachsen	North	Conventional	2022	June	Malvina
184	Greece	n.s.	n.s.	Conventional	2022	December	n.s.
185	Greece	West Greece	n.s.	Conventional	2023	January	n.s.
189	Spain	Andalusia	n.s.	Conventional	2023	January	n.s.
197	Spain	Andalusia	n.s.	Conventional	2023	January	Felicity
199	Egypt	n.s.	n.s.	Conventional	2023	January	n.s.
209	Egypt	n.s.	n.s.	Conventional	2023	January	Fortuna
212	Spain	n.s.	n.s.	Conventional	2023	Februray	Inspire
217	Spain	Andalusia	n.s.	Organic	2023	Februray	n.s.
224	Spain	Andalusia	n.s.	Organic	2023	March	Rociera
225	Spain	Andalusia	n.s.	Conventional	2023	March	Calinda
226	Spain	Andalusia	n.s.	Conventional	2023	March	n.s.
227	Spain	Andalusia	n.s.	Conventional	2023	March	n.s.
228	Greece	West Greece	n.s.	Conventional	2023	March	Victory
234	Greece	West Greece	n.s.	Conventional	2023	March	n.s.
236	Spain	Andalusia	n.s.	Conventional	2023	March	n.s.
237	Spain	Andalusia	n.s.	Conventional	2023	March	Calinda

238	Greece	West Greece	n.s.	Conventional	2023	March	n.s.
240	Spain	Andalusia	n.s.	Conventional	2023	March	n.s.
247	Spain	Andalusia	n.s.	Organic	2023	May	n.s.
252	Egypt	n.s.	n.s.	Conventional	2023	May	n.s.
265	Germany	Hessen	West	Conventional	2023	June	n.s.
267	Germany	Hessen	West	Organic	2023	June	Alegro
268	Germany	Hessen	West	Conventional	2023	June	Sonata
269	Germany	Hessen	West	Conventional	2023	June	Verdi
276	Germany	Nordrhein-Westfalen	West	Conventional	2023	June	Florentina
277	Germany	Nordrhein-Westfalen	West	Conventional	2023	June	n.s.
309	Germany	Baden-Württemberg	South	Conventional	2023	June	Faith
314	Germany	Baden-Württemberg	South	Organic	2023	June	Sonata
319	Germany	Bayern	South	Conventional	2023	June	Verdi
347	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
348	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
349	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
350	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
352	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
353	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
354	Egypt	n.s.	n.s.	Conventional	2023	December	n.s.
362	Spain	Andalusia	n.s.	Conventional	2024	Januray	n.s.
364	Spain	Andalusia	n.s.	Conventional	2024	Februray	Felicity
365	Spain	Andalusia	n.s.	Conventional	2024	Februray	n.s.
367	Spain	Andalusia	n.s.	Conventional	2024	Februray	n.s.
368	Greece	West Greece	n.s.	Conventional	2024	Februray	Fortuna
369	Spain	Andalusia	n.s.	Organic	2024	Februray	Rabida
372	Egypt	n.s.	n.s.	Conventional	2024	Februray	n.s.

Table S2: Water content [%] of intact, homogenized as well as homogenized and freeze-dried strawberries, measured using a dry scale.

Sample	Intact water content [%]	Homogenized water content [%]	Freeze-dried water content [%]
1	88.98	90.72	4.28
2	90.89	89.34	3.36
3	91.90	90.68	2.48
4	89.77	91.25	4.11
5	90.90	92.13	2.87
6	90.81	90.23	3.46
7	90.64	91.32	2.59
8	90.88	90.92	3.31

9	91.70	88.98	4.35
10	90.06	91.38	2.08
11	89.82	90.56	2.39
12	91.03	91.14	1.91
Mean	90.62	90.67	3.10

Table S3: ANOVA results for the data from the benchtop device.

Number	Feature [cm^{-1}]	p-Values
1	4891.8360000000002	6,31E-15
2	4881.6319999999996	4,95E-14
3	4361.2240000000002	3,13E-10
4	6902.0410000000002	5,45E-10
5	6912.2449999999999	5,70E-10
6	6891.8360000000002	6,28E-10
7	6881.6319999999996	6,68E-10
8	4371.4279999999999	7,45E-10
9	6871.4279999999999	8,26E-10
10	6922.4489999999996	8,46E-10
11	6861.2240000000002	8,64E-10
12	4351.0200000000004	8,75E-10
13	4902.0410000000002	1,02E-09
14	6932.6530000000002	1,42E-09
15	6851.02	1,60E-09
16	6840.8159999999998	1,93E-09
17	6942.857	2,05E-09
18	6830.6120000000001	3,02E-09
19	6953.0609999999997	3,60E-09
20	4340.8159999999998	3,84E-09
21	6820.4080000000004	5,36E-09
22	4381.6319999999996	5,60E-09
23	6963.2650000000003	8,07E-09
24	6810.2039999999997	8,34E-09
25	4330.6120000000001	9,79E-09
26	6973.4690000000001	1,47E-08
27	4320.4080000000004	1,54E-08
28	6800	1,60E-08
29	4310.2039999999997	2,47E-08
30	6789.7960000000003	3,17E-08
31	4391.8360000000002	3,31E-08
32	4300	3,54E-08
33	6983.6729999999998	4,15E-08
34	6779.5919999999996	5,52E-08
35	4289.7960000000003	8,11E-08
36	6769.3879999999999	1,07E-07

37	6993.8770000000004	1,14E-07
38	6759.1840000000002	2,33E-07
39	4402.0410000000002	2,56E-07
40	7004.0820000000003	2,95E-07
41	4279.5919999999996	3,23E-07
42	6748.9790000000003	4,76E-07
43	4208.1629999999996	8,85E-07
44	7014.2860000000001	9,05E-07
45	6738.7749999999996	1,01E-06
46	4218.3670000000002	1,06E-06
47	4197.9589999999998	1,08E-06
48	4269.3879999999999	1,10E-06
49	4228.5709999999999	1,37E-06
50	4187.7550000000001	1,72E-06
51	4238.7749999999996	1,88E-06
52	6728.5709999999999	2,25E-06
53	4412.2449999999999	2,31E-06
54	4248.9790000000003	2,85E-06
55	4259.1840000000002	3,09E-06
56	4871.4279999999999	3,20E-06
57	7024.49	3,57E-06
58	4177.5510000000004	3,85E-06
59	6718.3670000000002	4,89E-06
60	7034.6940000000004	8,97E-06
61	6708.1629999999996	1,02E-05
62	4422.4489999999996	1,28E-05
63	4167.3469999999998	1,29E-05
64	6697.9589999999998	2,28E-05
65	7044.8969999999999	3,70E-05
66	6687.7550000000001	4,58E-05
67	4432.6530000000002	4,73E-05
68	4912.2449999999999	5,16E-05
69	4157.143	5,70E-05
70	6677.5510000000004	8,96E-05
71	7055.1019999999999	1,04E-04
72	4442.857	1,12E-04
73	4453.0609999999997	1,51E-04
74	4463.2650000000003	1,62E-04
75	4473.4690000000001	1,80E-04
76	6667.3469999999998	1,95E-04
77	4483.6729999999998	2,20E-04
78	4493.8770000000004	2,93E-04
79	7065.3059999999996	3,69E-04
80	6657.143	4,07E-04
81	4504.0820000000003	4,41E-04

82	4146.9380000000001	6,52E-04
83	6646.9380000000001	9,06E-04
84	7075.51	9,22E-04
85	4514.2860000000001	1,12E-03
86	6636.7340000000004	1,90E-03
87	7085.7139999999999	2,29E-03
88	4524.49	2,71E-03
89	6626.53	3,79E-03
90	7095.9179999999997	7,18E-03
91	4534.6940000000004	7,66E-03
92	6616.326	8,25E-03
93	4136.7340000000004	1,27E-02
94	6606.1220000000003	1,59E-02
95	7106.1220000000003	1,67E-02
96	6595.9179999999997	3,40E-02
97	4544.8980000000001	3,93E-02
98	7116.326	4,61E-02
99	6585.7139999999999	6,35E-02
100	4922.4489999999996	6,80E-02
101	7126.53	9,13E-02
102	6575.51	1,04E-01
103	6565.3059999999996	1,52E-01
104	4555.1019999999999	2,15E-01
105	4126.53	2,28E-01
106	5177.5510000000004	2,62E-01
107	5167.3469999999998	2,72E-01
108	6555.1019999999999	2,74E-01
109	5187.7550000000001	3,04E-01
110	5157.143	3,60E-01
111	5197.9589999999998	3,91E-01
112	7136.7340000000004	3,95E-01
113	6544.8980000000001	3,98E-01
114	5146.9380000000001	4,89E-01
115	6534.6940000000004	5,53E-01
116	7146.9380000000001	6,27E-01
117	5208.1629999999996	6,62E-01
118	5136.7340000000004	7,66E-01
119	4861.2240000000002	8,28E-01
120	6524.49	8,36E-01
121	4565.3059999999996	1,04E+00
122	6514.2860000000001	1,11E+00
123	5218.3670000000002	1,26E+00
124	7157.143	1,37E+00
125	5126.53	1,39E+00
126	6504.0820000000003	1,54E+00

127	6493.8770000000004	2,20E+00
128	5116.326	2,61E+00
129	6483.6729999999998	2,89E+00
130	4116.326	3,30E+00
131	5228.5709999999999	3,58E+00
132	6473.4690000000001	3,86E+00
133	7167.3469999999998	4,28E+00
134	5106.1220000000003	5,49E+00
135	6463.2650000000003	5,56E+00
136	4932.6530000000002	5,86E+00
137	6453.0609999999997	7,69E+00
138	4575.51	8,28E+00
139	7177.5510000000004	8,59E+00
140	11075.51	1,16E+01
141	6442.857	1,21E+01
142	5238.7749999999996	1,28E+01
143	5095.9179999999997	1,35E+01
144	7187.7550000000001	1,38E+01
145	6432.6530000000002	1,54E+01
146	11004.08	1,74E+01
147	11095.92	1,78E+01
148	8708.1630000000005	1,82E+01
149	8718.3670000000002	2,02E+01
150	8687.7549999999992	2,37E+01

Table S4: ANOVA results for the data of the handheld device.

Number	Feature	p-Values
1	6734	3,95E-26
2	5938	2,00E-16
3	6046	3,84E-16
4	6882	1,52E-11
5	6579	2,84E-08
6	6180	3,88E-07
7	6301	1,67E-02
8	7052	5,05E+00
9	8846	2,46E+01
10	8217	1,12E+02
11	7210	8,51E+02
12	7391	1,74E+04
13	6447	3,55E+04
14	7794	2,75E+05
15	7994	3,21E+05
16	7576	5,09E+05

Table S5: Results for the test data sets (two class model) for the benchtop and the handheld NIR.

			Response class		
			Germany	Other origin	Number of samples
True class	Benchtop NIR	Germany	11	0	11
		Other origin	1	8	9
	Handheld NIR	Germany	11	0	11
		Other origin	3	6	9

33 Table S6: Results for the test data sets (4 class model) for the benchtop and the handheld
34 NIR.

				Response class			
A			Egypt	Germany	Greece	Spain	Number of Samples
Benchtop NIR	True class	Egypt	1	1	0	0	2
		Germany	0	11	0	0	11
		Greece	0	0	1	1	2
		Spain	0	0	2	3	5
B							
Handheld NIR	True class	Egypt	2	0	0	0	2
		Germany	2	9	0	0	11
		Greece	0	0	2	0	2
		Spain	0	2	2	1	5

35

36

37



Figure S1: Image of the benchtop NIR used (left) and the handheld NIR used.

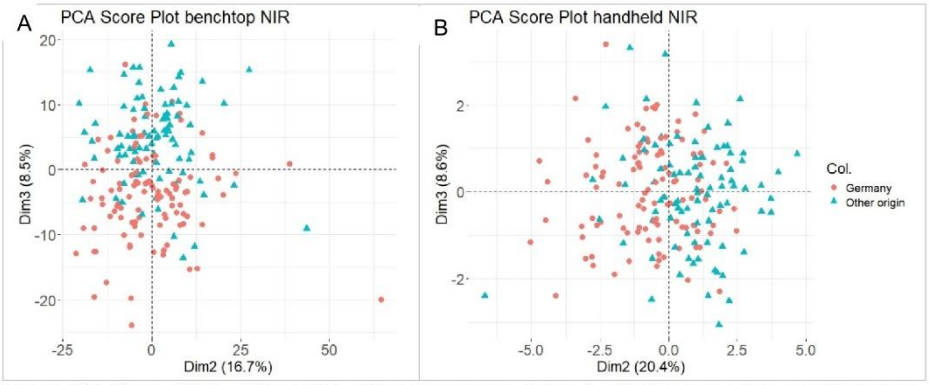


Figure S2: Score Plots of the PCA of the strawberry dataset with the two origins of Germany and the totality of all other samples. Showed are the PC 2 vs. 3 of the PCAs measured with the benchtop (A) and handheld NIR (B) device.

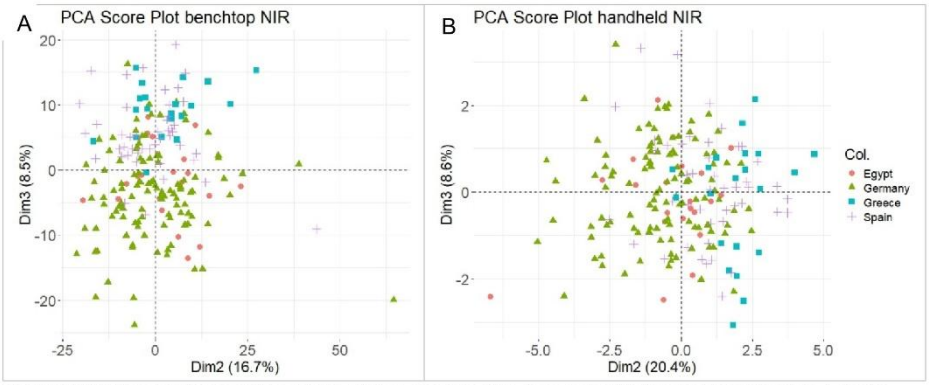


Figure S3: Score Plots of the PCA of the strawberry dataset with the four countries Egypt, Germany, Greece and Spain. Showed are the PC 2 vs. 3 of the PCAs measured with the benchtop (A) and handheld NIR (B) device.

9.5 Verwendete Gefahrstoffe nach GHS

Chemikalie	Piktogramm/Signalwort	H-Sätze	P-Sätze
Aceton	GHS 02 GHS 07 Gefahr	225, 319, 336	210, 233, 240, 241, 242, 305+351+338
Acetonitril	GHS 02 GHS 07 Gefahr	225, 302+312+332, 319	210, 280, 301+312, 303+361+353, 304+340+312, 305+351+338
Ameisensäure	GHS 02 GHS 05 GHS 06 Gefahr	226, 302, 331, 314	210, 280, 303+361+353, 304+340+310, 305+351+338
Ammoniumformiat	Kein gefährlicher Stoff nach GHS		
Butylhydroxytoluol	GHS 09	410	273, 391, 501
Chloroform	GHS 06 GHS 08 Gefahr	302, 331, 315, 319, 351, 361d, 336, 372, 412	201, 273, 301+312+330, 302+352, 304+340+311, 308+313
Diacylglycerid 32:0	Kein gefährlicher Stoff nach GHS		
Diacylglycerid 34:1	Kein gefährlicher Stoff nach GHS		
Diacylglycerid 36:2	Kein gefährlicher Stoff nach GHS		
Ethanol	GHS 02 GHS 07 Gefahr	225, 319	210, 233, 240, 241, 242, 305+351+338
Flüssiger Stickstoff	GHS 04 Achtung	281	282, 336+315, 403
Helium	GHS 04	280	403
Isopropanol	GHS 02 GHS 07 Gefahr	225, 319, 336	210, 240, 305+351+338, 403+233
Methanol	GHS 02 GHS 06 GHS 08 Gefahr	225, 301+311+331, 370	210, 233, 280, 301+310, 303+361+353, 304+340+311
Monoacylglycerid 18:1	Kein gefährlicher Stoff nach GHS		
Natriumcarbonat	GHS 07 Achtung	319	264, 280, 305+351+338, 337+313
Natriumchlorid	Kein gefährlicher Stoff nach GHS		
n-Hexan	GHS 02 GHS 07 GHS 08 GHS 09 Gefahr	225, 304, 315, 336, 361f, 373, 411	201, 210, 273, 301+330+331, 302+352, 308+313
Ölsäure	Kein gefährlicher Stoff nach GHS		
Phosphocholin 36:2	Kein gefährlicher Stoff nach GHS		

Chemikalie	Piktogramm/Signalwort	H-Sätze	P-Sätze
Phosphatidylethanolamin 36:0	Kein gefährlicher Stoff nach GHS		
Tocopherole	GHS 07	317	261, 272, 280, 302+352, 333+313, 362+364
Triacylglycerid 48:0	Kein gefährlicher Stoff nach GHS		
Triacylglycerid 54:3	Kein gefährlicher Stoff nach GHS		
Triacylglycerid 54:6	Kein gefährlicher Stoff nach GHS		
Trockeneis	Kein gefährlicher Stoff nach GHS		
Wasser	Kein gefährlicher Stoff nach GHS		

9.6 Zeitschriftenartikel: Food Authentication goes Green: Geographische Herkunftsbestimmung von Körnermais (*Zea mays* L.) mittels Direct Analysis in Real Time-Massenspektrometrie

DEUTSCHE LEBENSMITTEL-RUNDSCHAU
»DLR
120. Jahrgang Sep. 2024 Behr's Verlag I Hamburg I ZKZ 9982

Angewandte Wissenschaft » Originalarbeiten exklusiv für Sie vorgestellt

Food Authentication goes Green: Geographische Herkunftsbestimmung von Körnermais (*Zea mays* L.) mittels *Direct Analysis in Real Time*-Massenspektrometrie

Felix Schmauder, Marina Creydt, Markus Fischer
Hamburg School of Food Science – Institut für Lebensmittelchemie, Universität Hamburg, Grindelallee 117, 20146 Hamburg

Einleitung

In den letzten Jahren ist das Interesse an umweltfreundlichen, schnellen und kosteneffizienten Methoden zur Überprüfung der Authentizität von Lebens- und Futtermitteln stark gestiegen. Dies erfordert die Entwicklung neuer Ansätze, Verfahren und analytischer Plattformen, um diese Ziele zu erreichen. Diesen Anforderungen steht jedoch die Tatsache gegenüber, dass oft hochauflösende Methoden eingesetzt werden müssen, um verfälschte Waren zuverlässig zu erkennen, da die Fälschungen immer raffinierter werden. In diesem Zusammenhang haben insbesondere *non-targeted Metabolomics*-Strategien an Bedeutung gewonnen, welche darauf basieren, dass die Konzentrationen von Metaboliten in Organismen von verschiedenen standortbedingten exogenen und endogenen Faktoren beeinflusst werden. So kann beispielsweise auf die Herkunft von pflanzlichen Lebensmitteln zurückgeschlossen werden, weil klimatische Bedingungen oder die Zusammensetzung der Böden einen Einfluss auf bestimmte Metabolite aufweisen können. Um die entsprechenden Konzentrationsunterschiede detektieren zu können, werden üblicherweise Kernresonanzspektrometer oder Massenspektrometer, welche mit flüssigkeits- oder gaschromatographischen Anlagen gekoppelt sind, verwendet. Auf diese Weise wird gewährleistet, dass möglichst viele Metabolite (kleine organische Moleküle mit einem Molekulargewicht < 1 500 g/mol) detektiert werden und ein möglichst aussagekräftiges chemisches Profil erhalten wird.

Mittels hochauflösender massenspektrometrischer Analysen, welche oftmals über Elektronenspray-Ionisation-Quellen mit flüssigkeitschromatographischen Methoden gekoppelt werden, können abhängig vom Organismus bis zu 5 000 Metabolite erfasst werden. Allerdings müssen bei dieser Vorgehensweise häufig recht hohe Mengen an organischen Lösungsmitteln zum Einsatz kommen und der Zeitaufwand beträgt üblicherweise 10 bis 30 Minuten. Beides Parameter, die sich negativ auf die Umweltfreundlichkeit und die Wirtschaftlichkeit auswirken. Eine neuere Technologie, um schnellere, nachhaltigere und somit auch günstigere Analysen zu ermöglichen, stellt die Verwendung von *Direct Analysis in Real Time*-Massenspektrometrie (DART-MS) dar. DART-Ionenquellen bieten die Möglichkeit, Proben lösungsmittelfrei zu ionisieren [1]. Des Weiteren gewährleisten DART-MS-Geräte relativ niedrige Nachweisgrenzen und die Möglichkeit der Strukturaufklärung relevanter Verbindungen [2].

In der hier vorgestellten Studie wurde ein *non-targeted Metabolomics* Ansatz gewählt, um die Eignung von DART-MS zur Herkunftsbestimmung zu untersuchen. Als Datensatz diente ein bisher bereits mit Flüssigkeitschromatographie-Massenspektrometrie (LC-ESI-MS) vermessener Datensatz von 200 Körnermaisproben aus sieben verschiedenen Ländern [3]. Die Proben stammten aus Spanien, Frankreich, Deutschland, Peru, der Ukraine, Ungarn und den USA, welche den Großteil der Körnermais produzierenden Länder abdecken. Körnermais ist eine der wichtigsten Feldfrüchte und leistet einen bedeutenden Beitrag zur globalen Ernährungssicherheit [4]. Als wichtiges Futtermittel ist Körnermais besonders anfällig für betrügerische Ursprungserklärungen. Der Ort des Anbaus beeinflusst nicht nur die ernährungsphysiologische Qualität des Lebens- oder Futtermittels, sondern auch den Risikofaktor für die Kontamination mit Mykotoxinen [5].

Ein weiteres Ziel war die Anwendung von Open-Source-Softwaretools für die Datenanalyse, um einen Arbeitsablauf für die Datenauswertung aufzusetzen, der für potenzielle Endnutzer frei zugänglich ist. Für die Klassifizierung der Daten wurde der *Random Forest* Algorithmus, ein Verfahren, welches auf ma-

schinellem Lernen basiert, verwendet, da dieses gut geeignet für hochdimensionale Datensätze ist. Insbesondere vorteilhaft ist dieser Algorithmus, wenn sich Daten aus einer großen Anzahl von Variablen (Metaboliten) und einer vergleichsweise geringen Anzahl von Proben zusammensetzen [6]. Anschließend wurden die relevantesten Metaboliten, die am geeignetsten für eine Herkunftsunterscheidung sind, mittels der Variablenwichtigkeit bestimmt und identifiziert, um die Unterschiede der Klassen (Herkünfte) zu charakterisieren. Zudem sollte ein Eindruck von den Fragmentierungsmustern, welche mittels DART-MS entstehen, erhalten werden, um die relevantesten Metabolite in ihrer Struktur aufklären zu können. Neben Fragmentierungsprozessen, die bereits während des Ionisierungsprozesses ablaufen, wurden zudem Fragmentierungen mittels MS/MS-Experimenten untersucht. Für den Leistungsvergleich von DART-MS mit dem LC-ESI-MS Ansatz wurden nachhaltige Aspekte, Klassifizierungsergebnisse, Kosten und die identifizierten Markerverbindungen verglichen.

Arbeitsablauf *Metabolomics* Analyse, Datenakquise und Auswertung

Da der vorliegende Datensatz bereits mittels LC-ESI-MS vermessen wurde, konnte der entwickelte Arbeitsablauf (siehe Abbildung 1) auf die Analyse mittels DART-MS übertragen werden und so eine gute Vergleichbarkeit der Analysen gewährleistet werden.

Der erste Schritt, die Probenakquise, hat zum Ziel, möglichst viele authentische Referenzproben zu erhalten. Dabei wurden die Körnermaisproben aus der bereits erwähnten LC-ESI-MS Publikation von Schütz et al. verwendet und durch weitere Proben aus Deutschland ergänzt [3]. Insgesamt wurden 200 Proben zur Analyse eingesetzt, von diesen stammten 26 Proben aus Deutschland, 37 aus Frankreich, 50 aus Ungarn, 20 aus Spanien, 25 aus Peru, 24 aus den Vereinigten Staaten und 18 Proben aus der Ukraine. Die Homogenisierung der Proben erfolgte mittels einer Messermühle unter Verwendung von Trockeneis, um die Proben vor Prozesswärme zu schützen. Anschließend wurden die homogenisierten Proben gefriergetrocknet. Auf diese Weise wurde ein Pulver erhalten, welches über längere Zeiträume gelagert werden kann. Da sich in der vorangegangenen Studie zeigte, dass insbesondere die lipophilen Metabolite (bspw. Fettsäuen, Glyceride oder Phospholipide) geeignet sind, um die unterschiedlichen Herkünfte voneinander zu unterscheiden, standen diese ebenfalls im Mittelpunkt bei den DART-MS-Analysen. Die Extraktion erfolgte mittels einer Zweiphasenextraktion nach Bligh and Dyer mit Wasser, Methanol und Chloroform [8]. Die Detektion der Verbindungen wurde mit einer DART-Ionenquelle, welche mit einem *Quadrupol-Time-of-Flight* (QTOF)-Massenspektrometer (Ion Sense DART source und Agilent Technologies 6560 Q-TOF MS-System) gekoppelt wurde, durchgeführt. Die Probenaufgabe in die DART-Quelle erfolgte mittels DIP-it Probennehmern, wobei es sich um Glasstäbe handelt, welche in den Probenextrakt getaucht werden und anschließend automatisiert durch die DART-Quelle gefahren werden. Dieser Messaufbau hat den Vorteil, dass eine Messzeit von wenigen Sekunden pro Probe ausreichend ist. Die Extrakte wurden im positiven Ionenmodus vermessen, da hier eine gute Ionisierung der lipophilen Stoffe erreicht wird.

Die Datenprozessierung wurde mit der Umwandlung des Datenformats durch die Software *Proteowizard* begonnen. Die weitere Datenverarbeitung erfolgte mit der Software *MZmine2*. Die ungerichtete multivariate Auswertung wurde mittels der Software *Metabo Analyst 5.0* ausgeführt, während die gerichtete Klassifizierung mittels *R* und dem Paket *RandomForest* durchgeführt wurde. Die Markeridentifizierung erfolgte mittels der Software *Qualitative Analysis* der Firma Agilent anhand der aufgenommenen MS^2-Spektren sowie durch einen Datenbankabgleich [3].

Herkunftsbestimmung mittels DART-MS

Durch die unpolare Extraktion und die Vermessung der 200 Proben mittels DART-MS konnten 647 Signale (*Features*) erfasst werden. Abbildung 2 zeigt ein beispielhaftes Massenspektrum einer Körnermaisprobe. Insbesondere im Bereich der Triacylglyceride (TGs) von *m/z* 800-950 zeigen sich viele *Features* mit hoher Intensität sowie im Bereich der Diacylglyceride (DGs) von *m/z* 550-650. Die Signale in einem Bereich von *m/z* 230-350 sind unter anderem auf Monoacylglyceride (MGs) sowie Fettsäuren (FS) zurückzuführen.

Zur Auswertung mittels einer ungerichteten Analyse, also ohne Berücksichtigung der Herkünfte der Proben durch die Software, wurde eine Hauptkomponentenanalyse (PCA, *principal component analysis*) durchgeführt. Die erhaltenen *Scoreplots* der Hauptkomponenten (PC) 1–4 sind in Abbildung 3 dargestellt. Basierend auf der ersten PC ergibt sich eine klare Trennung in zwei Cluster: Einer der Cluster setzt sich aus den Proben aus Frankreich, Ungarn und Deutschland zusammen, während der andere Cluster die Stichproben aus Spanien, den USA, der Ukraine und Peru enthält. Die Aufteilung in zwei Cluster lässt sich dadurch erklären, dass die Proben aus Deutschland, Frankreich und Ungarn aus Chargen stammen, welche für die Stärkeproduktion bestimmt waren. Zudem wurde eine leicht andere Probenvorbereitung verwendet. Durch die vorhergehende LC-ESI-MS-Analyse wurde gezeigt,

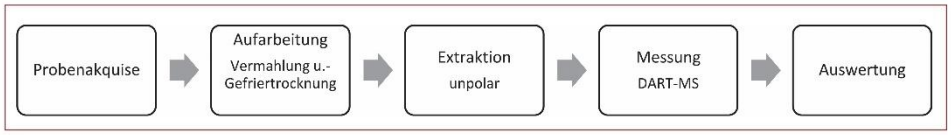


Abb. 1 Schematischer Arbeitsablauf einer Metabolomanalyse [7]

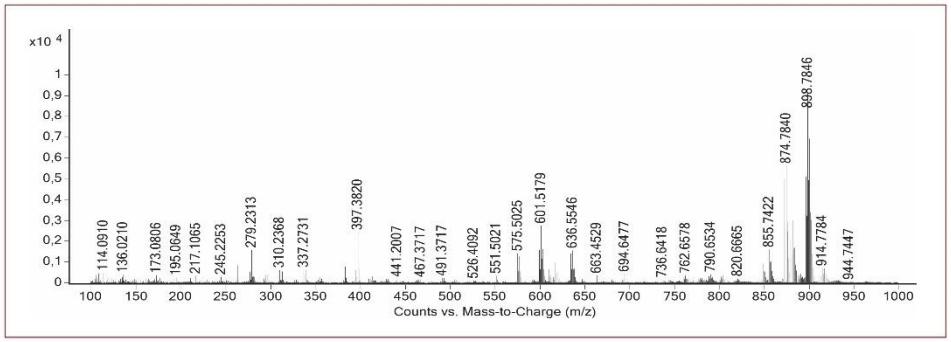


Abb. 2 Exemplarisches Massenspektrum eines unpolaren Extraktes einer Körnermaisprobe gemessen mittels des Dip-it Verfahrens und DART-MS.

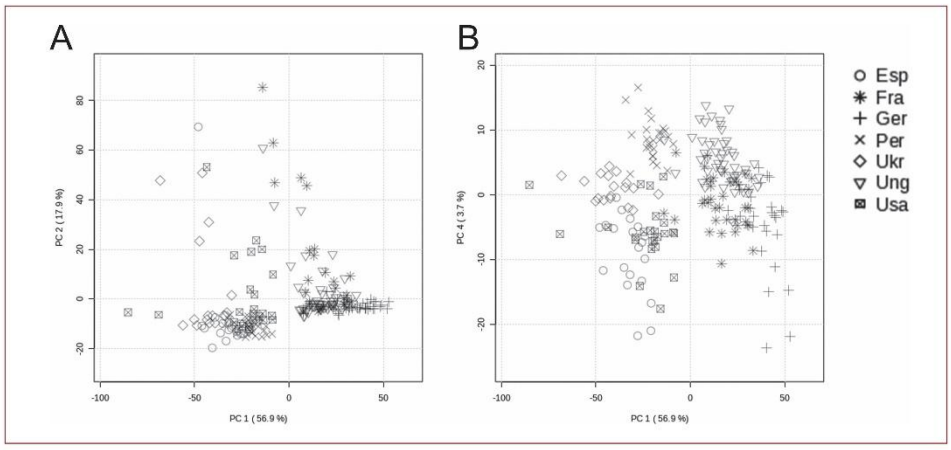


Abb. 3 Score Plots der PCA der mittels DART-MS vermessenen Körnermaisproben. Aufgetragen sind die PC 1 gegen PC 2 (A) und PC 1 gegen PC 4 (B).

dass abgesehen von der Konzentration von DGs nur ein geringer Unterschied zwischen dem Futtermais und Mais für die Stärkegewinnung bestand [3]. Innerhalb dieser beiden Cluster konnte dennoch eine Tendenz zur räumlichen Trennung nach Ländern beobachtet werden. Zum Beispiel zeigten die deutschen Proben höhere Werte für die erste Hauptkomponente als die französischen und ungarischen Proben. Weiter zeigte der *Scoresplot* der ersten und vierten PC eine gute Trennung der Proben nach ihrer Herkunft (siehe Abbildung 3B). Hieraus kann geschlossen werden, dass die Hauptvarianzen zwischen den Proben nicht für die Trennung nach der Herkunft relevant sind. Dennoch lässt sich in den weiteren Hauptkomponenten eine Clusterung der Proben erkennen, sodass darauf geschlossen werden kann, dass herkunftsrelevante Informationen im Datensatz vorhanden sind.

Darüber hinaus wurde zur Klassifizierung der Proben nach ihrer Herkunft eine gerichtete Analyse basierend auf dem *Random Forest* Algorithmus verwendet. Validiert wurde diese Analyse durch eine 10-fache Kreuzvalidierung mit 100-facher Wiederholung, um den Zufallseinfluss zu minimieren. Dabei wurde eine Richtigkeit der Testdatensätze von 84,4 % erzielt. Die Konfusionsmatrix in Abbildung 4 gibt die Zuordnungsraten der Proben in den Testdatensätzen nach Land an. Die Richtigkeit für die Proben der einzelnen Länder liegt zwischen 73,7 und 95,7 %. Die Proben aus Peru und Deutschland zeigten die höchsten Zuordnungsraten von 95,7 bzw. 92,7 %. Bei den Proben aus der Ukraine ist eine vergleichsweise niedrige Richtigkeit von 73,7 % gegeben, was auf die geringere Anzahl von Proben (17) zurückzuführen ist. Die ungarischen Proben wurden mit einer Genauigkeit von 80,1 % klassifiziert, wobei die Fehl-

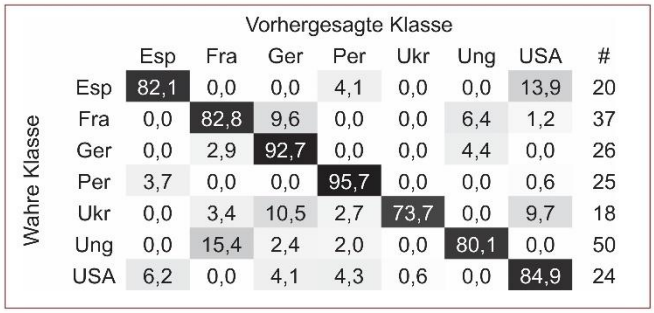


Abb. 4 Konfusionsmatrix der *Random Forest* Klassifizierung der mittels DART-MS vermessenen Körnermaisproben. Aufgezeigt sind die Zuordnungsraten zu den Ländern Spanien, Frankreich, Deutschland, Peru, der Ukraine, Ungarn und den USA sowie die Anzahl der Proben pro Klasse (#).

zuordnungen zwischen Frankreich und Ungarn auffällig sind. Beide Länder haben ein kontinentales und warmes Klima mit sehr ähnlichen Jahreswerten für Temperatur und Niederschlag, was die Unterscheidung zwischen diesen beiden Klassen wahrscheinlich erschwert hat. Darüber hinaus gab es ein paar Fehlklassifizierungen zwischen Proben aus Spanien und den USA, die Genauigkeiten von 82,1 beziehungsweise 84,9 % aufweisen. Hierfür könnte die große Varianz der vorhandenen Proben aus den USA verantwortlich sein, da diese aus verschiedenen Bundesstaaten von Georgia bis Texas kamen und somit ein deutlich größeres Gebiet abbilden, als es bei den anderen Ländern der Fall ist. Die Probennahme aus verschiedenen Bundesstaaten bietet den Vorteil einer besseren Abdeckung der USA im Klassifizierungsmodell.

Um die wichtigsten Verbindungen zu identifizieren, welche für eine Unterscheidung der Herkünfte relevant sind, wurden zunächst die Fragmentierungsmuster von ähnlichen Referenzsubstanzen analysiert. Dies war besonders wichtig, da bislang nur wenig Literatur über die Fragmentierung von lipophilen Molekülen mit DART-MS^2 vorliegt und daher keine Datenbanksuche durchgeführt werden konnte. Die Fragmentierungsmuster von verschiedenen Standards aus der Klasse der MGs, DGs, TGs sowie Glycerophosphoethanolamine und Glycerophosphocholine konnten detektiert und aufgeklärt werden. Dabei wurde deutlich, dass die MGs, DGs und TGs sehr ähnliche Fragmentierungsmuster zu einer Analyse mittels LC-ESI-MS^2 aufweisen. Im Gegensatz dazu unterliegen die mittels DART-MS^2 vermessenen Phospholipide einer *in-source* Fragmentierung und können nur begrenzt mittels DART-Ionenquellen erfasst werden.

Unter Verwendung von DART-MS^2 Spektren war es möglich, sieben Markersubstanzen zu identifizieren. Es handelte sich um die Verbindungen DG 34:2, DG 34:1, DG 36:3 und DG 36:2 sowie TG 50:1, TG 52:3 und TG 54:5. Damit konnten insbesondere Marker aus der Klasse der DGs und TGs als relevant für die Klassifizierung ermittelt werden. Auffällig ist zudem, dass die TGs als $[M+NH_4]^+$-Addukte ionisieren, obwohl dem Lösemittel kein Ammonium zugesetzt wurde. Dafür sind die in der Laboratmosphäre häufig vorhandenen Ammoniakdämpfe verantwortlich, welche bereits in niedrigen Konzentrationen ausreichen, um solche Addukte zu formen [9].

Die Intensitäten dreier ausgewählter Marker sind in Abbildung 5 in Form von Boxplots dargestellt. Die Boxplots zeigen, dass die Proben unterschiedlicher Herkünfte bereits bei der Betrachtung einzelner Marker Intensitätsunterschiede aufweisen. So ist es beispielsweise mittels DG 36:2 möglich, die Proben sämtlicher untersuchter Länder von den Proben aus Peru abzugrenzen, da dieser Marker in den Proben aus Peru sehr hohe Intensitäten aufweist. Die TG 50:1 und TG 52:3 sind dagegen gute Marker, um die zentraleuropäischen Länder Frankreich, Deutschland und Ungarn von den Proben der verbleibenden Herkunftsländer zu unterscheiden.

Vergleich von DART-MS und LC-ESI-MS

Zum Leistungsvergleich der Herkunftsunterscheidung mittels DART-MS wurde LC-ESI-MS als ähnliche und bereits für die Herkunftsbestimmung etablierte Methode gewählt [7].

Ein großer Vorteil der Messung mit einem DART-MS-Gerät ist die kurze Analysezeit von nur wenigen Sekunden pro Probe. Die LC-ESI-MS Methode, mit einer Analysezeit von durchschnittlich 30 Minuten pro Probe, führte zu Messzeiten von mehreren Tagen für den kompletten Datensatz. Mit der DART-MS-Plattform konnten die Messungen innerhalb eines Tages abgeschlossen werden, sodass Zeit und Personal entsprechend eingespart werden können. Da die Extrakte bei der Messung der Proben mittels LC-ESI-MS verdünnt wurden und dieser Schritt bei der DART-MS Analyse nicht notwendig ist, konnten der zeitliche Aufwand und der Lösungsmittelverbrauch reduziert werden.

Für den Endnutzer ist die Datenauswertung einer LC-ESI-MS-Analyse einfacher, da für diese mittlerweile zahlreiche Soft-



Felix Schmauder
Zur Person: Felix Schmauder studierte Lebensmittelchemie an der Universität Würzburg und promoviert seit 2021 an der Hamburg School of Food Science – Universität Hamburg zum Thema Herkunftsauthentifizierung von Lebensmitteln mittels DART-MS.

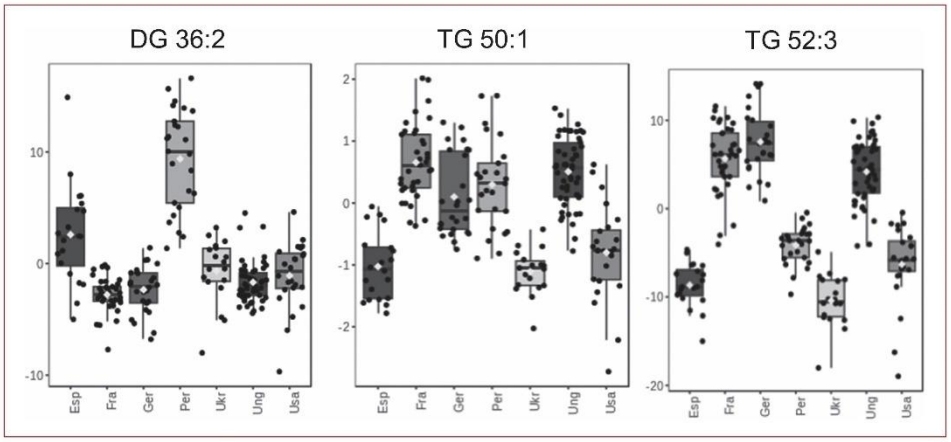


Abb. 5 Normalisierte Intensitäten von drei identifizierten Markersubstanzen DG 36:2, TG 50:1 und TG 52:3.

waretools zur Verfügung stehen. In der Regel können hierfür die Programme der Hersteller genutzt werden, weshalb keine Konvertierung der Datenformate erfolgen muss, und die Workflows zum Umgang mit derartigen Daten sind etabliert. Im Gegensatz dazu gibt es für DART-MS-Plattformen bisher nur wenig Software für den Endanwender. Es ist jedoch zu erwarten, dass spezialisierte Software entwickelt wird, wenn DART-MS-Plattformen sich in der Forschung und Routineanwendungen etablieren.

Da die Nachhaltigkeit beziehungsweise Umweltverträglichkeit der entwickelten Methoden in den letzten Jahren immer wichtiger geworden sind, spielt die Reduzierung der Umweltbelastung eine immer größere Rolle. Die Messung mittels LC verbraucht circa 10 mL an organischen Lösungsmitteln pro Probe, während die Menge an Lösungsmittelmenge bei der Extraktionsmethode mit 1 mL vergleichsweise gering ist. Hier bietet der DART-MS Ansatz einen großen Vorteil, da ohne Chromatographie etwa 90 % des Lösungsmittels eingespart werden können. Mit dem Wegfall des Großteils der Lösungsmittel werden circa ein Drittel der Kosten der Analyse eingespart. Somit handelt es sich bei DART-MS um ein ressourceneffizientes Verfahren.

Darüber hinaus wird bei der Verwendung von DART-Quellen nur der ionisierte Teil der Probe in das Analysengerät eingezogen. Hierdurch können Kontaminationen vermieden werden, welche bei LC-ESI-MS-Systemen des Öfteren ein Problem darstellen [1].

Mittels der LC-ESI-MS-Messungen wurden 1 022 Signale detektiert, das ist circa ein Drittel mehr gegenüber dem DART-MS Ansatz. Unter Verwendung des RF-Algorithmus wurde eine sehr gute Richtigkeit von 98,5 % erreicht. Eine Reduktion der Variablen auf 20 Marker führte zu einem Endergebnis mit einer Genauigkeit von 90,5 % [3].

Im Gegensatz dazu erreichte der DART-MS-Ansatz mit einer Genauigkeit von 84,4 % nicht ganz die Richtigkeit der LC-ESI-MS-Analyse, bietet aber die oben genannten Vorteile. Beim Vergleich der identifizierten Marker traten Überschneidungen zwischen der LC-ESI-MS- und der DART-MS-Plattform auf. Die Lipide DG 34:2, DG 34:1 und DG 36:3 wurden mit beiden Methoden als Marker detektiert und identifiziert, was die Bedeutung dieser Verbindungen für die Unterscheidung der verschiedenen Herkünfte unterstreicht. Außerdem wurden vier Moleküle, DG 36:2, TG 50:1, TG 52:3 und TG 54:5, die bisher nicht als Marker bekannt waren, durch die DART-MS²-Messungen identifiziert.



Dr. Marina Creydt
Zur Person: Marina Creydt studierte Lebensmittelchemie an der Hamburg School of Food Science – Universität Hamburg und forscht derzeit unter anderem an der Entwicklung von massenspektrometrischen Methoden im Rahmen von Fragestellungen zum Thema Food Fraud



Prof. Dr. Markus Fischer
Zur Person: Lebensmittelchemiker, seit 2006 Direktor des Instituts für Lebensmittelchemie der Universität Hamburg und Gründer/Direktor der Hamburg School of Food Science.

Die dargestellten Ergebnisse wurden aus der Studie „Schmauder F., Schütz D., Creydt M., Riedl J., Fischer M. Food authentication goes green: Determination of the geographical origin of grain maize (Zea mays L.) using direct analysis in real time mass spectrometry (DART-MS)" reproduziert.

Die Autoren danken David Schütz und Janet Riedl für die Bereitstellung der Proben.

Zusammenfassung und Ausblick

In dieser *Proof-of-Concept*-Studie wurde ein neuer Arbeitsablauf zur Bestimmung der Herkunft von Körnermais mittels einer unpolaren Extraktion in Kombination mit einem DART-MS-Ansatz vorgestellt. Diese kostengünstige und umweltfreundliche Massenspektrometrie-basierte Methode bietet den Vorteil kurzer Analysezeiten, um einen interpretierfähigen Fingerabdruck von Lebens- und Futtermittelproben zu erhalten.

Anhand der Analyse von 200 Proben, die mittels *Random Forest* klassifiziert werden konnten, wurde eine Richtigkeit von 84,4 % erreicht. Dies demonstriert, dass DART-MS sich für die Herkunftsbestimmung aufgrund der analysierten chemischen Fingerabdrücke eignet. Die mit *msConvert*, *MZmine2* und *R* entwickelte Auswertestrategie basiert rein auf *Open-Source*-Software und bietet den Vorteil, dass sie frei zugänglich und gut nachvollziehbar sind. Außerdem hat sich gezeigt, dass der *Random Forest* Algorithmus für die Auswertung von Daten ohne chromatographische Trennung geeignet ist. In weiteren Experimenten wurden geeignete Markersubstanzen für die Klassifizierung der geographischen Herkunft von Mais mit Hilfe von DART-MS^2 identifiziert. Neben neuen Markersubstanzen konnten auch bereits bekannte Markerverbindungen verifiziert werden. Die Lipidmarker gehören zu den Substanzklassen der MGs, DGs und TGs. Mithilfe des beschriebenen Workflows und dem kontinuierlichen Aufbau einer Datenbank kann ein wichtiger Beitrag zur Entwicklung weiterer Verfahren für die Authentifizierung von Lebensmitteln geleistet werden.

Referenzen

[1] Gross, J.H. Direct analysis in real time—a critical review on DART-MS. *Analytical and bioanalytical chemistry* **2014**, *406*, 63–80.
[2] Sisco, E.; Forbes, T.P. Forensic applications of DART-MS: A review of recent literature. *Forensic Chemistry* **2021**, *22*, 100294.
[3] Schütz, D.; Achten, E.; Creydt, M.; Riedl, J.; Fischer, M. Non-targeted LC-MS metabolomics approach towards an authentication of the geographical origin of grain maize (Zea mays L.) samples. *Foods* **2021**, *10*, 2160.
[4] United States Department of Agriculture. Grain: World Markets and Trade: 2023/24 Grain Production Exceeds Consumption. Available online: https://apps.fas.usda.gov/psdonline/circulars/grain.pdf (accessed on 24 May 2023).
[5] Khodaei, D.; Javanmardi, F.; Khaneghah, A.M. The global overview of the occurrence of mycotoxins in cereals: A three-year survey. *Current Opinion in Food Science* **2021**, *39*, 36–42.
[6] Seifert, S. Chemometric analysis of analytical fingerprints. *Spectroscopy* **2023**.
[7] Creydt, M.; Hudzik, D.; Rurik, M.; Kohlbacher, O.; Fischer, M. Food authentication: Small-molecule profiling as a tool for the geographic discrimination of German white asparagus. *Journal of agricultural and food chemistry* **2018**, *66*, 13328–13339.
[8] Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology* **1959**, *37*, 911–917.
[9] Hajslova, J.; Cajka, T.; Vaclavik, L. Challenging applications offered by direct analysis in real time (DART) in food-quality and safety analysis. *TrAC Trends in Analytical Chemistry* **2011**, *30*, 204–218.

Kritzmöllers Warenwelt



Unschuldsvermutung

„Ohne Zusätze. Rahmt natürlich auf, daher vor Gebrauch schütteln." Diese Handlungsanweisung formuliert „Berchtesgadener Land" auf seinem Demeter-Schlagrahm. Tatsächlich bildet sich auf dem schneeweißen Rahm eine dicke Fettschicht, die – sofern nicht verrührt oder prophylaktisch vernascht – uncharmant in den Kaffee zu plumpsen droht. Wie viel homogener zeigen sich da konventionelle Produkte, die in ihrer makellosen Konsistenz an einen Inbegriff unschuldiger Reinheit grenzen. An „Zusätze" dürften bei diesem Anblick bestenfalls Lebensmittelchemiker denken; Laiin und Laie wähnen sich hingegen im Gebrauch eines naturreinen Rohstoffs.

Carrageen, eine aus Rotalgen gewonnene Substanz, die auch in der Kosmetikindustrie Verwendung findet, sorgt dafür, dass Fett- und Milchbestandteile auch nach längeren Standzeiten treu verbunden bleiben. Der Stoff gilt als gesundheitlich unbedenklich – und dennoch trägt gerade der Verweis auf sein Fehlen zur Entzauberung (allzu) naiver Konsumentenvorstellungen bei.

PD Dr. Monika Kritzmöller
Forschungs- und Beratungsinstitut
Trends + Positionen
www.kritzmoeller.ch
mail@kritzmoeller.ch

9.7 Zeitschriftenartikel: Grüne Lebensmittelanalytik mittels DART-MS

Lebensmittelanalytik

Lebensmittelauthentifizierung

Grüne Lebensmittelanalytik mittels DART-MS

Umweltfreundlichkeit, Schnelligkeit und Kostenreduzierung sind heute wichtige Kriterien für die Entwicklung analytischer Verfahren. Vor diesem Hintergrund biete die DART-Massenspektrometrie Möglichkeiten für die Lebensmittelanalytik, so die Autoren des folgenden Artikels. Sie haben mit dieser Methode Untersuchungen zur Herkunftsanalyse von Körnermais durchgeführt.

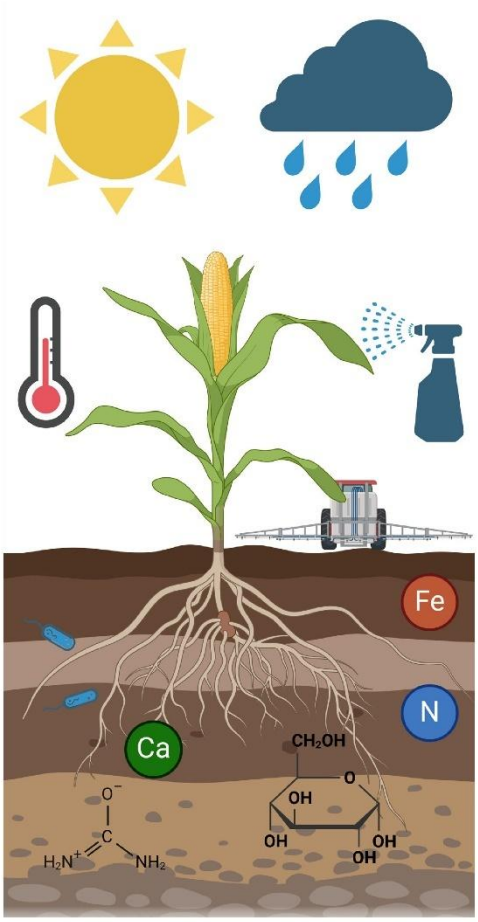


Bild 1: Einfluss exogener Faktoren auf das Metabolom. Bild: Felix Schmauder/Universität Hamburg; Grafik erstellt mit biorender.com

Direct analysis in real time (DART) stellt eine vergleichsweise neue Ionisationstechnik für die Massenspektrometrie (MS) dar. Diese Ionisationsmethode macht das direkte Einbringen und Ionisieren einer Probe oder eines Probenextraktes möglich, ohne vorherige chromatographische Trennung der Analyten. In Verbindung mit einem hochauflösenden Quadrupole-Time-of-Flight (qToF)-Massenanalysator oder einem in der Routineanalytik häufig eingesetztem Triple-Quadrupol (QQQ)-Massenspektrometer ergeben sich mit DART einige inhärente Vorteile. Der offensichtlichste Vorzug ist die Schnelligkeit der Methode, da durch die Einsparung der Chromatographie die Zeit der Messung von 15 – 30 Minuten auf weniger als eine Minute pro Probe reduziert werden kann. Darüber hinaus handelt es sich gegenüber einer Methode, die eine chromatographische Auftrennung erfordert, um eine umweltfreundlichere Methode, da keine organischen Lösungsmittel für die Messungen notwendig sind und auch kein Chemikalienabfall entsteht, welcher aufwändig entsorgt werden müsste. Zudem wird die Analytik zeitgleich günstiger, weil keine Kosten für Lösungsmittel, Säulen, Wartung der chromatographischen Anlage oder Abfallentsorgung auftreten. Auf diese Weise könnten circa 1/3 der Betriebskosten eingespart werden. Darüber hinaus wird das Massenspektrometer im Vergleich zu Chromatographie- oder Direkt-Infusionsexperimenten weniger verunreinigt, da lediglich die ionisierten Moleküle und keine neutralen Moleküle in das Gerät gelangen.

Anwendung: Herkunftsbestimmung von Lebensmitteln

Um die Eignung dieses relativ neuen und bislang wenig angewendeten Verfahrens für Fragestel-

lungen im Bereich Lebensmittelauthentizität zu evaluieren, wurde geprüft, ob die Methode eingesetzt werden kann, um die geographische Herkunft von Körnermais zu bestimmen. Der Nachweis der geographischen Herkunft von Körnermais ist relevant, da der Ursprung die Qualität durch Faktoren wie Anbaupraktiken und Umweltbedingungen beeinflussen kann, welche potenziell gesundheitliche Risiken mit sich bringen können. Hier kommen auch unterschiedliche gesetzliche Vorschriften zum Tragen. Im Fokus der Untersuchung standen Stoffwechselprodukte, anhand derer ein molekularer Fingerabdruck in Abhängigkeit der geographischen Herkunft erstellt wurde. Dieser Ansatz beruht darauf, dass die qualitative und quantitative Zusammensetzung des Metaboloms, also die Gesamtheit der Stoffwechselprodukte, durch verschiedene äußere Faktoren beeinflusst wird (Bild 1). Der im Folgenden vorgestellten Vorgehensweise mittels DART-MS ging eine massenspektrometrische Untersuchung mittels eines QToF-Massenspektrometers voraus, das über eine Elektronenspray-Ionisations-Quelle (ESI) mit einer Flüssigkeitschromatographie (LC) verbunden wurde (LC-ESI-QTOF-MS/(MS)). Denn diese massenspektrometrische Plattform wird üblicherweise für vergleichbare Fragestellungen eingesetzt, und so war es möglich, die erhaltenen Ergebnisse aus beiden Verfahren miteinander zu vergleichen.



Bild 2: Arbeitsablauf einer Metabolomanalyse mittels DART-MS. Bild: Universität Hamburg

Beispiel: Authentifizierung von Körnermais

In der vorgestellten Studie wurde Körnermais (Zea mays L.) untersucht. Zunächst erfolgte die Probenakquise von 200 Referenzproben aus sieben Ländern. Diese Maisproben wurden homogenisiert und gefriergetrocknet (Bild 2). Für die Extraktion der Metabolite wurde eine Abwandlung der unpolaren „Bligh and Dyer"-Extraktion angewandt, und die Extrakte wurden mittels einer DART-Quelle ionisiert und mittels eines qToF-Analysators analysiert (Bild 3).

Dabei wurden zeitgleich 647 Signale (Features) erfasst. Die Analyse der Massenspektren zeigte hohe Signalintensitäten, insbesondere im Bereich der Triacylglyceride (TG, m/z 800 – 950) und der Diacylglyceride (DG, m/z 550 – 650). Darüber hinaus fanden sich Signale von Monoacylglyceriden (MG) und Fettsäuren (FS) im Bereich von m/z 230 – 350 (m/z = Masse-zu-Ladungsverhältnis). Zur Bewertung der Herkunft der Proben wurde eine Hauptkomponentenanalyse durchgeführt, in welcher die Daten ohne Berücksichtigung der Herkunft nach ihrer Varianz aufgetrennt wurden. Die Analyse der Hauptkomponenten ergab insbesondere in der Darstellung der Hauptkomponente 1 gegen 4 eine Tendenz zur räumlichen Auftrennung der Proben nach ihrer Herkunft (Bild 4A). Dies deutet darauf hin, dass herkunftsrelevante Informationen in den Daten vorhanden sind.

Für eine gezielte Klassifizierung der Proben zur Herkunftsanalyse wurde der „Random Forest"-Algorithmus genutzt. Die Analyse wurde durch zehnfache Kreuzvalidierung mit 100 Wiederholungen verifiziert, was zu einer Genauigkeit von 84,4 % führte (Bild 4B). Die Genauigkeit variierte dabei zwischen 73,7 % für Proben aus der Ukraine und 95,7 % für die Proben aus Peru. Besonders bei den ungarischen Proben zeigten sich mit 80,1 % eine geringere Genauigkeit und häufig Verwechslungen mit französischen Proben, was auf Ähnlichkeiten im Klima und den geographischen Bedingungen zwischen Ungarn und Frankreich zurückzuführen sein könnte. Bei den Proben aus den USA und Spanien traten ebenfalls einige Fehlklassifizierungen auf, die möglicherweise aus der heterogenen Herkunft der amerikanischen Proben resultierten.

Herkunfts-Marker identifiziert

Um die signifikantesten Verbindungen zur Herkunftsbestimmung zu identifizieren, wurden die Fragmentierungsmuster mittels MS/MS-Experimenten von verschiedenen Standards wie MG, DG und TG analysiert. Die Fragmentierungsmuster dieser Verbindungen zeigten Ähnlichkeiten mit denen, die in der vorangegangen LC-ESI-QTOF-MS/(MS)-Analyse beobachtet wurden. Allerdings unterlagen die Phospholipide, die mittels DART-

MS/MS gemessen wurden, einer In-source-Fragmentierung, was die Analyse erschwerte. Insgesamt konnten sieben Marker identifiziert werden, die zur Herkunftsklassifizierung der Proben relevant sind, darunter DG 34:2, DG 34:1, DG 36:3 und DG 36:2 sowie TG 50:1, TG 52:3 und TG 54:5 (angegeben sind hier die Kurzformen der Verbindungsnamen mit Zahl der Kohlenstoffatome:Zahl der Doppelbindungen). Diese Marker, insbesondere aus den Klassen der DG und TG, erwiesen sich als besonders geeignet für die Unterscheidung der Herkunft.

Besonders bemerkenswert war die Ionisierung der Triacylglyceride in Form von $[M + NH_4]^+$-Addukten (M = Molekülmasse), bedingt durch in der Laboratmosphäre vorhandene Ammoniakdämpfe. Die Intensitäten ausgewählter Marker wurden mittels Boxplots (Bild 4C) untersucht, welche verdeutlichen, dass bereits bei einzelnen Markern Unterschiede zwischen den Proben aus verschiedenen Herkunftsländern bestehen. Der Marker DG 36:2 war besonders effektiv, um die Proben aus Peru abzugrenzen, während TG 50:1 und TG 52:3 gute Indikatoren zur Differenzierung zwischen den zentral- und osteuropäischen Ländern boten. Die Ergebnisse dieser Studie unterstreichen das Potenzial von DART-MS für die Herkunftsbestimmung von Körnermais, indem sie sowohl qualitativ als auch quantitativ wertvolle Informationen zu den Metaboliten liefern. Zudem verdeutlichen die Ergebnisse, dass mittels DART-MS sowohl Konzentrationsunterschiede der einzelnen Marker erfasst werden konnten als auch eine erfolgreiche Klassifizierung durchgeführt werden konnte.

Zusammenfassung und Ausblick

Bereits erfolgreich eingesetzt werden DART-MS-Kopplungen für die Drogen- und Sprengstoffdetektion an Flughäfen, da für diese eine minutenschnelle Analytik für den reibungslosen Ablauf der Sicherheitskontrollen nötig ist. In der Routineanalyse für Lebensmittel sind sie derzeit noch nicht sehr verbreitet. In diesem Rahmen werden vorrangig LC-ESI-QQQ-Geräte eingesetzt, welche vergleichsweise kostengünstig sowie sensitiv und robust sind. Im Vergleich zu einem qTOF-Analysator, welcher in der beschriebenen Studie für die Identifizierung der Markerverbindungen eingesetzt wurde, weisen QQQ-Analysatoren eine deutlich geringere Massenauflösung auf und sind deshalb nicht für die Identifizierung von Markerverbindungen geeignet (qTOF-Massenspektrometer = Quadrupol-Flugzeit-Massenspektrometer). Sind die Analyten allerdings bekannt, eignen sich auch QQQ-Analysatoren für Authentifizierungsexperimente. So könnten die im obigen Beispiel identifizierten Markerverbindungen auf ein LC-ESI-QQQ-MS-Gerät übertragen und validiert werden. Auf diese Weise kann der Transfer von Forschung in die direkte Anwendung sichergestellt

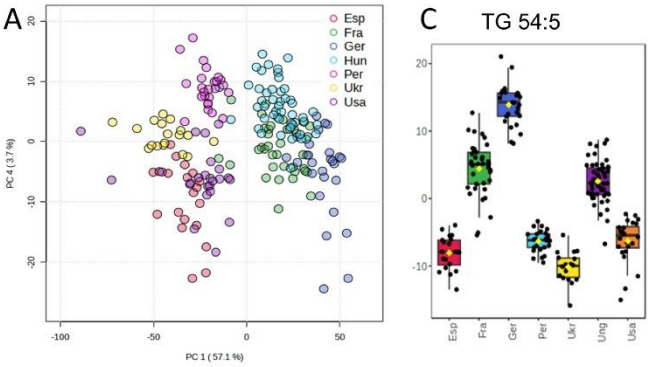


Bild 3: Die verwendete DART-Ionenquelle mit „DIP-It"-Probenaufgabesystem. Bild: Universität Hamburg

werden. Auch aus diesem Grund arbeiten die Hersteller für Massenspektrometer derzeit an der Entwicklung von DART-QQQ-Kopplungen für die Routineanalytik, um die beschriebenen Vorteile von DART-Ionisierung und QQQ-Analysator auch für die Routineanalytik zugänglich zu machen. Hierdurch könnten sich diese Bauweisen zukünftig in der Routineanalytik für viele Authentifizierungsfragestellungen etablieren, solange eine überschaubare Anzahl von Markerverbindungen für die Klassifikation ausreichend ist.

Ein weiterer Parameter, welcher voraussichtlich die Verbreitung von lösungsmittelfreien Techniken beschleunigen wird, ist das 1,5-Grad-Ziel des Pariser Klimaabkommens. Aufgrund der steigenden CO_2-Abgaben und der Verknappung von fossilen Ausgangsprodukten ist eine Preissteigerung für organische Lösemittel erwartbar. Dieser Marktmechanismus begünstigt die Etablierung von lösungsmittelfreien Analysen unter der Voraussetzung, dass die Verfügbarkeit analytischer interner Standards gegeben ist.

Ein weiteres Zukunftsfeld im Rahmen von DART-MS-Applikationen ist die Automatisierung, insbesondere in der Probenvorbereitung. Hier könnten die Kosten für die Analytik durch den Einsatz von Pipettierrobotern weiter reduziert werden, da die Probenvorbereitung derzeit die größte Fehlerquelle der Analytik darstellt und die meiste Zeit in Anspruch nimmt. Optimal wäre eine Kombination aus einem automatisiertem Pipettierroboter und DART-MS, um einen möglichst hohen sowie reproduzierbaren Probendurchsatz zu gewährleisten. Eine solche Vorgehensweise, die auf einem umweltfreundlichen Messverfahren basiert, könnte die Analytik auf ein neues und nachhaltiges Niveau heben.



B

Vorhergesagte Klasse

Wahre Klasse	Esp	Fra	Ger	Per	Ukr	Ung	USA	#
Esp	82,1	0,0	0,0	4,1	0,0	0,0	13,9	20
Fra	0,0	82,8	9,6	0,0	0,0	6,4	1,2	37
Ger	0,0	2,9	92,7	0,0	0,0	4,4	0,0	26
Per	3,7	0,0	0,0	95,7	0,0	0,0	0,6	25
Ukr	0,0	3,4	10,5	2,7	73,7	0,0	9,7	18
Ung	0,0	15,4	2,4	2,0	0,0	80,1	0,0	50
USA	6,2	0,0	4,1	4,3	0,6	0,0	84,9	24

Bild 4: Ergebnisse der Herkunftsbestimmung von Körnermais mittels DART-MS. Abkürzungen für die Herkunftsländer: Esp (Spanien), Fra (Frankreich), Ger (Deutschland), Hun u. Ung (Ungarn) Per (Peru), Ukr (Ukraine), Usa (USA). A: Scores-Plot der 1. und der 4. Hauptkomponente. B: Konfusionsmatrix (aus „Random Forest"-Algorithmus) der gemittelten Testdatensätze zur Darstellung der Vorhersagen des Modells im Vergleich zu den tatsächlichen Klassen. Die Angaben sind in Prozent aufgeführt. Zudem kann der Abbildung die Anzahl der gemessenen Proben (#) entnommen werden. C: Box-Plot der relativen Konzentration eines Triacylglycerids (TG 54:5), welches als geeignete Markerverbindung identifiziert werden konnte. Bilder: Elsevier; bearbeitet Felix Schmauder/Universität Hamburg

Danksagung:
Die Autoren danken Janet Riedl vom Bundesinstitut für Risikobewertung und David Schütz für die Bereitstellung der Proben.

Publikationen:
Felix Schmauder, David Schütz, Marina Creydt, Janet Riedl, Markus Fischer: Food authentication goes green: Determination of the geographical origin of grain maize (Zea mays L.) using direct analysis in real time mass spectrometry (DART-MS), Food Control, Volume 163, 2024; https://doi.org/10.1016/j.foodcont.2024.110489
David Schütz, Elisabeth Achten, Marina Creydt, Janet Riedl, and Markus Fischer: Non-Targeted LC-MS Metabolomics Approach towards an Authentication of the Geographical Origin of Grain Maize (Zea mays L.) Samples, Foods 2021, 10(9), 2160; https://doi.org/10.3390/foods10092160

AUTOREN
Felix Schmauder, Dr. Marina Creydt, Prof. Dr. Markus Fischer
Hamburg School of Food Science – Institut für Lebensmittelchemie
Universität Hamburg
leitung.chemie@uni-hamburg.de
www.chemie.uni-hamburg.de

10 Danksagung

Ich möchte herzlich allen Menschen danken, die mich bis hierher begleitet haben.

Zunächst möchte ich Herrn Prof. Dr. Fischer für die Möglichkeit der Anfertigung dieser Dissertation danken. Vielen Dank für das spannende und neue Thema, sowie die Hilfe bei der Umsetzung und die angeregten fachlichen Diskussionen.

Danke an Dr. Marina Creydt, die mich tatkräftig bei der Ideenfindung, fachlichen Problemen und insbesondere bei dem Verfassen von meinen ersten wissenschaftlichen Publikationen unterstützt hat.

Darüber hinaus möchte ich allen Mitarbeitern und Doktoranden für die schöne Zeit an der Hamburg School of Food Science danken. Die gemeinsamen Praktikumsaufsichten, das Mittagessen und die Diskussionen haben meine Zeit erheblich bereichert.

Herausheben möchte ich hier mein Büro mit Joey, Nils, Henri, Christian und Timothy. Ihr wart eine große Hilfe und moralische Unterstützung.

Daneben möchte ich Marie-Sophie Müller, Dr. David Schütz, Dr. Janet Riedl, Dr. René Bachmann und Dr. Sven Klockmann für die Bereitstellung der Haselnuss- und Körnermaisproben danken. Für die Hilfe bei der Akquise der Erdbeerproben sowie der Probenaufarbeitung möchte ich Lina Cvancar, Robin Dammann, Kerstin Blum, Andreas Heinze und Mirjam Rose danken. Zudem gilt meinen Dank den Erdbeererzeugern und dem Forschungskreis der Ernährungsindustrie, die durch ihre Unterstützung dieses Projekt möglich gemacht haben.

Zudem danke ich Herrn Prof. Seifert, Soeren Wenck, Jule Hansen und Kim Brettschneider für die Bereitstellung des Handheld NIR-Gerätes und für die gute Zusammenarbeit bei der Datenauswertung.

Abschließend möchte ich meiner Familie und meinen Freunden für die Unterstützung danken. Ohne meine Eltern, meine Familie und die große Zahl an Freunden, wäre diese Arbeit so nicht möglich gewesen.

Danke!

11 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die KI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, 13.12.2025

Felix Schmauder