# Entwicklung von LC-MS-basierten Metabolomics-Applikationen zur Bestimmung der geographischen Herkunft von Haselnüssen (*Corylus avellana*)

### Dissertation

Zur Erlangung des akademischen Grades

#### Doctor rerum naturalium

Dr. rer. nat.

aus dem Fachbereich Chemie der Universität Hamburg, Hamburg School of Food Science

> vorgelegt von Sven Klockmann

Hamburg 24. Dezember 2017

Hamburg, 24. Dezember 2017

Die vorliegende Arbeit wurde im Zeitraum von April 2014 bis Februar 2017 an der Hamburg School of Food Science des Fachbereichs Chemie der Fakultät für Mathematik, Informatik und Naturwissenschaften der Universität Hamburg unter der Anleitung von Herrn Prof. Dr. Markus Fischer angefertigt.

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

2. Gutachter der Dissertation: Herr Prof. Dr. José A. C. Broekaert

Tag der Disputation: 15.12.2017 Tag der Druckfreigabe: 15.12.2017

# Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei der Erstellung der vorliegenden Arbeit unterstützt haben:

Mein Dank gilt zunächst meinem Doktorvater Herrn Prof. Dr. Markus Fischer für die stetige wissenschaftliche und organisatorische Unterstützung und das mit entgegengebracht Vertrauen. Für die wissenschaftliche Freiheit und dir mir übertragene Verantwortung möchte ich mich zudem ganz besonders bedanken.

Meinen Diplomandinnen Jasmin Wrage, Silke Vatterodt und Eva Reiner für die Unterstützung und Mitarbeit in Teilen dieser Arbeit und die schöne gemeinsame Zeit.

Meinen wunderbaren Kolleginnen und Kollegen für die großartige und unterhaltsame Zeit. Insbesondere René Bachmann und Nicolas Cain gilt mein besonderer Dank für die tolle Unterstützung und Zusammenarbeit und die vielen umfangreichen (wissenschaftlichen) Gespräche. Euch habe ich viele großartige und spaßige Stunden innerhalb und abseits der Arbeitszeit zu verdanken. Ich habe euch als Freunde sehr schätzen gelernt.

Meinem Diplombetreuer und späterem Kollegen Philipp Werner für das geteilte Wissen und die angenehme Einarbeitung in die komplexe Thematik.

Abschließend danke ich allen, die mich während meiner gesamten Promotion und dem Weg dorthin besonders unterstützt haben:

Meiner Familie, die zu jeder Zeit für mich da ist und mich in meinem Handeln zweifelsfrei unterstützt.

Meiner Freundin Christin für ihren unerschütterlichen moralischen Rückhalt und ihre bedingungslose Liebe. Mit dir gemeinsam durch diese Zeit zu gehen war eine großartige Unterstützung und hat diesen Erfolg erst möglich gemacht!

### I Publikationsliste

Klockmann, S.; Reiner, E.; Bachmann, R.; Hackl, T.; Fischer, M. Food Fingerprinting: *Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS.* J. Agric. Food Chem. **2016**, *64*, 9253-9262.

Klockmann, S.; Reiner, E.; Cain, N.; Fischer, M. Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application. J. Agric. Food Chem. 2017, 65, 1456-1465.

Fischer, C.; Kallinich, C.; Klockmann, S.; Schrader, J.; Fischer, M. Automated Enrichment of Sulfanilamide in Milk Matrices by Utilization of Aptamer-Linked Magnetic Particles. J. Agric. Food Chem. **2016**, *64*, 9246-9252.

Fischer, C.; Klockmann, S.; Wessels, H.; Hünniger, T.; Schrader, J.; Paschke-Kratzin, A.; Fischer, M. *Aptamer-based trapping of phytosphingosine in urine samples*. J. Biotechnol. **2016**, *238*, 30-34.

Fischer, M.; Creydt, M.; Felbinger, C.; Fischer, C.; Klockmann, S.; Werner, P.; Klare, J.; Hünniger, T.; Hackl, T. *Food Profiling – Strategien zur Überprüfung der Authentizität von Rohstoffen*. Journal für Verbraucherschutz und Lebensmittelsicherheit. **2014**, *9*, 400-404.

Bachmann, R.; Klockmann, S.; Fischer, M.; Hackl, T. Metabolic Profiling: Geographische Herkunftsanalytik von Haselnüssen mittels NMR-Spektroskopie. Dtsch. Lebensm.-Rundsch.
2017, 113, 6-11.

# II Inhaltsverzeichnis

Da	DanksagungI						
I	Publika	Publikationsliste II					
II	Inhaltsv	Inhaltsverzeichnis					
III	Abkürzı	Abkürzungsverzeichnis					
1	Zusamn	nenfassung	1				
2	Abstrac	t					
3	Einleitu	ng	5				
	3.1 Ha	aselnuss	5				
	3.2 M	etabolomics	12				
	3.2.1	Definition	12				
	3.2.2	Probenvorbereitung und zu berücksichtigende Faktoren	13				
	3.2.3	Apparative Methoden	16				
	3.2.4	Food Metabolomics	17				
	3.3 LO	C-MS	20				
	3.3.1	Flüssigchromatographie	20				
	3.3.1	.1 Umkehrphase	22				
	3.3.1	.2 Silica Hydrid Phase					
	3.3.2	Massenspektrometrie	24				
	3.3.2	2.1 Electrospray Ionisation	24				
	3.3.2	2.2 Triple Quadrupol Massenspektrometer	26				
	3.3.2	2.3 Time-of-Flight Massenspektrometer	29				
	3.3.2	2.4 Kopplung von Flüssigchromatographie und Massenspektrometrie	31				
	3.4 M	ultivariate Datenanalyse	31				
	3.4.1	Datenvorbearbeitung	33				
	3.4.2	Erstellung und Bewertung von Klassifizierungs- und Vorhersagemodellen	ı 34				
	3.4.3	Multivariate Analysenmethoden	36				
	3.4.3	Multiple Lineare Regression (MLR)	36				
	3.4.3	Hauptkomponentenanalyse (PCA)	37				
	3.4.3	Partial Least Squares Regression bzw. Partial Least Squares Discri	minant				
		Analysis (PLS-R bzw. PLS-DA)	39				
	3.4.3	S.4 Soft Independent Modelling of Class Analogy (SIMCA)	39				
	3.4.3	Lineare Diskriminanzanalyse (LDA)	41				

	3.4.3.6 Support Vector Machines (SVM)	42				
4	Zielsetzung der Arbeit					
5	Ergebnisse und Diskussion					
	5.1 Methodenentwicklung	47				
	5.1.1 Non-targeted LC-MS	48				
	5.1.1.1 Methodenentwicklung für die Analyse von unpolaren Metaboliten	48				
	5.1.1.1.1 UPLC-ESI-QTOF-MS-Methode	49				
	5.1.1.1.2 Extraktionsmittel	49				
	5.1.1.1.3 Extraktionsmethode	54				
	5.1.1.1.4 Membranfiltration und Stabilität der Extrakte	57				
	5.1.1.2 Methodenentwicklung für die Analyse von polaren Metaboliten	60				
	5.1.1.2.1 UPLC-ESI-QTOF-MS-Methode	60				
	5.1.1.2.2 Extraktionsmittel, Stabilität der Extrakte und Membranfiltration	61				
	5.1.1.2.3 Extraktionsmethode	62				
	5.1.2 Targeted LC-MS	63				
	5.1.2.1 Extraktion	63				
	5.1.2.2 Flüssigchromatographische Trennung	63				
	5.1.2.3 Massenspektrometrische Detektion	65				
	5.1.2.4 Validierung	65				
	5.2 Geographische Herkunftsdiskriminierung mittels non-targeted LC-MS	70				
	5.2.1 Vergleich der Methoden NTNP-pos, NTNP-neg, NTP-pos und NTP-neg	70				
	5.2.2 Analyse der authentischen Haselnussproben von 2014 und 2015	unter				
	Anwendung von NTNP-pos	75				
	5.2.2.1 Bestimmung der Markersubstanzen	78				
	5.2.2.2 Strukturaufklärung der Markersubstanzen	82				
	5.2.2.3 Metabolit-Verteilung in unterschiedlichen Ländern	84				
	5.3 Geographische Herkunftsdiskriminierung mittels targeted LC-MS	87				
	5.3.1 Motivation und Zielsetzung	87				
	5.3.2 Quantifizierung der Markersubstanzen	89				
	5.3.3 Vergleich mit Literaturwerten	97				
	5.4 Vorhersagemodelle und Klassifizierung nicht-authentischer Proben	98				
	5.4.1 Non-targeted LC-MS	98				
	5.4.2 Targeted LC-MS	103				
	5.5 Messung von gemischten Haselnussproben	110				

	5.6	Über	berprüfung der Lagerstabilität 1					
	5.7	Schlu	Schlussfolgerungen und zukünftige Herausforderungen					
6	Mate	Material und Methoden						
	6.1	.1 Proben						
	6.1.	1 P	Probenmaterial					
	6.1.2	2 P	robenaufarbeitung	125				
	6.1.	3 E	Extraktion	125				
	6.	.1.3.1	Initialmethode zur Methodenoptimierung	125				
	6.	.1.3.2	Extraktion der unpolaren Metabolite für die non-targeted Analyse	125				
	6.	.1.3.3	Extraktion der unpolaren Metabolite für die targeted Analyse	126				
	6.	.1.3.4	Extraktion der polaren Metabolite für die non-targeted Analyse	126				
	6.	.1.3.5	Qualitätskontrolle	126				
	6.2	Stam	mlösungen und Multistandards	126				
	6.2.	1 N	Aarkersubstanzen	126				
	6.2.2	2 S	tammlösungen	127				
6.2.3 Mu		3 N	Aultistandards	127				
6.2.3.1 Methodenentwicklung und -validierung		.2.3.1	Methodenentwicklung und -validierung	127				
6.2		.2.3.2	3.2 Interner Standard Mix					
	6.	.2.3.3	Extrakt-Verdünnungsreihe	129				
	6.2.4	4 E	Blindwert	130				
	6.3	Non-	Targeted LC-MS (UPLC-ESI-QTOF-MS)	130				
	6.3.	1 L	C-MS Methoden für die Messung der unpolaren Metabolite (NTNP-p	os und				
		N	NTNP-neg)	130				
	6.3.2	2 L	C-MS Methoden für die Messung der polaren Metabolite (NTP-pos une	d NTP-				
		n	eg)	131				
	6.3.	3 K	Kalibration	133				
	6.3.4	4 A	Auswertung	133				
	6.3.	5 P	Probenmessungen	135				
	6.3.0	6 S	oftware-basierte Identifizierung von Markersubstanzen zur geograph	nischen				
		H	Ierkunftsbestimmung	135				
	6.3.	7 S	trukturaufklärung relevanter Metabolite	135				
	6.3.8	8 N	Aultivariate Datenanalyse	136				
	6.4	Targ	eted LC-MS (HPLC-ESI-QqQ-MS/MS)	137				
	6.4.	1 T	Cargeted LC-MS Methode	137				

	6.4.2	Validierung				
	6.4.3					
	6.4.4	Berechnung der Korrekturfaktoren				
	6.4.5	Probenmessungen				
	6.4.6	Mischungen				
	6.4.7	Univariate und multivariate Datenanalyse	144			
	6.5 La	gerungsversuche				
7	7 Literaturliste					
8	Anhang					
	8.1 Material					
	8.1.1	Chemikalienverzeichnis				
	8.1.2	Verzeichnis über verwendete Lösungen und Puffer				
	8.1.3	Verzeichnung über verwendete Geräte und Software				
8.1.4 Verzeichnis über verwendete Verbrauchsmaterialien						
	8.1.5 Formeln zur statistischen Auswertung					
	8.2 Er	gänzungen zum Abschnitt 5 – Ergebnisse und Diskussion	177			
9	Eidesstattliche Versicherung					

# III Abkürzungsverzeichnis

Ø	Durchschnitt
%	Prozent
°C	Grad Celsius
α	Irrtumswahrscheinlichkeit
γ	Kernel Parameter
μ	Micro
A	Acetonitril
А	Peakfläche
ACN	Acetonitril
acyl	Fettsäurerest
AF	Ammoniumformiat
ANOVA	Analysis of Variance
ANP	wässrige Normalphase (engl.: aqueous normal phase)
API	Atmosphärendruckionisation (engl.: <i>atmospheric pressure ionization</i> )
b	Mittleres Begleitsignal
bar	Bar
BHT	Butylhydroxytoluol
с	Konzentration
с	Centi
С	Soft Margin Parameter
С	Chloroform
С	Kohlenstoff
CAS	Chemical Abstracts Service
CAWG	Chemical Analysis Working Group
CE	Kapillarelektrophorese (engl.: <i>capillary electrophoresis</i> )
CE	Kollisionsenergie (engl.: collision energy)
CEP	Collision Cell Entrance Potential
CH <sub>3</sub> OH	Methanol
CH <sub>3</sub> COOH	Ethanol
CHCl <sub>3</sub>	Chloroform
CID	Collision Induced Dissociation
Cl	Chlor
$CO_2$	Kohlenstoffdioxid
counts	Zähleinheiten pro Zeitintervall
CRM	Charge Residue Model
CV	Variationskoeffizient
СХР	Collision Cell Exit Potential
d6 bzw. d7	6-fach bzw. 7-fach deuteriert
Da	Dalton
DAD	Diodenarraydetektor
DAG	Directed Acyclic Graph
DE	Deutschland
DG	Diacylglycerol
DIN	Deutsches Institut für Normung
DP	Declustering Potential
E	Einwaage
E	Energie

eV	Elektronenvolt
EO	Erzeugerorganisation
EP	Entrance Potential
ES	Spanien
ESI	Electrospray Ionization
EtOH	Ethanol
EU	Europäische Union
FA	Ameisensäure (engl.: <i>formic acid</i> )
FA	Fettsäure (engl.: <i>fatty acid</i> )
FDA	Food and Drug Administration
FDR	False Discovery Rate
FID	Flammenionisationsdetektor
FP	Focussing Potential
FR	Frankreich
FT	Fourier Transformation
FWER	Family-wise Error Rate
g	Gramm
σ	Erdbeschleunigung (= 9.81 m·s <sup>-2</sup> )
5 G	Gehalt
σσΑ	geographisch geschützte Angabe
σΙ	geschützte Ursnrungsbezeichnung
GC	Gaschromatogranhie
GE	Georgien
GHS	Globally Harmonized System
Н	Wasserstoff
H <sub>2</sub>	elementarer Wasserstoff
Н2	A meisensäure
HILIC	Hydronhilic Interaction Liquid Chromatography
HMDR	Human Metabolome Database
H	Wasser
Höhe üher NN	Höhe über Normal Null
HDC	High Precision Calibration
HDI C	Hochlaistungeflüssigehromatographia (angl: high norformance liquid
	abromatography)
Ца	Hortz
I IZ I	Isopropagal
ICD	induktiv gekonneltes Plasma (angl: inductive counled plasma)
ICP	Fourier Transform Ion Cyclotron Resonance
IEM	Ion Evanoration Model
	Isopropagal
	Isopropanoi
IQK ID	Inferrent (snaktroskonia)
IK IS	Interner Stenderd
IS IT	Italian
11 1 <sub>2</sub>	Vila
K V	Kilo
N Ironl	Kallulli
NUAI VE	Kilokalolitil
KI' VM	Kunalmühla
Γ.IVI I	
	Linu Linaarar Arbaitsharajah
LA	

LC	Flüssigchromatographie (engl.: <i>liquid chromatography</i> )
LDA	Lineare Diskriminanzanalyse
LLOO	Lower Limit of Ouantification ( $\triangleq$ Bestimmungsgrenze)
m	Masse
m	Milli
m	Meter
m	Steigung der Kalibriergeraden
M	Methanol
M	Molare Masse
M	Molar
M	Molekül
m/z	Vorhältnis von Masso zu Ladung
	Maturix Assisted Lasar Desorption Ionization
	Mairix Assisted Laser Desorption Ionization Metabolite Confidence Internal Model
	Melabolile Confidence Interval Model
MeOH .	Methanol
min	Minute
M10.	Millionen
MLR	Multiple Lineare Regression
mol	Mol
MRM	Multiple Reaction Monitoring
MS	Massenspektrometer
MSI	Metabolomics Standards Initiative
n	Nano
n	Anzahl einer Grundgesamtheit
Ν	Stickstoff
$N_2$	elementarer Stickstoff
Na	Natrium
neg	negativ
NH <sub>4</sub>	Ammonium
NH <sub>3</sub>	Ammoniak
NIPALS	Nonlinear Iterative Partial Least Sauares
NIR	Nahinfrarot(-spektroskopie)
NMR	Kernresonanzspektroskopie (engl · <i>nuclear magnetic resonance</i> )
NTP	Non-targeted polar
NTNP	Non-targeted unnolar
NWG	Nachweisgrenze
0	Sauerstoff
D	Dhogphor
r nnm	r Hospiloi Dant Day Million
ppin	Cuedminel
Q	Quadrupol Or aline Contra Dasha
QU	Quality Control Probe
QqQ	
QIOF	Quadrupol Time of Flight
QTrap	Quadrupol-Ionenfalle
р	Pico
Pa	Pascal
PC	Phosphatidylcholin
PC	Hauptkomponente
PCA	Hauptkomponentenanalyse (engl.: principal component analysis)
PE	Phosphatidylethanolamin
PG	Prüfgröße

pН	pondus Hydrogenii
PLS-DA	Partial Least Squares Discriminant Analysis
PLS-R	Partial Least Squares Regression
pos	positiv
psi	Pounds Per Square Inch
PTFE	Polytetraflourethylen
Qal	Qualifier
Qnt	Quantifier
r	Wiederholgrenze
R	PEARSON'scher Korrelationskoeffizient
$\mathbb{R}^2$	Bestimmtheitsmaß
Rt	Retentionszeit
RF	Radio Frequency
RMSE	Root-Mean-Square Error
RP	Umkehrphase (engl.: reversed phase)
rpm	Umdrehungen pro Minute
S	Sekunde
S	Standardabweichung
$S_R$	Reststandardabweichung der linearen Regression
S <sub>RQ</sub>	Reststandardabweichung der quadratischen Regression
$S_V$	Verfahrensstandardabweichung
SIMCA	Soft Independent Modelling of Class Analogy
SVD	Singulärwertzerlegung
SVM-C	Support Vector Machine Classification
SVM-R	Support Vector Machine Regression
t	Tonne
TG	Triacylglycerol
ToF	Flugzeit (engl.: time of flight)
TR	Türkei
UPLC	Ultra High Performance Liquid Chromatography
US	Ultraschall
USA	Vereinigte Staaten von Amerika
USD	US-Dollar
UV	Ultraviolett
V	Geschwindigkeit
$\mathbf{v}/\mathbf{v}$	Volumenanteile
V	Variationsbreite
V	Volt
V	Volumen
VKv	Verfahrensvariationskoeffizient
Vpp	Voltage Peak-to-Peak
WDH	Wiederholung
Z	Ladung

## 1 Zusammenfassung

Haselnüsse sind ein wichtiges internationales Handelsgut mit einer Weltjahresernte von 713.000 t (2014), wobei die wirtschaftlich relevanten Anbaugebiete über diverse Länder verteilt sind. Der mengenmäßig größte Teil stammt aus der Türkei, gefolgt von Italien, Georgien, USA sowie Aserbaidschan. Es bestehen zum Teil erhebliche Qualitäts- und damit verbunden Preisunterschiede für Haselnüsse aus den verschiedenen Anbauländern, sodass eine Verfälschung von Waren für kriminell motivierte Unternehmer zur Gewinnmaximierung potentiell sinnvoll erscheinen kann. Derzeit existieren jedoch keinerlei analytische Methoden zur Authentizitätsprüfung in Bezug auf die geographische Herkunft von Haselnüssen, sodass Verfälschungen nicht valide detektiert werden können.

In der vorliegenden Arbeit erfolgte daher die Entwicklung von massenspektrometrischen Metabolomics-Applikationen zur Bestimmung der geographischen Herkunft von Haselnüssen. Zu diesem Zweck wurden 207 authentische Haselnussproben sowie 59 Haselnussproben der Süßwarenindustrie aus den Erntejahren 2014 und 2015 aus wirtschaftlich relevanten Anbauländern akquiriert.

Zu Beginn erfolgte die Entwicklung von non-targeted UPLC-ESI-QTOF-MS Methoden zum hochaufgelösten Screening von Haselnussproben in Bezug auf das polare und unpolare Metabolom. Beim Vergleich verschiedener Methoden erwies sich die Analyse des unpolaren Metaboloms im positiven Ionenmodus als am besten geeignet für die vorliegende Fragestellung. Es wurden 20 Markersubstanzen mit hochsignifikanten Unterschieden zwischen den verschiedenen Ländern identifiziert. Dabei handelt es sich um Lipide der Stoffklassen Phosphatidylethanolamine, Phosphatidylcholine, Diacylglycerole und Triacylglycerole mit unterschiedlichen Fettsäure-Seitenketten sowie  $\gamma$ -Tocopherol.

Auf Basis dieser 20 Markersubstanzen wurde eine targeted HPLC-ESI-QqQ-MS/MS Methode entwickelt, welche die Anforderungen der Routineanalytik in Überwachungs- und Qualitätssicherungslaboratorien in hervorragender Weise erfüllt, wodurch eine spätere Implementierung ohne großen Aufwand ermöglicht wird. Eine umfangreiche Validierung gemäß den Richtlinien der FDA und der DIN 32645 belegte die Robustheit und Sensitivität der entwickelten Methode, welche mit einer Gesamtanalysenzeit (inkl. aller notwendigen Aufarbeitungs- und Messprozesse) von ca. 30-40 min für eine *high-throughput* Analytik im Metabolomics-Bereich besonders geeignet ist. Abschließend wurden anhand der authentischen Haselnussproben diverse multivariate Datenanalyseverfahren zur Entwicklung von zuverlässigen statistischen Modellen zur Vorhersage des Herkunftslandes unbekannter Proben getestet und evaluiert. Die besten Ergebnisse konnten mit einem Vorhersagemodell mittels *Support Vector Machine Classification* erzielt werden, wobei eine Genauigkeit von 100 % erreicht werden konnte. 80 % der Haselnussproben der Süßwarenindustrie konnten auf diese Weise korrekt vorhergesagt werden. Erste Messungen deuteten darüber hinaus darauf hin, dass diese Methode auch auf geröstete Haselnüsse anwendbar ist.

## 2 Abstract

Hazelnuts are an important commodity in the international trade with an annual world crop size accounting for about 713.000 t (2014), whereas commercially relevant harvest areas are spread across several countries. The main production is provided by Turkey, followed by Italy, Georgia, USA and Azerbaijan. To some extent, the the quality and consequently, the price for hazelnuts from these countries may vary considerably, affording a potential incentive for some criminal entrepreneurs to adulterate products for profit maximization. However, there are still no existing analytical methods for verifying the authenticity related to the geographical origin of hazelnuts and thus, no data is available.

Hence, this work focussed on the development of mass spectrometric metabolomics applications for the determination of the geographical origin of hazelnuts. For this purpose, 207 authentic hazelnut samples and 59 hazelnut samples for confectionary industry were acquired from harvest years 2014 and 2015 out of commercially relevant countries.

Initially, non-targeted UPLC-ESI-QTOF-MS methods for high-resolution screening of the polar and nonpolar metabolome of hazelnuts were developed. After the assessment of these methods, the analysis of the nonpolar metabolome in positive ion mode emerged to be best suited for the present issue. Overall 20 marker substances comprising highly significant differences between the different countries were determined and structurally identified. They belong to the lipid classes phosphatidylethanolamines, phosphatidylcholines, diacylglycerols and triacyclglycerols containing different fatty acid side chains, as well as  $\gamma$ -tocopherol.

Based on these 20 marker substances, a targeted HPLC-ESI-QqQ-MS/MS method was developed that perfectly meets the requirements of routine analytics in surveillance or quality control laboratories. Therefore, this method could be easily implemented into routine analytics. A comprehensive validation in accordance with guidelines of the FDA and DIN 32645 has proven the robustness, reproducibility and sensitivity of this method. The whole analysis process only lasts about 30-40 min, including all sample preparation and analysis steps and thus, it is perfectly suited for high-throughput analytics.

Based on authentic hazelnut samples, several multivariate data analysis methods for creating reliable statistical models predicting the geographical origin of unknown hazelnut samples were tested and evaluated. Best results could be achieved using support vector machine

classification, gaining 100 % accuracy in the training data. Applying this model to the hazelnut samples for confectionary industry, 80 % could be predicted correct. Furthermore, preliminary results indicate a general applicability for roasted hazelnuts, too.

#### 3 Einleitung

#### 3.1 Haselnuss

Haselnüsse sind die einsamigen Nussfrüchte verschiedener Pflanzenarten der Gattung Hasel (Corvlus) aus der Familie der Birkengewächse (Betulaceae). Heutzutage sind zehn verschiedene Spezies bekannt, die sich in ihrer Morphologie unterscheiden können, aber als Gemeinsamkeit essbare Haselnüsse produzieren. Unter dem Begriff "Haselnuss" im wirtschaftlichen Sinn werden gemeinläufig die Früchte der Spezies Corylus avellana verstanden, die den Großteil der kommerziellen Produktion ausmachen. Obwohl die nahverwandte Lambertshasel (Corylus maxima Mill.) früher als eine eigenständige Spezies aufgefasst wurde, zeigen aktuelle Forschungsergebnisse auf Basis von DNA-Sequenz- und Mikrosatteliten-Analysen, dass keine wissenschaftliche Grundlage für die Betrachtung als eigenständige Spezies existiert und daher beide in einer großen, polymorphen Spezies -C. avellana - zusammengefasst werden sollten.<sup>1-3</sup> Nichtsdestotrotz ist die Bezeichnung von Haselnüssen aus dem Schwarzmeergebiet als C. maxima heutzutage noch weit verbreitet und auch in der Wirtschaft häufig anzutreffen.

Die Pflanzen von C. avellana wachsen im Regelfall als vielstämmige, sympodial verzweigte, aufrechte Sträucher von bis zu sechs Metern Höhe, in seltenen Fällen aber auch baumartig bis zehn Meter hoch. In einigen Fällen werden die Pflanzen auf Stämme der Baumhasel (C. colurna) gepfropft. Die Blüten der Hasel sind monözisch verteilt in dichasialen Teilblütenständen. Dabei sind die männlichen Blüten zu vielen in Kätzchen und die weiblichen Blüten zu mehreren in Infloreszenzen vereinigt. Die Bestäubung erfolgt durch den

Wind, wobei die Befruchtung erst mehrere Wochen danach erfolgt. Die Haselnuss ist selbst-inkompatibel, wodurch eine Selbstbestäubung verhindert wird. Daraus ergibt sich die Besonderheit in großen Haselnussplantagen, dass neben dem eigentlichen Kultivar stets zusätzlich vereinzelt in unmittelbarer Nähe kompatible Bestäuberpflanzen von einem bis mehreren Abbildung 3-1: (A) Cupula (Fruchthülle); (B) Nussanderen Kultivaren stehen müssen.



frucht (Haselnuss in-shell); (C) Samen (Haselnuss-In kern); (D) Blatt.

traditionellen, eher kleineren Plantagen erfolgt meist eine Wildbestäubung durch Pflanzen in näherer Umgebung. Aus dem weiblichen Blütenstand entwickeln sich 1 - 4 einsamige Nussfrüchte, die jeweils von einer glockenförmigen, vielfach gerissen gezähnten Fruchthülle (Cupula) umgeben sind.<sup>1, 4-6</sup> Je nach Anbaugebiet und klimatischen Gegebenheiten erreichen die Haselnüsse ihre volle Reife von August bis November. Während die Ernte in der Schwarzmeerregion üblicherweise noch händisch erfolgt, wird in den meisten westlichen Ländern sowie neueren Anlagen mittlerweile eine maschinelle Ernte bevorzugt. Die gesammelten Nüsse werden anschließend durch maschinelle Trocknungsanlagen oder auf Planen im Freien durch die Sonne bis zu einem Restwassergehalt von ca. 3,5 - 4,5 % Kernfeuchtigkeit getrocknet.<sup>7-9</sup> Die getrockneten Nüsse werden geschält als Haselnusskerne oder ungeschält als sogenannte "in-shell" Nüsse verkauft, wobei die geschälte Form mit 79 % im Weltmarkt dominiert.<sup>10</sup>

Die größte Menge der weltweiten Jahresernte (ca. 90 %) wird in der Süßwarenindustrie weiterverarbeitet, meist in gerösteter Form als ganze, gehackte oder gemahlene Nusskerne.<sup>11</sup> Insgesamt belief sich die Weltjahresernte 2014 auf 713.000 t, wobei mit ca. 63 % der größte Anteil von der Türkei produziert wird, gefolgt von Italien (11 %), Georgien, USA, Aserbaidschan, China, Iran, Spanien und Frankreich (jeweils  $\leq 5$  %).<sup>38</sup> Für den deutschen Markt spielen die Haselnüsse aus China und dem Iran allerdings nur eine untergeordnete Rolle, während immer mehr aus deutschen Plantagen stammen.<sup>12</sup> Die Hauptanbaugebiete der jeweiligen Länder fokussieren sich überwiegend auf sehr kleine, lokal begrenzte Gebiete (siehe Abbildung 3-2). In den "klassischen" Anbauländern (Türkei, Italien, Georgien,



Abbildung 3-2: Hauptanbaugebiete der wirtschaftlich relevantesten Anbauländer Europas und der USA.

Aserbaidschan, Spanien) haben sich größtenteils spezifische Haselnusssorten etabliert, die für die jeweiligen Gebiete typisch sind und traditionell bevorzugt angebaut werden<sup>13</sup>, während Länder, in denen der Haselnussanbau erst seit einigen Jahren oder Jahrzehnten forciert wird (z.B. Frankreich, USA, Deutschland und seit kurzem auch Chile und Australien), ein breiteres Spektrum an kultivierten Sorten aufweisen.<sup>14-17</sup> Tabelle 3-1 gibt eine Übersicht über die wirtschaftlich relevantesten Haselnusssorten und deren vorwiegende Anbaugebiete. Dabei ist zu beachten, dass in einigen Ländern der Name der Sorte zum Teil gleichbedeutend mit der Bezeichnung der jeweiligen Region ist (z.B. in Aserbaidschan und Georgien), wobei es sich dann bei den kommerziell erhältlichen Haselnüssen aus diesen Regionen fast ausschließlich um Mischungen diverser Sorten handelt, bedingt durch die Erntepraktiken bei meist noch kleinbäuerlichen Strukturen.<sup>1, 13</sup> Nach der Ernte werden die Haselnüsse zahlreicher Bauern gesammelt, bevor sie von Großbetrieben verarbeitet (geschält, sortiert) und/oder exportiert werden, wobei es meist zur Vermischung der Nüsse kommt.<sup>18</sup> Im Gegensatz dazu ist die Praxis in Italien dahingehend ausgelegt, möglichst sortenreine Haselnüsse zu verarbeiten.

Sorte	Land
Anakliuri	Georgien (Mingrelien und Oberswanetien)
Ata Baba	Aserbaidschan
Barcelona	Deutschland, Frankreich (Korfu, g.g.A.; Aquitanien; Midi-Pyrénées), USA
Butler	Frankreich, USA
Çakildak	Türkei (Ordu)
Corabel	Frankreich
Emoal	Deutschland
Ennis	Frankreich, USA
Foșa	Türkei (Trabzon)
Gulshishvela	Georgien (Gurien)
Karafindik	Türkei (Düzce)
Khachmaz	Aserbaidschan
Lewis	USA
Mincane	Türkei (Trabzon)
Mortarella	Italien (Kampanien, g.g.A.)
Negret	Spanien (Tarragona, g.U.), Frankreich
Palaz	Türkei (Ordu)
Pauetet	Spanien (Tarragona, g.U.), Frankreich
Rote Zellernuss	Deutschland
Segorbe	Deutschland, Frankreich
Shvesliskura	Georgien (Gurien)
Sivri	Türkei (Giresun)
Tombul	Türkei (Giresun)
Tonda Romana	Italien (Lazio, g.U.)
Tonda di Giffoni	Deutschland, Frankreich, Italien (Kampanien, g.g.A.)
Tonda Gentile Trilobata	Italien (Piemont, g.g.A.)

Tabelle 3-1: Wirtschaftlich relevante Haselnusssorten und ihre Hauptanbaugebiete.

So werden in der Region Lazio nahezu ausschließlich Tonda Romana und in Piemont nur Tonda Gentile Trilobata angebaut. Die italienischen Haselnüsse der drei Hauptanbaugebiete werden traditionell mit einer hohen Qualität assoziiert, sodass Produkte dieser Herkunft inzwischen von der Europäischen Union (EU) mit einem Gütesiegel besonders geschützt werden. Die Bezeichnung "Nocciola di Giffoni" und "Nocciola del Piemonte" sind geographisch geschützte Angaben (g.g.A.), wobei Anbau und Ernte in einem begrenzten geographischen Gebiet erfolgen müssen und nur die dort spezifizierten Sorten verwendet werden dürfen. Bei "Nocciola Romana" handelt es sich um eine geschützte Ursprungsbezeichnung (g.U.), bei der neben Anbau und Ernte auch alle weiteren Verarbeitungsschritte (z.B. Sortieren, Schälen, Rösten oder Verpacken) in dem räumlich begrenzten Gebiet erfolgen müssen. Neben diesen Produkten genießen zusätzlich "Noisette de Cervione g.g.A." aus Korsika (Frankreich) und "Avellana de Reus g.U." aus Tarragona (Spanien) besonderen Schutz durch die EU.<sup>19</sup>

Die ölhaltigen Samen der Haselnuss dienen seit prähistorischer Zeit als Nahrungsmittel der Menschen. Sie zeichnen sich durch ihren hohen Nährstoffgehalt von 650 kcal pro 100 g aus, welcher vor allem auf den hohen Fettgehalt von etwa 60 % zurückzuführen ist. Das Fett besteht überwiegend aus ungesättigten Triacylglycerolen. Dabei handelt es sich größtenteils um Öl- und Linolsäure mit Anteilen von 80 % bzw. 14 % an der Summe der Fettsäuren. Daneben enthält die Haselnuss als weitere Hauptinhaltsstoffe 16 % Kohlenhydrate, 14 % Proteine, 5 % Wasser und 2 % Mineralstoffe.<sup>20, 21</sup> Die exakte chemische Zusammensetzung ist allerdings unter anderem von der geographischen Herkunft der Pflanze abhängig und variiert deshalb in Bezug auf die Provenienz. Tabelle 3-2 zeigt eine Übersicht über die stoffliche Zusammensetzung von Haselnüssen aus verschiedenen Anbaugebieten.

Zusätzlich zur geographischen Abhängigkeit wird die chemische Zusammensetzung noch von dem zugrundeliegenden Genotyp (Sorte) beeinflusst. Es ist bereits bekannt, dass bestimmte Haselnusssorten in verschiedenen Gebieten unterschiedlich erfolgreich kultiviert werden können. Aus diesem Grund werden in Gebieten, in dem der Haselnussanbau erst seit einiger Zeit kommerziell ausgebaut wird, umfangreiche Züchtungsversuche durchgeführt. In Sortenversuchsanstalten wird nach denjenigen Sorten gesucht, die an dem jeweiligen Standort am besten wachsen können, resistent gegen etwaige einheimische Krankheiten sind und die höchsten Erträge liefern.

#### Einleitung

Bestandteil	Türkei	Italien	USA	Portugal	Neuseeland	Spanien	Frankreich	Polen
Fett <sup>a</sup>	$61,2 \pm 1,0$	$64,1 \pm 0,4$	$60,8 \pm 0,4$	64,0	58,4	56,1	-	$5\overline{6,5\pm 2,3}$
				(59,2 - 69,0)	(54,6 - 63,2)	(49,3 - 65,4)		
Proteine <sup>a</sup>	$15,4 \pm 0,4$	$13,7 \pm 0,5$	$15,0 \pm 0,2$	10,9	16,3	-	-	-
				(9,3 - 12,7)	(14,3 - 18,2)			
Kohlenhydrate <sup>a</sup>	$17,3 \pm 0,5$	$15,5 \pm 0,6$	$16,7 \pm 0,6$	17,6	17,5	-	-	-
				(12,1 - 21,1)	(15,6 - 20,5)			
Palmitinsäure <sup>b</sup>	$4,9 \pm 0,0$	5,6	$5,8 \pm 0,0$	5,6	4,9	5,8	-	$4,5 \pm 0,1$
		(5,0 - 6,0)		(4,8 - 6,8)	(4,1 - 5,9)	(5,4 - 6,5)		
Stearinsäure <sup>b</sup>	$2,7 \pm 0,0$	2,7	$3,1 \pm 0,1$	2,7	1,9	2,0	-	$1,8 \pm 0,1$
		(2,5 - 2,9)		(2,1 - 3,7)	(1,6 - 2,0)	(1,6 - 2,3)		
Ölsäure <sup>b</sup>	$82,7\pm0,0$	81,6	$83,0 \pm 0,0$	80,4	77,5	79,1	-	$80,8 \pm 0,5$
		(78,9 - 82,8)		(76,7 - 82,8)	(73,8 - 80,1)	(77,1 - 84,1)		
Linolsäure <sup>b</sup>	$8,9\pm0,0$	9,5	$7,6 \pm 0,0$	9,2	13,7	12,6	-	$12,1 \pm 0,5$
		(8,2 - 12,8)		(7,2 - 11,4)	(12,0 - 16,5)	(6,6 - 15,0)		
Stigmasterol <sup>c</sup>	1,5 - 2,8	2,0 - 5,4	1,0	$1,9 \pm 0,1$	1,7	1,5 - 1,9	1,4 - 2,7	0,8
			(0,2 - 3,3)		(1,5 - 2,3)			(0,7 - 0,9)
Clerosterol <sup>c</sup>	1,8 - 3,8	1,8 - 2,8	-	$1,2 \pm 0,1$	-	1,9 - 4,0	1,5 - 2,4	1,0
								(0,7 - 1,3)
β-Sitosterol <sup>c</sup>	84,5 - 143,1	116,4 - 138,8	105,0	$156 \pm 0,2$	151,5	92,9 - 149,1	112,5 - 161,2	116,9
			(82,6 - 130,6)		(141,6 - 169,3)			(107,5 - 126,2)
Campesterol <sup>c</sup>	5,8 - 10,7	6,7 - 8,1	6,8	$10,1 \pm 0,2$	9,1	5,9 - 9,7	7,3 - 10,1	7,8
			(5,3 - 9,4)		(7,8 - 11,4)			(6,7 - 8,9)
α-Tocopherol <sup>c</sup>	$40,4 \pm 0,6$	11,8 - 44,0	33,5	29,0	33,9	31,4 - 59,0	34,3 - 54,5	$33,6 \pm 0,0$
			(21,1 - 44,3)	(15,9 - 38,8)	(19,9 - 40,9)			
β-Tocopherol <sup>c</sup>	$1,5 \pm 0,0$	n.d 1,0	-	1,2	1,1	n.d 1,5	n.d 1,0	$1,2 \pm 0,0$
				(0,5 - 2,2)	(0,6 - 1,7)			
γ-Tocopherol <sup>c</sup>	$8,3 \pm 0,2$	n.d 1,0	-	0,8	4,5	n.d 4,2	n.d 1,0	$3,4 \pm 0,2$
				(0,2 - 3,3)	(1,9 - 14,9)			
n d · night datakti	arhar							

Tabelle 3-2: Übersicht über einige wesentliche Inhaltsstoffe von Haselnüssen aus verschiedenen Anbauländern; angegeben als Mittelwerte mit Konfidenzintervallen ( $\alpha = 0,05$ ) sofern vorhanden, ansonsten als Schwankungsbereiche.<sup>20, 22-28, 214</sup>

n.d.: nicht detektierbar

<sup>a</sup>g/100 g

<sup>b</sup>g/100 g extrahiertes Öl

<sup>c</sup>mg/100 g extrahiertes Öl

Bestandteil	Barcelona	Butler	Ennis	Negret	Pauetet	Tonda di Giffoni	Tonda Gentile Trilobata
Fett <sup>a</sup>	$62,7 \pm 0,2$	$61,8 \pm 1,9$	$60,2 \pm 0,1$	$69,0 \pm 0,2$	$65,7 \pm 0,1$	$65,3 \pm 0,1$	$66,7 \pm 0,0$
Proteine <sup>a</sup>	$10,9\pm0,1$	$11,3 \pm 0,1$	$9,5 \pm 0,1$	$10,6\pm0,0$	$10,7\pm0,1$	$9,3 \pm 0,2$	$10,9 \pm 0,1$
Kohlenhydrate <sup>a</sup>	$17,5 \pm 0,4$	$19,9 \pm 1,5$	$21,1 \pm 0,1$	$14,5 \pm 0,3$	$15,5 \pm 0,1$	$16,4 \pm 0,2$	$15,8 \pm 0,0$
Wasser <sup>a</sup>	$6,3 \pm 0,2$	$4,4 \pm 0,0$	$6,4 \pm 0,0$	$3,5 \pm 0,2$	$5,6 \pm 0,3$	$6,3 \pm 0,0$	$4,2 \pm 0,0$
Mineralstoffe <sup>a</sup>	$2,7 \pm 0,1$	$2,7 \pm 0,0$	$2,8 \pm 0,1$	$2,4 \pm 0,0$	$2,6 \pm 0,0$	2,6 ± 0,1	$2,4 \pm 0,0$
Palmitinsäure <sup>b</sup>	$5,4 \pm 0,0$	$5,8 \pm 0,0$	6,1 ± 0,0	$5,1 \pm 0,0$	$5,8 \pm 0,0$	$5,1 \pm 0,0$	$6,0 \pm 0,0$
Stearinsäure <sup>b</sup>	$2,2 \pm 0,0$	$3,1 \pm 0,0$	$2,8 \pm 0,0$	$2,6 \pm 0,0$	$2,5 \pm 0,0$	$2,5 \pm 0,0$	3,0 ± 0,0
Ölsäure <sup>b</sup>	$79,1 \pm 0,1$	$79,7 \pm 0,1$	$80,4 \pm 0,1$	$79,8 \pm 0,3$	$79,3 \pm 0,0$	$80,1 \pm 0,1$	$80,0 \pm 0,1$
Linolsäure <sup>b</sup>	$11,4 \pm 0,0$	$9,1 \pm 0,0$	$8,4 \pm 0,0$	$10,4 \pm 0,0$	$9,9 \pm 0,0$	$10,3 \pm 0,0$	$8,8 \pm 0,0$
Stigmasterol <sup>b</sup>	$1,60 \pm 0,1$	$1,4 \pm 0,0$	$2,0 \pm 0,0$	$1,7 \pm 0,0$	$2,3 \pm 0,1$	$1,8 \pm 0,0$	$1,4 \pm 0,0$
Clerosterol <sup>b</sup>	$1,7 \pm 0,1$	$1,5 \pm 0,0$	$1,6 \pm 0,0$	$1,5 \pm 0,1$	$2,3 \pm 0,1$	$2,0 \pm 0,2$	$1,8 \pm 0,0$
β-Sitosterol <sup>b</sup>	$136 \pm 0$	$132 \pm 0$	$154 \pm 0$	$153 \pm 0$	$153 \pm 0$	$163 \pm 0$	$108 \pm 0$
Campesterol <sup>b</sup>	$9,5 \pm 0,1$	$7,8 \pm 0,1$	$8,1 \pm 0,1$	$9,9 \pm 0,0$	$11,0 \pm 0,2$	$9,7 \pm 0,1$	$7,2 \pm 0,0$

Tabelle 3-3: Übersicht über einige wesentliche Inhaltsstoffe von verschiedenen wirtschaftlich relevanten Haselnusssorten aus einer Sortenversuchsanstalt in Portugal; angegeben als Mittelwerte mit Konfidenzintervall ( $\alpha = 0,05$ ).<sup>36</sup>

<sup>a</sup>g/100 g

<sup>b</sup>g/100 g extrahiertes Öl

Diese Selektionsverfahren werden unter anderem in Deutschland<sup>15</sup>, Slowenien<sup>29</sup>, Chile<sup>16</sup>, Südafrika<sup>30</sup> und Australien<sup>31</sup> betrieben. Treibende Kraft für den Ausbau des Haselnusswirtschaft in den drei letztgenannten Ländern ist der weltweit größte Verarbeiter von Haselnüssen (Ferrero), der mit den eigens dafür gegründeten Firmen AgriChile (seit 1991), AgriSudafrica (seit 2009) und AgriAustralia (seit 2011) lokale Farmer durch Forschungsprojekte und agrarwirtschaftlicher Beratung fördert.<sup>32</sup> Für die Bewertung der Eignung einer Haselnusssorte für eine Kultivierung werden dabei vor allem folgende Faktoren berücksichtigt: Ertrag, Aromaprofil, Form und Größe, Schälbarkeit und Krankheits-/Verderbnisanfälligkeit.<sup>3, 10, 33, 34</sup> Tabelle 3-3 zeigt eine Übersicht über kompositionelle Unterschiede von verschiedenen Haselnusssorten.

Da verschiedene Anbauregionen bevorzugt regionaltypische Sorten anbauen, folgt daraus eine starke herkunftsabhängige Variation des Metaboloms der Haselnusskerne. Obwohl die "Piemonteser" Haselnuss (Nocciola del Piemonte g.g.A.) in der Süßwarenindustrie bereits seit Mitte des 19. Jahrhunderts ein "unumstrittenes Prestige" besitzt und häufig als aromatischste Haselnuss der Welt bezeichnet wird, scheiterten Versuche, die zugehörige Sorte Tonda Gentile Trilobata in anderen Gebieten einzuführen, weil sich "die Sorte nicht anpassen konnte



Abbildung 3-3: Erzeugerpreise für Haselnüsse mit Schale.<sup>38</sup>

bzw. nicht die Waren- und organoleptischen Merkmale entwickelte, die diese Haselnuss in ihrem Herkunftsgebiet auszeichnen".<sup>37</sup> Dies verdeutlicht die Relevanz der geographischen Herkunft für die chemische Zusammensetzung der Haselnuss bei identischem Genotyp.

Zusätzlich zur genetischen und geographischen Abhängigkeit wird die chemischen Zusammensetzung und dadurch resultierend die organoleptischen Eigenschaften zusätzlich noch durch agrarwirtschaftliche Aspekte (Kultivierung, Ernte, Nachernte) beeinflusst.<sup>18</sup> Durch die herkunftsbedingten Unterschiede sowie etwaige Qualitätssiegel wie g.g.A. und g.U. variieren die zu erzielenden Preise je nach Anbauland. Während Italien 2014 durchschnittlich einen Erzeugerpreis von 5.191 USD/t (ungeschälte Haselnüsse) erzielen konnte, lagen die Preise des Weltmarktführers Türkei um 18 % und die der restlichen Länder zwischen 24 % und 56 % niedriger. Im Jahr 2015 wiederum lag die Türkei mit 5.526 USD/t an der Spitze, während die anderen Länder 14 % bis 68 % günstiger waren. (Abbildung 3-3) Dabei ist seit Jahren ein kontinuierlicher Anstieg der durchschnittlichen Haselnusspreise zu beobachten, wobei die Preisspanne zwischen den unterschiedlichen produzierenden Ländern ebenfalls stetig wächst.<sup>38</sup> Solche preislichen Unterschiede sind im Allgemeinen ein Anreiz für kriminell motivierte Lebensmittelverfälschungen z.B. durch Strecken oder Ersetzen von hochwertigen Produkten durch niedrigpreisige Ware um eine höhere Marge erzielen zu können. Die bisherige Praxis der Überprüfung der geographischen Herkunft verläuft anhand der Nachverfolgung von Lieferwegen auf Basis von Lieferdokumenten, welche allerdings durch geschickte Manipulationen verhältnismäßig einfach überlistet werden kann. Aus diesem Grund werden analytische Methoden benötigt, mit denen die geographische Herkunft

zweifelsfrei nachgewiesen werden kann. Einen möglichen Ansatz stellen dabei Metabolomics-Applikationen dar.

#### 3.2 Metabolomics

#### 3.2.1 Definition

Der Begriff Metabolomics wurde erstmals 2001 von Oliver Fiehn vorgeschlagen und beschreibt die Charakterisierung und Quantifizierung möglichst aller Metabolite eines Organismus.<sup>39</sup> Metabonomics hingegen wurde bereits 1999 durch Nicholson als quantitative Messung der dynamischen, multiparametrischen Änderung von lebenden Systemen als Antwort auf pathophysiologische Stimuli oder genetische Modifikationen definiert.<sup>40</sup> In der Praxis werden diese beiden Begriffe jedoch meistens synonym verwendet.<sup>41</sup> Im Folgenden wird allerdings einheitlich der Begriff Metabolomics verwendet.

Metabolomics gehört neben Genomics, Transkriptomics und Proteomics zu den vier Hauptgruppen der "-omics"-Disziplinen, welche sich als Teil eines als "Systems Biology" bezeichneten Forschungszweiges mit der Studie verschiedener Kollektive von Molekülen, biologischen Prozessen oder physiologischen Funktionen und Strukturen als System und deren Veränderungen durch äußere Einflüsse beschäftigen.<sup>42</sup> Der Phänotyp eines Organismus resultiert entsprechend der Omics-Kaskade aus der konsekutiven Abfolge der genetischen Expression, ausgehend vom Genom über das Transkriptom und Proteom bis hin zum Metabolom gemäß des zentralen Dogmas der Molekularbiologie (Abbildung 3-4).



Abbildung 3-4: Erweitertes zentrales Dogma der Molekularbiologie als Omics-Kaskade.

Metabolite sind per Definition kleine Moleküle (typischerweise < 1000  $Da^{43, 44}$ ), die ein Organismus selbst synthetisiert oder aufnimmt, metabolisiert oder katabolisiert. Für die Klassifiierung eines Stoffes als Metabolit müssen bestimmte Kriterien erfüllt sein<sup>45</sup>:

- Metabolite sind chemische Verbindungen, die sich innerhalb von Zellen befinden
- Metabolite werden durch Enzyme erkannt und umgesetzt
- Produkte von Metaboliten müssen in Folgereaktionen eintreten können
- Metabolite haben meist eine begrenzte biologische Halbwertszeit
- Viele Metabolite sind Regulatoren, die Gleichgewichte beeinflussen und regeln
- Metabolite übernehmen nützliche biologische Funktionen in Zellen

Dazu zählen unter anderem Lipide, Aminosäure, Nukleinsäuren, Kohlenhydrate, organische Säuren, Vitamine, Mineralstoffe, Steroide, sekundäre Pflanzenstoffe und Toxine.<sup>46</sup> Die Gesamtheit aller Metabolite eines Organismus wird als Metabolom bezeichnet. Es wird davon ausgegangen, dass im Pflanzenreich etwa 0,2 - 1 Mio. verschiedene Metabolite existieren.<sup>47, 48</sup>

#### 3.2.2 Probenvorbereitung und zu berücksichtigende Faktoren

Neben der genetischen Prädisposition eines Organismus, wird dessen Metabolom zusätzlich wesentlich direkt oder indirekt durch exogene Faktoren beeinflusst. Bei Pflanzen im Allgemeinen, die in der Regel ortsgebunden sind, zählen hauptsächlich Temperatur, Wasserverfügbarkeit, Sonnenstunden, Bodenbeschaffenheit, Höhe über NN und Alter zu diesen Faktoren, da sich der Stoffwechsel der Pflanze den entsprechenden Möglichkeiten und Gegebenheiten anpasst (siehe Abbildung 3-5).<sup>49-51</sup> Im Fall von landwirtschaftlich genutzten Pflanzen spielen darüber hinaus agrarwirtschaftliche Aspekte wie z.B. Pestizidbehandlung, Düngung, künstliche Bewässerung, Wachstumskontrolle, Schutzmaßnahmen (z.B. Folienbedeckung) oder die Züchtung in Glashäusern eine Rolle.<sup>51-53</sup> Viele dieser Faktoren können abhängig von klimatischen und agrarwirtschaftlichen Bedingungen jährlichen Schwankungen unterworfen sein.54, 55 Bei Haselnüssen im Speziellen als Produkt der geschlechtlichen Fortpflanzung könnte darüber hinaus die Sorte der Bestäuberpflanze als männlicher Part des Samen-Genoms auf genetischer Ebene Einfluss nehmen. Auch die einzelnen Schritte im Workflow eines Metabolomics-Experiments haben direkten Einfluss auf dessen Ergebnisse und sind daher von essentiellem Stellenwert. Die Schritte unterteilen sich in Ernte und Nacherntebehandlung, Probenvorbereitung, Extraktion und apparative Analytik.



Abbildung 3-5: Einflussfaktoren auf das Metabolom von Nutzpflanzen.

Schon ab dem Moment der Ernte können sich diverse Faktoren in Veränderungen des Metaboloms widerspiegeln. Es ist z.B. bekannt, dass bereits der Tag und sogar die Uhrzeit der Ernte zu signifikanten Unterschieden in Metabolit-Gehalten führen können.<sup>56, 57</sup> Je nachdem, an welcher Position im Strauch bzw. in der Plantage sich eine Haselnuss befindet, erhält sie mehr oder weniger Sonnenlicht, Wind, Regen oder Schadstoffe, wodurch sich selbst innerhalb eines Haselnussstrauches Unterschiede ergeben können.<sup>43</sup> Im weiteren Verlauf übt sowohl die Trocknung der Nüsse (Sonne oder maschinell) sowie die Lagerung generell einen erheblichen Einfluss auf das Metabolom aus. Nach der Ernte bleiben viele Stoffwechselprozesse aktiv, sodass sich das Metabolom kontinuierlich verändert. Aus diesen Gründen sollte das Probenmaterial für Metabolomics-Analysen idealerweise unter standardisierten Bedingungen geerntet und sofort nach der Ernte mit Flüssigstickstoff bei -196 °C schockgefroren werden, um jegliche Stoffwechselaktivität unverzüglich zu unterbinden.<sup>58, 59</sup> In der Praxis ist diese Vorgehensweise allerdings aus logistischen und finanziellen Gründen oft nicht bzw. nur sehr schwer realisierbar. Darüber hinaus spiegelt diese Prozedur bei der vorliegenden Fragestellung nicht die marktrealen Bedingungen wider. Im Gegensatz dazu sind z.B. bei medizinischbiologisch motivierten Metabolomics-Experimenten jegliche metabolische Veränderungen nach der Probenahme unerwünscht, da die Studien häufig dem Verständnis von physiologischen oder zellulären Prozessen dienen und daher konservierte metabolische Zustände zum Zeitpunkt der Probenahme analysieren werden müssen.<sup>60</sup> Vielmehr sollte die Entwicklung

#### Einleitung

von Methoden zur Authentizitätsbestimmung von Lebensmitteln unter Berücksichtigung realer Praktiken im Laufe der Wirtschaftskette möglichst realistisch erfolgen, damit sie überhaupt für Proben im Zuge der Lebensmittelüberwachung oder –kontrolle anwendbar sind. Daraus folgt allerdings, dass wegen der relativ hohen möglichen Variabilität der Metabolome ein ausreichend großer statistischer Umfang an Probenmaterial zugrunde gelegt werden sollte, damit die Gesamtheit der Haselnussproben des jeweiligen Standortes erfasst wird und das Ergebnis nicht durch eine biologische Verzerrung verfälscht wird. (z.B., weil alle Haselnüsse von einem Strauch in randständiger, schattiger und besonders feuchter Position gesammelt wurden)

Nach der Probenahme muss das Probenmaterial im nächsten Schritt (Probenvorbereitung) sinnvoll homogenisiert werden, um repräsentative Messungen durchführen zu können. Idealerweise erfolgt die Homogenisierung ebenfalls unter gekühlten Bedingungen um unerwünschte Reaktionen durch die entstehende Reibungswärme gering zu halten. Bei sehr fetthaltigen Matrices wie der Haselnuss bietet dies darüber hinaus den Vorteil, dass das Homogenisat keine breiige Konsistenz aufweist, sondern pulverförmig bleibt. Die größere Oberfläche verbessert gleichzeitig den Extraktionsprozess.<sup>61</sup> Eine weitere bzw. zusätzliche Möglichkeit zur Konservierung des Metaboloms stellt die Gefriertrocknung des Probenmaterials dar, wobei durch den Entzug von Wasser die Enzymaktivität verringert werden kann. Der Nachteil ist allerdings, dass die Extraktionseffizienz durch eine größere Adsorption von Metaboliten am Pflanzenmaterial verringert werden kann.<sup>62</sup>

Die Extraktion als nächster Schritt im Workflow ist ein weiterer essentieller Punkt, der die Resultate von Metabolomics-basierten Experimenten maßgeblich beeinflusst. Sie sollte möglichst nur von kurzer Dauer sein um möglicherweise auftretende biochemische Prozesse zu verhindern.<sup>59</sup> Generell sollte der Extraktionsprozess für eine größtmögliche Reproduzierbarkeit so schnell und so einfach wie möglich durchgeführt werden und gleichzeitig die Zersetzung, Veränderung und der Verlust von Metaboliten verhindert werden.<sup>63, 64</sup> Im Metabolomics Bereich ist die Flüssig-Extraktion derzeit am weitesten verbreitet, wobei die exakte Prozedur und Extraktionsmittelzusammensetzung für jede Fragestellung individuell bewertet und optimiert werden muss, da aufgrund der hohen chemischen Diversität kein Lösungsmittel/-gemisch in der Lage ist alle Metabolite gleichzeitig quantitativ zu extrahieren.<sup>56, 61</sup> Zu berücksichtigende Parameter sind dabei die Eigenschaften der Lösungsmittel (Selektivität, Polarität, Siedepunkt, Toxizität, Reinheit und Umweltverträglichkeit), der pH-Wert, das Verhältnis von Extraktionsmittel und Probenmaterial, die Extraktionszeit, die

Temperatur, die Zellaufschlussverfahren (z.B. Ultraschall, Mikrowellen oder Kugelmühle) sowie ggfs. die Pufferkapazität.<sup>59, 64</sup> Üblicherweise werden Methanol, Acetonitril, Chloroform und Wasser sowie Mischungen dieser verwendet.<sup>43, 65-68</sup> Bei der Verwendung eines Gemisches aus Wasser und organischem Lösungsmittel erfolgt die Zugabe des organischen Lösungsmittels üblicherweise zuerst, um eine Proteinpräzipitation zu erreichen, damit eventuell noch aktive Enzyme inaktiviert werden und gleichzeitig Proteine als störende Substanzen in der apparativen Analytik aus dem Extrakt entfernt werden.<sup>69</sup>

#### 3.2.3 Apparative Methoden

Im Gegensatz zu Genomics, Transcriptomics und Proteomics, welche sich mit Makromolekülen ähnlicher chemischer Eigenschaften beschäftigen, zeichnet sich Metabolomics durch eine Vielzahl von unterschiedlichen Verbindungen mit geringem Molekulargewicht und einer großen chemischen Diversität aus.<sup>70</sup> Aus diesem Grund gibt es derzeit keine analytische Plattform, mit der alle Metabolite simultan erfasst werden können.<sup>71</sup> Es haben sich daher verschiedene Applikationen etabliert, von denen die Massenspektrometrie (MS), die Kernresonanzspektroskopie (NMR) und die Infrarotspektroskopie (IR) zu den meistgenutzten Verfahren zählen. Im Fall der Massenspektrometrie ist in der Regel zur Erhöhung der Detektionsleistung eine kapillarelektrophoretische (CE), flüssig- (LC) oder gaschromatographische (GC) Trennung der meist komplexen Stoffgemische vorgeschaltet.<sup>49, 72, 73</sup> Während sich die Massenspektrometrie in Hinsicht auf Metabolomics-Analysen vor allem durch die Eignung für quantitative Analysen bei hoher Selektivität und Sensitivität sowie die Identifizierung von Metaboliten mittels MS/MS-Experimenten auszeichnet, liegt der große Vorteil der NMR-Analytik in dem hohen Probendurchsatz bei minimaler Probenvorbereitung sowie der nicht-destruktiven Detektion. Die Vorteile der Infrarotspektroskopie hingegen sind vor allem die verhältnismäßig niedrigen Gerätekosten sowie der geringe apparative Aufwand und dadurch einfache Handhabung. Zu den Nachteilen des MS zählen insbesondere der teilweise hohe Aufwand für die Probenvorbereitung, die lange Analysendauer (bei Kopplung mit chromatographischer Trennung) sowie die eingeschränkte Reproduzierbarkeit, während die geringere Sensitivität, die komplexen Spektren, die schwierige Identifizierung von einzelnen Metaboliten und die starke Matrixabhängigkeit bei der NMR-Spektroskopie sowie der geringe Informationsgehalt, die Notwendigkeit der Probentrocknung und die fehlende Identifizierbarkeit von Metaboliten bei der Infrarot-Spektroskopie ins Gewicht fallen.<sup>71, 74</sup>

#### Einleitung

Prinzipiell existieren unabhängig von der verwendeten apparativen Methodik zwei verschiedene Verfahren zur Analyse des Metaboloms: Das Metabolic Fingerprinting und das Metabolic Profiling.<sup>75</sup> Beim Metabolic Fingerprinting handelt es sich um ein Verfahren zur qualitativen und relativ quantitativen Erfassung möglichst vieler Metabolite eines biologischen Systems, mit dem Ziel ein charakteristisches Metabolit-Profil ("Fingerprint") zu erhalten und durch den Vergleich der erhaltenen Muster Veränderungen des Metaboloms unterschiedlichen Probenpopulationen (z.B. zwischen geographische Herkunft) festzustellen.<sup>71, 76</sup> Dieser Ansatz wird auch als non-targeted Metabolomics bezeichnet, da keine gezielte Messung von vorher festgelegten Analyten erfolgt.<sup>77</sup> Metabolite, denen ein signifikanter Konzentrationsunterschied der Fragestellung betreffend zwischen unterschiedlichen Probenpopulationen zugrunde liegt, werden als Schlüsselmetabolite oder Markersubstanzen bezeichnet. Eine Identifikation oder absolute Quantifizierung der korrespondierenden Schlüsselmetabolite ist dabei nicht zwingend erforderlich.<sup>43</sup> Allerdings ist es in Hinsicht auf die Anwendung von Routineapplikationen in der Forschung oder Lebensmittelüberwachung sinnvoll, die Struktur von potentiellen Biomarker aufzuklären, um ein besseres Verständnis über die möglichen Ursachen der identifizierten Unterschiede zu erhalten sowie sie im Rahmen von zielgerichteten Assays (Metabolic Profiling) gezielt quantifizieren zu können.<sup>78</sup> Als Metabolic Profiling, oder auch targeted Metabolomics, wird die gezielte Analyse von zuvor definierten Schlüsselmetaboliten bezeichnet. Die Auswahl der zu untersuchenden Metabolite beruht typischerweise auf bestimmten biochemischen Fragestellungen oder Hypothesen und beschränkt sich teilweise auf spezifische Stoffwechselwege oder Substanzklassen.<sup>77</sup> Der Vorteil des Metabolic Profiling liegt vor allem darin begründet, dass die Möglichkeit der Erstellung von validierten, guantitativen Analysemethoden besteht, durch die absolute Gehalte der zu untersuchenden Metabolite bestimmt werden können. Die absolute Quantifizierung ist vor allem in Hinsicht auf eine methoden- und laboratorienübergreifende Vergleichbarkeit als besonders wertvoll einzuschätzen.

#### 3.2.4 Food Metabolomics

Neben den Metabolomics-basierten Ansätzen hat sich auch die Stabilisotopen-Analytik (Isotopolomics) seit einigen Jahren stark entwickelt und stellt einen komplementären Ansatz dar.<sup>79, 80</sup> Die Analyse beruht auf der regional unterschiedlichen Verteilung von stabilen Isotopen (überwiegend <sup>12</sup>C/<sup>13</sup>C, <sup>16</sup>O/<sup>18</sup>O, <sup>1</sup>H/<sup>2</sup>H, <sup>14</sup>N/<sup>15</sup>N, <sup>32</sup>S/<sup>34</sup>S), die auf biochemischen, klimatischen, anthropogenen und geologischen Faktoren beruhen. Aufgrund der Massenunterschiede der Isotope kommt es bei unvollständigen physikalischen und chemischen Prozessen

(z.B. globaler Wasserkreislauf) zu Fraktionierungen und damit geringfügigen Unterschieden in der Isotopenverteilung. Allerdings kann diese Zusammensetzung insbesondere durch agrarwirtschaftliche Praktiken wie künstliche Bewässerung, Düngung oder Pestizidbehandlung derart beeinflusst werden, dass eine sichere Vorhersage der Herkunft oft schwer ist.<sup>81, 82</sup>

Im Gegensatz dazu weisen Metabolomics-basierte Ansätze oft einen Vorteil gegenüber diesen traditionellen Methoden auf.<sup>49</sup> Die Anwendungsmöglichkeiten der Metabolomics-basierten Authentizitätsbestimmung von Lebensmitteln (Food Metabolomics) sind bereits in zahlreichen Veröffentlichungen aufgezeigt worden und demonstrieren die hervorragende Eignung des Metaboloms für zahlreiche unterschiedliche Fragenstellungen.<sup>44, 49, 50, 53</sup> Insbesondere die Überprüfung der Authentizität von Lebensmitteln in Bezug auf deren geographische Herkunft ist in zunehmendem Maße von Relevanz, was sich in der steigenden Anzahl an wissenschaftlichen Publikationen widerspiegelt. Tabelle 3-4 zeigt dazu eine Auswahl relevanter Veröffentlichungen seit 2014 zu dieser Thematik am Beispiel verschiedener Lebensmittel.

In Bezug auf Haselnüsse existieren ebenfalls bereits einige Veröffentlichungen, die sich allerdings überwiegend auf die Sortenunterscheidung oder Authentizitätsbestimmung von ausgewählten Produkten geschützter geographischer Herkunft (Nocciola del Piemonte, g.g.A.; Nocciola Romana, g.U.) beschränken. Hauptaugenmerk liegt dabei allerdings meist auf der Identifizierung der jeweiligen Haselnussorte, während die geographische Herkunft teils unbeachtet bleibt oder zusammen mit der Sorte betrachtet wird. Die verwendeten apparativen Methodiken sind vorwiegend NIR, <sup>1</sup>H-NMR, GC-FID sowie HPLC-UV sowie darüber hinaus ein Ansatz zum Vergleich der Spurenelemente mittels ICP-MS. Ein echter, konkreter Ansatz zur Bestimmung der geographischen Herkunft von Haselnüssen unabhängig der Haselnusssorte existiert bislang nicht. Während der Vergleich von Haselnüssen aus verschiedenen Anbauregionen oder verschiedener Sorten in Bezug auf bestimmte Inhaltsstoffe bereits vielfach erfolgte und per Definition ebenfalls als Metabolic Profiling zu verstehen ist, sind diese Studien häufig auf vergleichende agrarwirtschaftliche oder qualitative Fragestellungen beschränkt und haben nicht die tatsächliche analytische Bestimmung der geographischen Herkunft zum Ziel (siehe Tabelle 3-2 und Tabelle 3-3, Abschnitt 3.1). Im Zuge der Entwicklung von Methoden zur Bestimmung der geographischen Herkunft sollte unbedingt darauf geachtet werden, dass die Methoden und die entsprechenden Schlüsselmetabolite möglichst sortenunabhängig sind, da selbst für bestimmte Regionen charakteristische Sorten (z.B. Tonda di Giffoni, Kampanien, Italien) mittlerweile auch in an-

Lebensmittel	Analysenmethode	Zielmoleküle	Fragestellung	Quelle
Haselnuss	RP-HPLC-DAD	Phenole	Einfluss der Röstung auf	Locatelli et al. (2015)
	GC-FID	Fettsäuren	geographische Herkunfts-	
			bestimmung	
Haselnuss	NIR	non-targeted	Authentizität Nocciola	Moscetti et al. (2014)
			Romana (g.U.)	
Haselnuss	HPLC-UV	Phenole	Sortenunterscheidung	Ciarmiello et al. (2014)
	MALDI-TOF-MS	Proteine		
Haselnuss	<sup>1</sup> H NMR	non-targeted	Authentizität Nocciola del	Caligiani et al. (2014)
		(polare Metabolite)	Piemonte (g.g.A.)	
Haselnuss	<sup>1</sup> H NMR	non-targeted	Sortenunterscheidung	Sciubba et al. (2014)
Haselnuss	RP-HPLC-UV	Phenole	Authentizität Nocciola del	Locatelli et al. (2011)
	GC-FID	Fettsäuren	Piemonte (g.g.A.)	
Haselnuss	ICP-MS	Spurenelemente	Authentizität Nocciola del	Oddone et al. (2009)
			Piemonte (g.g.A.)	
Geleepalme	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft/	Hoffmann et al. (2017)
			Genotypen	
Reis	<sup>1</sup> H- <sup>13</sup> C NMR	non-targeted	Geographische Herkunft	Chae and Kim (2016)
Himbeere	UPLC-ESI-Orbitrap-MS	non-targeted	Geographische Herkunft	D'Urso et al. (2016)
	UPLC-ESI-QTrap-MS	Phenole		
Kaffee	NIR	non-targeted	Geographische Herkunft/	Marquetti et al. (2016)
			Genotypen	
Olivenöl	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft	Gil-Solsona et al. (2016)
Kaffee	GC-MS	non-targeted	Geographische Herkunft	Arana et al. (2016)
	<sup>1</sup> H NMR			
Apfel	<sup>1</sup> H NMR	non-targeted	Geographische Herkunft	Tomita et al. (2015)
Goji Beere	NIR	non-targeted	Geographische Herkunft	Tingting et al. (2015)
Tabak	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft	Li et al. (2015)
Tee	<sup>1</sup> H NMR	non-targeted	Geographische Herkunft	Lee et al. (2015)
Kakao	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft	Hori et al. (2015)
Pistazie	<sup>1</sup> H NMR	non-targeted	Geographische Herkunft	Sciubba et al. (2014)
		(polare Metabolite)		
Orange	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft	Diaz et al. (2014)
Goji Beere	UPLC-ESI-QTOF-MS	non-targeted	Geographische Herkunft	Bondia-Pons et al. (2014)
Butter	FT-IR	non-targeted	Geographische Herkunft	Bassbasi et al. (2014)

# Tabelle 3-4: Veröffentlichungen zu Metabolomics-Applikationen zur Bestimmung der Authentizität von Haselnüssen sowie zur Bestimmung der geographischen Herkunft von anderen Lebensmitteln seit 2014.

deren Ländern bzw. Regionen angebaut werden und es sonst zu Verzerrungen der statistischen Modelle durch einen länderübergreifenden Sortenanbau kommen kann. Darüber hinaus ist ersichtlich, dass bereits existierende Ansätze zur geographischen Herkunftsbestimmung bei anderen Lebensmitteln fast ausschließlich auf non-targeted Analysen beschränkt sind, während echte targeted Metabolomics-Applikationen derzeit noch fehlen. Der in dieser Forschungsarbeit verfolgte Ansatz der Entwicklung von LC-MS-basierten targeted Methoden zur Quantifizierung ausgewählter Schlüsselmetabolite in einfachen und robusten Methoden nach vorangegangener non-targeted Analyse zur Identifizierung geeigneter Schlüsselmetabolite ist daher bislang einzigartig.

### 3.3 LC-MS

### 3.3.1 Flüssigchromatographie

Die Flüssigchromatographie ist ein auf chemisch-physikalischen Grundlagen basierendes Trennverfahren für Stoffgemische. Durch unterschiedlich starke Wechselwirkungen mit einer immobilisierten, stationären Phase und einer diese durchströmende mobile Phase werden die zu trennenden Substanzen unterschiedlich retardiert. Voraussetzung für die Anwendung dieser Methode ist die Löslichkeit der Analyten in der mobilen Phase. Die Hochleistungs-Flüssigchromatographie (engl.: high performance liquid chromatography, HPLC) sowie die Ultrahochleistungs-Flüssigchromatographie (engl.: *ultra performance liquid chromatography*, UPLC) zeichnen sich im Gegensatz zur normalen Flüssigchromatographie durch eine Miniaturisierung aller Komponenten und die Anwendung von Arbeitsdrücken bis 600 bar (HPLC) bzw. mittlerweile sogar 1500 bar (UPLC) bei kommerziell erhältlichen Geräten aus.<sup>105</sup> Verschiedene Parameter der VAN-DEEMTER-Gleichung, die zur Verbesserung der Trennleistung aber zur Erhöhung des Rückdrucks führen, sind dadurch optimierbar. Sie beschreibt die Effizienz einer chromatographischen Trennung und gibt die Bodenhöhe H (engl.: height equivalent to a theoretical plate) als Funktion der Fließgeschwindigkeit unter Berücksichtigung der zu einer Peakverbreiterung beitragenden Effekte der EDDY-Diffusion, der Longitudinaldiffusion und des Massentransfers an.<sup>106</sup> Neben der Erhöhung der Auflösung und der Sensitivität können dadurch kürzere Analysenzeiten erreicht werden. Einer der wichtigsten Faktoren ist dabei die Partikelgröße der stationären Phase, die mit abnehmender Größe zu einer Vergrößerung der Oberfläche und dadurch Verbesserung der Trennleistung führt. Partikelgrößen im sub-2-µm-Bereich sind daher in hochaufgelösten Metabolomics-Applikationen mittlerweile weit verbreitet. Allerdings vergrößert sich bei abnehmender Partikelgröße auch der Arbeitsdruck.

Eine zusätzliche Effizienzsteigerung wird durch die Verwendung teilporöser Partikel erreicht, bei der eine poröse Schicht der stationären Phase auf einen festen, unporösen Kern aufgebracht ist. Einerseits wirkt sich diese Art der Partikelbeschaffenheit durch die Verkleinerung von Diffusionswegen positiv auf die Trennleistung aus. Andererseits verringert sich gleichzeitig aber auch der Arbeitsdruck bei gleichbleibendem Fluss. Daraus resultiert, dass teilporöse stationäre Phasen in der Lage sind, in kürzeren Retentionszeiten eine höhere chromatographische Auflösung bei geringerem Gegendruck zu erzielen.<sup>107</sup> Abbildung 3-6 zeigt eine schematische Darstellung einer UPLC-Anlage. Die mobile Phase wird meist durch



Abbildung 3-6: Schematische Darstellung einer UPLC.

eine Kurzhub-Doppelkolbenpumpe aus dem Vorratsgefäß gefördert und zunächst über poröse Teflonschläuche im Vakuum entgast. Enthaltene Gase stören beim Freiwerden den konstanten Fluss der mobilen Phase und sorgen für ein erhöhtes Grundrauschen. Es wird je nach Anzahl der Pumpeinheiten eines Systems zwischen einfachen, binären, tertiären und quaternären Pumpen unterschieden. Wird über die Dauer des chromatographischen Laufes eine konstante Laufmittelzusammensetzung gefördert, spricht man von einer isokratischen Elution. Werden mehrere Laufmittel in sich verändernden Verhältnissen gefördert, wird je nach Einstellung von diskontinuierlicher oder kontinuierlicher Gradientenelution gesprochen. Die Gradientenelution kann sich durch Variation der Polarität positiv auf die Trennungseigenschaften und -Analysendauer auswirken.

Die Probeninjektion erfolgt im Injektor durch ein Mehr-Wege-Ventil, wodurch Druckschwankungen geringgehalten werden. Zum Schutz der eigentlichen Trennsäule vor Schmutzpartikeln oder unlöslichen Bestandteilen kann eine Vorsäule vorgeschaltet werden. Die Trennsäule besteht meist aus einem Edelstahlrohr, in dem die stationäre Phase immobilisiert eingebracht ist. Die üblichen Säulenmaße belaufen sich auf eine Länge von 5 bis 30 cm und einen Innendurchmesser von bis zu 5 mm. Die Partikelgröße reicht von 10 bis mittlerweile unter 2 µm. Nach dem Verlassen der Säule werden die zu analysierenden Substanzen in einem geeigneten Detektor erfasst. Zur Optimierung eines chromatographi-
schen Systems können Parameter wie Art und Zusammensetzung der mobilen Phase (Lösungsmittel, Gradient, Zusatz von Puffern, pH-Wert), Fließgeschwindigkeit, Säulenofentemperatur oder Injektionsvolumen herangezogen werden. Den letztendlich entscheidenden Einfluss hat aber die Art und Beschaffenheit der stationären Phase, welche maßgeblich für die Wechselwirkungen mit den Analyten verantwortlich ist.

### 3.3.1.1 Umkehrphase

Die stationäre Phase besteht bei der Reversed Phase (RP, dt.: Umkehrphase) aus chemisch modifizierten Silicagelen mit einer lipophilen Oberfläche. Zur Modifizierung wird meistens eine Silvlierung der endständigen Hydroxylgruppen des Silicagels mit unverzweigten Alkylketten mit Kettenlängen von C-18 und C-8 genutzt. Diese Arten von stationären Phasen sind die meistgenutzte Form in diversen Anwendungsbereichen der Metabolomics-Disziplin.<sup>108</sup> Sie zeichnen sich durch gute Trenneigenschaften von unpolaren und mittelpolaren Substanzen, eine hohe Reproduzierbarkeit, sowie geringe Anforderungen an die mobile Phase aus. Als meistgenutztes Laufmittel finden sich häufig Kombinationen aus wässrigen Systemen mit organischen Anteilen wie Methanol oder Acetonitril. Durch die Zugabe von Isopropanol oder Tetrahydrofuran kann die Elution von sehr unpolaren Substanzen beschleunigt werden. Der Trennmechanismus basiert dabei vor allem auf der Verteilungschromatographie, bei der die Analyten durch unterschiedlich starke, auf Löslichkeitsunterschieden basierenden Wechselwirkungen mit flüssiger mobiler und stationärer Phase getrennt werden. Jedoch wird zu einem gewissen Teil auch der Mechanismus der Adsorptionschromatographie beobachtet, welcher unterschiedliche Affinitäten der Analyten zu Adsorptions- und Desorptionsvorgängen mit fester stationärer und flüssiger mobiler Phase zugrunde gelegt werden. Grundsätzlich sind dabei hydrophobe (van-der-Vaals-) Wechselwirkungen für die Retention verantwortlich.

### 3.3.1.2 Silica Hydrid Phase

Aufgrund des unpolaren Charakters der stationären Phase und der Nutzung von polaren bis mittelpolaren Lösungsmitteln oder Lösungsmittelgemischen ist die Anwendbarkeit von Umkehrphasensystem hinsichtlich polarer Analyten deutlich begrenzt. Diese erfahren keine oder nur eine geringe Retardation.<sup>109</sup> Herkömmliche Normalphasensysteme besitzen eine polare stationäre Phase und werden mit organischen wie Lösungsmitteln wie n-Hexan oder Chloroform betrieben, die nicht mit Wasser mischbar sind. Die fehlende Kompatibilität dieser Lösungsmittel mit MS-Kopplungen - und speziell mit ESI-MS-Kopplungen - sowie die

Unlöslichkeit von biologischen Proben in der mobilen Phase haben die Entwicklung von alternativen Methoden erforderlich gemacht. Aus diesem Grund werden für polare Substanzen im Metabolomics-Bereich häufig andere Systeme wie die Hydrophilic Interaction Liquid Chromatography (HILIC) genutzt.<sup>110</sup> Diese basiert wie die klassische Normalphasenchromatographie auf Silica-, Amino- oder Cyano-Säulen mit verschiedenen Modifikationen und kombiniert diese mit den positiven Eigenschaften der typischen RP-Laufmittel, wodurch die Trennung von polaren Substanzen mit MS-Kopplungen ermöglicht werden.<sup>111</sup> Sie ist zudem in der Lage Ionen zu trennen, ohne die bei der Ionenaustauschchromatographie üblichen, für das MS schädlichen, hohen Salzkonzentrationen zu benötigen. Der Trennmechanismus basiert auf der Ausbildung einer teilweise immobilisierten, wässrigen Oberflächenschicht auf der stationären Phase, mit der die Analyten nach dem Verteilungsmechanismus mit der mobilen Phase wechselwirken.<sup>112, 113</sup> Während bei der Ionenchromatographie vor allem die hohen Pufferkonzentrationen sowie der notwendige ionische Charakter der Metabolite als Nachteile zu nennen sind, sind bei der HILIC insbesondere die schlechte Reproduzierbarkeit sowie die langen Equilibrierungszeiten wegen der sich ausbildenden immobilisierten Wasserschicht störend.114,115

Sandoval und Pesek stellten daher bereits 1989 eine neuartige Silica-Hydrid Phase vor, die viele der Vorteile bisheriger Phasenmechanismen für polare Substanzen verbindet und gleichzeitig keine der oben genannten Nachteile aufweist.<sup>116</sup> Herkömmliche Normalphasensysteme sind meist Silica-basiert und besitzen an der Oberfläche freie Silanol-Gruppen, während diese bei Silica Hydrid Phasen zu Silanen reduziert sind.<sup>113</sup> Im Gegensatz zur klassischen Normalphasenchromatographie können dadurch wie bei der HILIC auch bei der Silica Hydrid Phase wässrig/organische Lösungsmittelgemische verwendet werden, wie sie auch in der RP-Chromatographie Anwendung finden, weshalb der Mechanismus auch als wässrige Normalphasenchromatographie (engl.: aqueous normal-phase, ANP) bezeichnet wird.<sup>117</sup> Anders als bei der HILIC ist die Adsorption von Wasser aber auf ein Minimum reduziert, wodurch sich die Reproduzierbarkeit und Equilibrierungsgeschwindigkeit erhöhen.<sup>118</sup> Für die Analyse von polaren Metaboliten beginnt die Gradientenelution bei diesen Methoden mit einem hohen Anteil an organischem Lösungsmittel. Die höhere Flüchtigkeit der Lösungsmittel gegenüber Wasser birgt den zusätzlichen Vorteil der erhöhten Sensitivität bei ESI-MS-Kopplungen.<sup>111</sup> Für die Trennung von geladenen Substanzen können zusätzlich Puffer zur pH-Stabilisierung eingesetzt werden. Gängige Puffer bilden bspw. Ameisensäure/Ammoniumformiat und Essigsäure/Ammoniumacetat in einem pH-Wert-Bereich von 3-8, welche im Zuge der Ionisation zu flüchtigen Stoffen reagieren und dadurch wenig Ionensuppressionseffekte oder Salzverkrustungen verursachen.

### 3.3.2 Massenspektrometrie

### 3.3.2.1 Electrospray Ionisation

Die Electrospray Ionisation (ESI) bildet heutzutage den bedeutendsten Vertreter der Atmosphärendruck Ionisationstechniken (engl.: atmospheric pressure ionization, API) für LC-MS Kopplungen bei biologischen Proben.<sup>119</sup> Dabei wird die Analytlösung unter Atmosphärendruck durch eine dünne Kapillare (Sprayer Needle oder Sprayer Capillary) in einem elektrischen Feld versprüht. Die Elektroden bestehen aus der Kapillare selbst und einer zylindrischen Gegenelektrode. Durch Anlegen eines elektrischen Gleichspannungsfeldes von typischerweise 3-5 kV werden die Analyten einerseits in der Kapillare ionisiert und andererseits einer Beschleunigungskraft hin zur Gegenelektrode unterworfen. Der gewählte Ionenmodus entscheidet darüber, welche Elektrode positiv oder negativ geladen ist, sodass sowohl Anionen als auch Kationen erzeugt werden können. Aus einem negativen Ionenmodus entstehen demnach Anionen und umgekehrt. Die Ladungen werden meist durch Protonierung  $[M + nH]^{n+}$  bzw. Deprotonierung  $[M - nH]^{n-}$  einer funktionellen Gruppe gebildet. Diese Arten der Ionen werden als Pseudomolekülionen bezeichnet. Im positiven Ionenmodus können zudem teilweise verschiedene Addukte wie beispielsweise Natrium- [M + nNa]<sup>n+</sup>, Kalium-Addukte  $[M + nK]^{n+}$  gebildet werden, die meist durch deren ubiquitäres Vorhandensein zu erklären sind.<sup>120</sup> Ammonium-Addukte  $[M + nNH_4]^{n+}$  bilden sich im Speziellen bei Anwesenheit von Ammoniumformiat oder -acetat Puffern in den entsprechenden Laufmitteln. Bei kleinen Molekülen entstehen in der Regel einfach geladene Ionen, wohingegen die Tendenz bei größeren Molekülen mit vielen funktionellen Gruppen zu mehrfach geladenen Ionen steigt. Da bei der ESI im Vergleich zu anderen Ionisierungsverfahren sehr wenig In-Source-Fragmentierungen der Moleküle auftreten, wird die Methode als weiche Ionisation bezeichnet.<sup>121</sup> Die Funktionsweise der *Electrospray Ionization* und die Entstehung der Ionen sind in Abbildung 3-7 schematisch dargestellt. Aufgrund des starken elektrostatischen Feldes an der Spitze der Sprühkapillare driften die negativen Gegenionen im positiven Ionenmodus (und umgekehrt im negativen Ionenmodus) von der Oberfläche der Lösung in Richtung der Wand der Kapillare und werden dort entladen, während die positiven Ionen in Richtung Lösungsmittelfront gelangen.<sup>119</sup> Daraus folgend bildet sich durch elektrostatische Wechselwirkungen an der Spitze der Sprühkapillare ein sogenannter Taylor-Kegel (engl.: taylor cone),



Abbildung 3-7: Schematische Darstellung der *Electrospray Ionisation*.

an dessen Front sich aufgrund der steigenden Instabilität durch die Abstoßungskräfte gleichartig gerichteter Ladungen geladene Tröpfchen abspalten. Die Bildung des Aerosols wird durch das Einleiten eines *Nebulizer Gases* (meist Stickstoff) unterstützt. Infolge stetiger Verdunstung des Lösungsmittels durch eingeleitetes *Drying Gas* (ebenfalls meist Stickstoff) entstehen aufgrund des Ladungsüberschusses in den Tröpfchen immer größere Abstoßungskräfte, bis es bei Überschreitung des Rayleigh-Limits zu einer Coulomb-Explosion kommt, wobei sich weitere kleinere Tröpfchen bilden.<sup>122</sup> Der exakte Vorgang des Übergangs der Ionen aus der flüssigen in die Gasphase ist bis heute nicht eindeutig geklärt.<sup>123</sup> Es existieren dazu zwei Modelle, deren Gültigkeit diskutiert wird. Das *Ion Evaporation Model* (IEM) geht davon aus, dass es nach der Coulomb-Explosion zu einer Desorption der Ionen in die Gasphase kommt, wohingegen das *Charge Residue Model* (CRM) eine stetige Verdunstung des Lösungsmittels bis zum Zurückbleiben der freien Ionen zugrunde legt.<sup>124, 125</sup>

Eine der größten Herausforderung bei LC-MS-basierten Applikationen mit ESI als Ionisationsmethode stellt die Ionensuppression bei Multikomponentenproben dar. Kompetitive Effekte der Probenmatrix, des Lösungsmittel oder von koeluierenden Substanzen haben einen Einfluss auf die Ionisation der Analyten in der Ionenquelle, sodass die Signalintensität gemindert oder ausgelöscht werden kann.<sup>126</sup> Quantitative Messungen werden dadurch teilweise erheblich beeinträchtigt. Die zur Ionensuppression führenden Vorgänge können durch eine Vielzahl von Effekten ausgelöst werden. Als maßgeblich ausschlaggebend wird die Beeinträchtigung der Bildung und der Evaporierung der Lösungsmitteltröpfchen angesehen. Durch die Anwesenheit von nicht oder kaum flüchtigen Substanzen kann die Effizienz des Überganges der Analyten aus den Lösungsmitteltröpfchen in die Gasphase gehemmt werden. Zusätzlich tragen Ladungsübertragungsreaktionen der Analytionen auf andere Moleküle und Ionenpaarbildungen mit in der Lösung enthaltenen, entgegengesetzt geladenen Salzen zu einer Verringerung der absoluten Ionenkonzentration im Massenanalysator bei.<sup>127</sup> Die Konzentration an Analyten und koeluierenden Substanzen, deren Polaritäten, Massen, Ladungen, Flüchtigkeiten, Oberflächenaktivitäten und Löslichkeiten in der mobilen Phase sowie die Pufferkonzentration, der pH-Wert, die Zusammensetzung, Polarität und Flüchtigkeit der mobilen Phase und deren Fließgeschwindigkeit haben einen erheblichen Einfluss auf den Ionisierungsvorgang.<sup>126, 128</sup> Es wurde beobachtet, dass Moleküle großer Masse bevorzugt kleinere in ihrer Ionisation unterdrücken und dass unpolare und oberflächenaktive Moleküle weniger anfällig für Ionensuppression sind als polare. Dies wird dadurch erklärt, dass sich unpolare und oberflächenaktive Moleküle bevorzugt an der Oberfläche der meist polareren Lösungsmitteltropfen aufhalten und von dort leichter in die Gasphase desorbieren können. Polare Substanzen hingegen verbleiben im Inneren der Tröpfchen und können schwieriger isolierte Ionen in der Gasphase ausbilden.<sup>129, 130</sup>

Als effiziente Maßnahme zur Verringerung der Ionensuppressionseffekte bei komplexen, matrixreichen Proben hat sich eine vorgeschaltete chromatographische Trennung der Analytlösung etabliert. Durch diese Kopplung wird eine zeitlich aufgelöste Auftrennung der Bestandteile erreicht, wodurch die Komplexität der gleichzeitig in die Ionenquelle eingebrachten Moleküle deutlich verringert wird. Als weitere Maßnahme kann sich eine Verringerung des Probenvolums als vorteilhaft erweisen. Als vielversprechend hat sich in dieser Hinsicht die Entwicklung von Nano-Flow-Infusion-Techniken erwiesen, mit der die Ionensuppression weitestgehend reduziert werden kann.<sup>127</sup>

### 3.3.2.2 Triple Quadrupol Massenspektrometer

Triple Quadrupol Massenanalysatoren (QqQ) zeichnen sich im Vergleich zu anderen Massenspektrometern vor allem durch ihre hohe Sensitivität, Selektivität und Reproduzierbarkeit aus. Wegen ihrer hervorragenden Quantifizierungseigenschaften werden sie bevorzugt

### Einleitung



Abbildung 3-8: Schematische Darstellung eines Triple Quadrupol Massenspektrometers.

bei targeted Analysen eingesetzt.<sup>43, 131</sup> Es handelt sich dabei um sogenannte "Tandem-in-Space"-Applikationen, bei denen mehrere Massenanalysatoren physisch hintereinander geschaltet sind.<sup>76</sup> Wie der Name schon ausdrückt, sind bei Triple Quadrupol Massenspektrometern drei dieser Quadrupole in Reihe geschaltet, wobei der zweite lediglich als Kollisionszelle und nicht als massenselektiver Analysator dient. Generell wird die Aneinanderreihung von mehreren Massenanalysatoren mit zwischengeschalteter Fragmentierung als MS/MS oder MS<sup>n</sup> bezeichnet.

Ein Quadrupol besteht aus vier parallelen Metallstäben in einem Vakuum, welche äquidistant zu einer gemeinsamen Mittelachse angeordnet sind. Die beschleunigten Ionen werden durch ein Gleichstromfeld entlang der Mittelachse beschleunigt und gemäß ihres Masse-zu-Ladungs-Verhältnisses (*m*/*z*-Verhältnis) durch Anlegen eines überlagerten, hochfrequenten Wechselspannungsfeldes mehr oder weniger stark abgelenkt. Durch Variation der elektrischen Parameter können einzelne Ionen eines bestimmten *m*/*z*-Verhältnisses den Quadrupol passieren, während alle anderen abgelenkt, an den Metallstäben entladen und durch die Vakuumpumpe abgesaugt werden. Ebenso ist es möglich Ionen von bestimmten Massenbereichen oder alle einfallenden Ionen hindurch zulassen.<sup>121</sup> Der prinzipielle Aufbau eines Triple Quadrupol Massenspektrometers ist in Abbildung 3-8 dargestellt. Im Anschluss an die Ionisation werden die erzeugten Ionen in das Massenspektrometer geleitet. Orifice und Skimmer dienen hierbei als Fokussierungsvorrichtungen zur Erzeugung eines diskreten Ionenstrahls. Zum Schutz vor Verunreinigungen befindet sich davor das Curtain Plate. Entgegengesetzt zur Richtung der einfallenden Ionen strömt das *Curtain Gas* (meist Stickstoff) durch die Öffnung des *Curtain Plate*, um überwiegend ungeladene Verunreinigun-

### Einleitung

gen am Eindringen in die hochsensiblen Bereiche zu hindern. Je nach Intensität des Gasstromes wird jedoch auch ein Signalverlust der detektierten Ionen beobachtet. Nach der Fokussierung erfolgt im ersten Quadrupol (Q1) die erste massenselektive Trennung der einfallenden Ionen. Die dort selektieren Ionen werden auch Precursor-Ionen genannt. Die Precursor-Ionen werden im Q2 durch Einleiten eines Gasstroms durch Kollisionen mit den Gaspartikeln reproduzierbar fragmentiert. Dieser Vorgang wird als Collision Induced Dissociation (CID) bezeichnet.<sup>132</sup> Im Q2 ist der Druck durch den Gaseinstrom mit etwa 10<sup>-1</sup> Pa höher als im Rest des Massenanalysators. Die Intensität der Fragmentierung kann durch Variation der anliegenden Spannung in der Kollisionszelle, sowie geringfügig durch die Stärke des Gasstroms beeinflusst werden. Anschließend erfolgt eine zweite massenselektive Trennung der entstandenen Fragment-Ionen (Produkt-Ionen) im Q3. In der Regel erreichen auch noch Reste von undissoziierten Precursor-Ionen den Q3. Die Detektion erfolgt meist mithilfe des sogenannten Faraday-Bechers, bei dem die Ionen auf eine Oberfläche mit konstantem Potential auftreffen, wodurch Elektronen über einen Widerstand zu- oder abfließen müssen und die resultierenden Spannungsschwankungen gemessen werden können. Alternativ kommt häufig ein Sekundärelektronenvervielfacher zur Anwendung, bei dem durch das Aufprallen eines Ions auf eine Dynode mehrere Elektronen herausgeschlagen werden. Die Anzahl der emittierten Sekundärelektronen ist proportional zur Intensität des auftreffenden Ionenstrahls. Die emittierten Elektronen werden zu einer weiteren Dynode beschleunigt, um dort wiederum weitere Sekundärelektronen zu erzeugen. Je nach Anzahl der Wiederholungen dieses Vorgangs kann eine Signalverstärkung von bis zu 10<sup>8</sup> Elektronen pro auftreffendem Ion erreicht werden.

Es können sowohl Precursor-Ion- als auch Produkt-Ion-Spektren aufgenommen werden, je nachdem welche der beiden massenselektiven Quadrupole für alle einfallenden Ionen durchlässig gestellt wird. Die Spannung des anderen Quadrupols wird so variiert, dass nacheinander alle m/z-Verhältnisse detektiert werden. Zur Quantifizierung von Substanzen wird meistens ein als *Multiple Reaction Monitoring* (MRM) bezeichnetes Verfahren angewendet. Es beruht auf der Detektion von mehreren ausgewählten Produkt-Ionen eines bestimmten Precursor-Ions, wodurch eine extrem hohe Selektivität erreicht wird.<sup>133</sup> Auch die Sensitivität wird durch die Verbesserung des Signal-zu-Rausch Verhältnisses in erheblichem Maße gesteigert.<sup>134</sup> Selbst isobare Verbindungen können durch unterschiedliche Fragmentierungsmuster eindeutig unterschieden werden. Allerdings stößt auch diese Methode bei

Stereoisomeren und Konsitutionsisomeren mit identischen Fragmentierungsmustern an ihre Grenzen.

Die Vorteile von Triple Quadrupol Massenspektrometern gegenüber anderen MS-Systemen sind vor allem die einfache Handhabung in Kombination mit einer hohen Selektivität und Sensitivität sowie der Fähigkeit auch sehr kleine Ionen zu analysieren. Es handelt sich um eine robuste Applikation mit der Fähigkeit zur Generierung von reproduzierbaren Ergebnissen, die eine gute Vergleichbarkeit mit spektralen Datenbänken ermöglicht. Als Nachteile sind vor allem der nach oben stark limitierte Massenbereich von maximal 4000 Da und die vergleichsweise schlechte Auflösung zu berücksichtigen.<sup>135, 136</sup>

### 3.3.2.3 Time-of-Flight Massenspektrometer

Im Unterschied zum Quadrupol Massenanalysator wird bei einem Time-of-Flight (TOF) Massenspektrometer die tatsächliche Flugzeit von Ionen in einem definierten Hochvakuumbereich gemessen um Informationen über das m/z-Verhältnis zu erhalten. Dabei bildet die Tatsache, dass Ionen unterschiedlicher m/z-Verhältnisse bei der Zufuhr gleicher kinetischer Energie unterschiedlich stark beschleunigt werden und somit verschiedene Zeiten für das Zurücklegen derselben Strecke benötigen, die Grundlage für die Trennung der Ionen. Die Ionen werden zunächst in einem sogenannten Ion-Cooler verlangsamt und zu Ionenpaketen gebündelt. Diese Ionenpakete treten anschließend mit bis zu 20 kHz in den orthogonalen Beschleuniger ein, wo sie durch ein darin befindliches elektrisches Feld einen starken Impuls erhalten. Durch die orthogonale Beschleunigung sollen longitudinale Diffusions-Effekte vermieden werden.<sup>137</sup> Die Ionen erhalten entsprechend dem lokalen Potential eine kinetische Energie ( $E = \frac{1}{2} \text{ m} \cdot v^2$ ) und werden in einen feldfreien Raum hinein beschleunigt. Ionen kleinerer m/z erhalten eine größere Geschwindigkeit als größere m/z und erreichen daher bei gleicher Strecke früher den Detektor. Die Trennung der Ionen unterschiedlicher m/z-Verhältnisse und damit das Auflösungsvermögen sind umso genauer, je länger der zurückgelegte Weg ist. Um die Größe der TOF-Geräte zu limitieren wurden daher Reflektoren eingeführt, die die einfallenden Ionen durch ein umgekehrt gerichtetes elektrisches Feld, stoppen und schließlich entgegengesetzt in Richtung Detektor beschleunigen und damit die Flugstrecke verdoppeln. Zum Zeitpunkt des Starts befinden sich die Ionen in unterschiedlicher Nähe zum Ursprung des Potentialgefälles im orthogonalen Beschleuniger, sodass Ionen gleicher m/z-Verhältnisse verschiedene kinetische Energien erhalten und sich dementsprechend unterschiedlich schnell fortbewegen. Da im Reflektor allerdings ein





Abbildung 3-9: Schematische Darstellung einer QTOF Massenspektrometers.

gegenteiliger Effekt auftritt und schnelle Ionen tiefer eindringen als langsame, werden diese Dispersions-Effekte wieder kompensiert, wodurch die Auflösung zusätzlich gesteigert wird.<sup>120, 121</sup> Zur Detektion werden meist dieselben Verfahren wie beim QqQ eingesetzt (siehe Abschnitt 3.3.2.2).

Eine weit verbreitete Erweiterung stellt die Kopplung mit Quadrupolen zu sogenannten Hybrid Massenspektrometern (Tandem-Applikation aus mind. zwei verschiedenen Massenanalysatoren), genannt QTOF, dar. Der Aufbau ist vergleichbar zu einem Triple Quadrupol Gerät, wobei der dritte Quadrupol durch ein TOF ersetzt ist. Auch hier können die Ionen zunächst massenselektiv im Q1 getrennt und im Q2 fragmentiert werden. Die Trennung und Detektion erfolgt anschließend gemäß dem oben beschriebenen Prinzip des TOF-MS. Durch diese Kopplung wird die Akquisition von hochaufgelösten MS/MS Spektren ermöglicht, die einen großen Beitrag zur Identifizierung von unbekannten Substanzen liefern. Darüber hinaus können quantitative Messung im MRM-Modus durchgeführt werden.<sup>140</sup> Der Aufbau eines QTOF ist in Abbildung 3-9 schematisch dargestellt. Als Vorteile von TOF-MS sind insbesondere die extrem hohe Massenauflösung und -genauigkeit sowie der theoretisch nach oben unbegrenzte Massenbereich, der große dynamische lineare Bereich und die hohe Scanrate zu nennen. Nachteilig jedoch sind meist die vergleichsweise schlechte Robustheit und Reproduzierbarkeit, schlechtere Sensitivität und die höheren Kosten in Bezug auf Triple Quadrupol Geräte.<sup>136, 138, 139</sup>

### 3.3.2.4 Kopplung von Flüssigchromatographie und Massenspektrometrie

Die Kombination aus Massenspektrometrie und Flüssigchromatographie hat sich in den letzten Jahren als wegweisende Technik zur Analyse von komplexen und matrixreichen Analytgemischen erwiesen. Vor allem bei der Untersuchung von biologischen Proben, wie sie im Metabolomics Bereich häufig anzutreffen sind, ist diese Analysenmethode inzwischen weit verbreitet.<sup>72, 141</sup> Insbesondere bei der Reduktion der Ionensuppressionseffekte, ausgelöst durch koeluierende Verbindungen, spielt die chromatographische Trennung von komplexen Probengemischen eine wichtige Rolle. Sowohl eine Erniedrigung der Nachweisgrenze und des Hintergrundrauschens, als auch eine Vergrößerung des linearen Arbeitsbereiches kann durch eine gute analytische Trennung erzielt werden.<sup>71</sup> Die besten Ergebnisse können dabei bei sehr niedrigen Volumina bzw. Fließgeschwindigkeiten erreicht werden, sodass Nano-Flow-Applikationen ein Maximum an Sensitivität generieren.<sup>142</sup> Zusätzlich können durch das Auffangen diskreter Fraktionen weitere Analysen (z.B. NMR) angeschlossen werden.<sup>143</sup> Daneben bietet die Einführung der Dimension Zeit neben der Masse der Precursor- und Produkt-Ionen ein weiteres Identifikationsmerkmal. Isobare Verbindungen mit gleichem Fragmentierungsmuster können somit zeitlich getrennt voneinander detektiert werden.<sup>144</sup> Sogar Enantiomere sind mithilfe der chiralen Chromatographie prinzipiell trennbar.<sup>145</sup> Das breite Spektrum an kommerziell erhältlichen Säulen und die diversen Kombinationsmöglichkeiten mit verschiedenen mobilen Phasen und Gradienten erlauben es heutzutage, eine maßgeschneiderte Trennungsmethode für die gewünschte Verbindung zu entwickeln, wodurch die meisten Substanzen für LC-MS-basierte Analysen zugänglich sind.<sup>78</sup>

# **3.4** Multivariate Datenanalyse

Die Rohdaten von Metabolomics-Experimenten enthalten letztendlich sämtliche, mit der jeweiligen Applikation generierbare Informationen aus einer metabolischen Momentaufnahme und spiegeln somit das äußerst komplexe Resultat sämtlicher Einflussfaktoren auf die zu untersuchende Probe wider. Im Idealfall werden dabei hunderte bis tausende von Metaboliten gleichzeitig erfasst. Um aus dieser unübersichtlichen Datenmasse die für die jeweilige Fragestellung relevanten Informationen in kürzester Zeit extrahieren zu können, ist die Verwendung von multivariater Statistik unumgänglich, da einfache univariate Methoden schnell an die Grenzen der Zweckmäßigkeit stoßen. Ziel der multivariaten Datenanalyse ist es daher, die Datenmenge zu reduzieren, relevante Informationen herauszuarbeiten und darzustellen sowie nicht-Informationen wie bspw. das Grundrauschen zu entfernen, wodurch nicht direkt messbare Größen erfasst werden können.<sup>146, 147</sup> Bspw. wird das Metabolom von Haselnüssen neben der geographischen Herkunft zusätzlich durch andere Einflussfaktoren wie z.B. die Sorte bedingt (vgl. Abschnitt 3.2.2), wodurch die Konzentration einzelner Metabolite durch mehrere Informationsursprünge summiert bzw. überlagert sein kann. Mit Hilfe der multivariaten Datenanalyse kann gezielt nach denjenigen Metaboliten gesucht werden, die repräsentativ für den geographischen Ursprung sind. Die Auswertung wird dabei in zwei Schritte unterteilt: die Datenvorbearbeitung sowie die eigentliche statistische Analyse, welche mit verschiedensten Methoden durchgeführt werden kann.<sup>148</sup> In der vorliegenden Arbeit wurden folgende im Metabolomics-Bereich häufig verwendete multivariate Analysenmetho-den verwendet (exemplarisch sind aktuelle Applikationsbeispiele angeführt):

- Hauptkomponentenanalyse (engl.: *Principal Component Analysis*; PCA)<sup>100, 149, 150</sup>
- Partial Least Squares Discriminant Analysis (PLS-DA) bzw. Partial Least Squares Regression (PLS-R)<sup>93, 96, 151</sup>
- Multiple Lineare Regression (MLR)<sup>261-263</sup>
- Soft Independent Modelling of Class Analogy (SIMCA)<sup>101, 152, 153</sup>
- Lineare Diskriminanzanalyse (LDA)<sup>154-156</sup>
- Support Vector Machines (SVM)<sup>84, 97, 157</sup>

Prinzipiell werden multivariate Auswerteverfahren in zwei unterschiedliche Kategorien unterteilt, die *supervised* und *unsupervised* Methoden. Während bei *unsupervised* Methoden (z.B. PCA) keine Informationen im Vorfeld vorgegeben werden und die statistischen Methoden ungerichtet nach Zusammenhängen suchen, werden bei den *supervised* Methoden (z.B. PLS, MLR, SIMCA, LDA und SVM) Gruppenzugehörigkeiten (in diesem Fall Herkunftsländer) definiert, woraufhin die Modelle gerichtet nach Unterschieden oder Gemeinsamkeiten suchen können. Dies hat den Vorteil, dass gezielt nach Informationen gesucht werden kann, wodurch die identifizierten Unterschiede meist größer sind. Gleichzeitig birgt diese Vorgehensweise jedoch das Risiko, dass vermeintliche Zusammenhänge identifiziert werden, die in Wirklichkeit nicht oder nicht in der vermeintlichen Intensität existieren.<sup>158</sup>

### 3.4.1 Datenvorbearbeitung

Vor der statistischen Analyse müssen die Rohdaten von Metabolomics Messungen zur Generierung zweckmäßiger Datenmatrices eine Reihe von Prozessierungsschritten unterlaufen, die je nach analytischer Plattform unterschiedliche Anforderungen erfüllen. Das sogenannte Data Pre-Processing ist maßgeblich ausschlaggebend für die Ergebnisse der statistischen Auswertung und muss daher stets individuell für die vorliegende Fragestellung betrachtet werden.<sup>159</sup> Beim LC-MS zählen hierzu insbesondere Peakdetektion, *Alignment*, Normalisierung, *Scaling* und *Centering*.<sup>148</sup>

Die generierten Chromatogramme enthalten neben der Signalintensität auch Informationen über die Retentionszeit und das m/z-Verhältnis einer Vielzahl von potentiellen Signalen. Diese müssen zunächst zu sogenannten "*Features*" zusammengefasst werden, während gleichzeitig falsch positiven Zuordnungen vermieden werden sollten. Der erste Schritt ist daher die Peakdetektion (oder auch *Feature Detection*), bei der die Software ausgehend von definierten Parametern chromatographische Signale als Peaks identifiziert, jedem *Feature* ein definiertes m/z-Verhältnis und eine Retentionszeit zuordnet, dessen Intensität bestimmt und Störsignale dabei möglichst ignoriert. Ionen mit gleicher Retentionszeit aber unterschiedlicher Masse (bzw. umgekehrt) müssen dabei getrennt voneinander erfasst werden. Zusätzlich dazu werden einem *Feature* mehrere Massenspuren zugeordnet, die entweder durch deren natürliche Isotope, Addukte oder Isotope der Addukte verursacht werden, sodass diese Signale nicht als jeweils eigenständige "Metabolite" betrachtet werden und die statistische Auswertung somit nicht verfälschen.<sup>61, 148</sup>

Beim Vergleich mehrerer Messungen ist ausgehend von minimalen Retentionszeitverschiebungen und Massenungenauigkeiten ein *Alignment* der Datensätze erforderlich, welches diese Unterschiede erfasst und ausgleicht, wodurch Signale gleicher Metabolite weiterhin korrekt zugeordnet werden können.<sup>72</sup> Solche minimalen Messabweichungen sind unvermeidbar und können unter anderem durch Temperaturschwankungen, pH-Wert Unterschiede, das Säulenmaterial oder Probenverschleppung verursacht werden.<sup>159</sup>

Ziel der anschließenden Normalisierung ist die Beseitigung von systematischen Messabweichungen, die beispielsweise durch unterschiedliche Wassergehalte, Probenkonzentrationen oder instrumentelle Drifts verursacht werden können und keinerlei chemische Informationen enthalten.<sup>158, 160</sup> Typischerweise wird dies durch Multiplikation jeder Probe mit einer probenspezifischen Konstante (z.B. Summe oder Mittelwert aller Signalintensitäten oder Intensität eines speziellen oder des größten Signals) erreicht, wobei die Verwendung der Summe aller Probensignale meist am zweckmäßigsten ist.<sup>158</sup>

Ein zentraler Punkt der Datenvorbehandlung mit einem enormen Einfluss auf den Output der statistischen Analyse stellt das *Scaling* dar, bei dem jede Variable (*Feature*) durch einen variablenspezifischen Faktor, den *Scaling* Faktor, geteilt wird.<sup>161</sup> Aus der generellen Beschaffenheit von Metabolomics Daten biologischer Proben resultiert die Problematik, dass die Signale abundanter Metabolite sehr große Werte annehmen können, während weniger konzentrierte Metabolite um ein vielfaches niedrigere Werte aufweisen. Daraus ergibt sich, dass die abundanten Metabolite bei der statistischen Auswertung deutlich stärker ins Gewicht fallen, weil selbst geringe Konzentrationsunterschiede größere absolute Signalunterschiede verursachen, als große Konzentrationsunterschiede bei wenig abundanten Metaboliten. Daher ist es notwendig, die Skalierung für alle *Features* gleich zu gestalten und dadurch eine konzentrationsunabhängige Gewichtung zu erhalten. Es gibt diverse Verfahren wie z.B. *Autoscaling, Range Scaling, Interquartile Range Scaling, Level Scaling, Pareto Scaling* oder *Vast Scaling*. Die Wahl der Methode muss in jedem Fall mit Bedacht erfolgen, da sie das Endergebnis in erheblichen Maß beeinflussen kann und gleichzeitig, abhängig von der Datenbeschaffenheit, nicht jede Methode für jeden Datensatz geeignet ist.<sup>161, 162</sup>

Schließlich können die Daten mit Hilfe des *Centering* zusätzlich auf eine einheitliche Null-Linie gebracht werden, um den Fokus auf den fluktuierenden Teil der Daten und nicht die absoluten Werte zu setzen.<sup>161, 163</sup>

# 3.4.2 Erstellung und Bewertung von Klassifizierungs- und Vorhersagemodellen

Zur Erstellung und Bewertung von multivariaten Modellen wird der Datensatz typischerweise in verschiedene Gruppen aufgeteilt. Die zur Modellerstellung verwendeten Daten werden als Training-Daten bezeichnet. Neue (unbekannte) Proben, die mithilfe der zuvor erstellten Modelle vorgesagt werden sollen, sind die Vorhersage-Daten (engl.: *prediction data*). Während die Modellgenauigkeit (engl.: *training accuracy*; Anteil richtig zugeordneter Proben des Training-Datensatzes) die Güte des Modells und damit die Fähigkeit zur Diskriminierung der zugrundeliegenden Daten erfasst, beschreibt die Vorhersagegenauigkeit (engl.: *prediction accuracy*; Anteil richtig zugeordneter Proben des Vorhersage-Datensatzes) die Effizienz des Modells in Bezug auf neue Proben. Diese beiden Parameter gleichen sich mit steigender Probenanzahl des Training-Datensatzes immer stärker an und erreichen idealerweise 100 %. Umgekehrt erhält die Vorhersagegenauigkeit bei geringeren Probenmengen meist niedrigere

Werte, weil die Datengrundlage des Training-Datensatzes nicht ausreicht um sämtliche biologische Varianzen zu umfassen, wodurch die Vorhersagemodelle mit sinkender Probenzahl immer unrepräsentativer für die Grundgesamtheit werden. Dadurch werden Proben, die etwas stärker vom Mittelwert abweichen, nur unzureichend oder gar nicht durch die Modellgrundlage der jeweiligen Gruppen beschrieben und folglich falsch klassifiziert. Daher sollte das Ziel stets die Verwendung einer größtmöglichen Probenmenge sein, damit die entsprechenden Modelle bestmögliche Resultate erzielen können. Aus diesem Grund stellen die in der vorliegenden Arbeit erstellten Modelle nur eine Momentaufnahme dar, die zukünftig durch die Implementierung weiterer Proben aus weiteren Erntejahren und/oder Ländern kontinuierlich verbessert werden können. Einen weiteren wichtigen Qualitätsparameter stellt die sogenannte Validierungsgenauigkeit (engl.: validation accuracy) dar, die typischerweise durch Kreuzvalidierung des Training-Datensatzes errechnet wird. Dabei wird der Training-Datensatz in n zufällige Segmente aufgeteilt, von denen anschließend n-1 zur Berechnung des entsprechenden Modells verwendet werden und das jeweils fehlende Segment (Test-Datensatz) mit diesem vorhergesagt wird. Dieses wird für jede mögliche Kombination durchgeführt. Je niedriger die Anzahl der Segmente gewählt wird, desto repräsentativer ist das resultierende Validierungsergebnis, da weniger Proben zur Modellerstellung verwendet werden, und als desto robuster ist das Modell bei einer guten Validierungsgenauigkeit anzusehen. Andererseits ist insbesondere für sehr kleine Probenmengen die Anwendung von wenigen Segmenten nicht sinnvoll, da teilweise komplette Probengruppen für die Modellerstellung fehlen könnten, weshalb für diese Falle häufig eine Full Cross Validation (komplette Kreuzvalidierung) genutzt wird, bei der alternierend jeweils eine einzige Probe nicht zur Modellerstellung verwendet und anschließend vorhergesagt wird. Die Validierung von Vorhersagemodellen ist wichtig um ein sogenanntes Overfitting zu



Abbildung 3-10: Unterschiedliche statistische Fits desselben imaginären Datensatzes zur Klassifizierung von zwei Probengruppen (blau und rot). (modifiziert nach Srivastava 2015)

vermeiden, welches das Modell so exzessiv kompliziert berechnet und spezifiziert, dass es die Training-Daten eher reproduziert statt einen generellen Trend widerzugeben. Daraus resultiert, dass das entsprechende Modell zwar sehr gut die Training-Daten beschreiben kann, aber nur eine sehr eingeschränkte Vorhersagekraft für neue Proben besitzt.<sup>164, 165</sup> Overfitting tritt insbesondere dann auf, wenn die Anzahl der beschreibenden Variablen die Anzahl der Proben übersteigt und/oder wenn das Modell oder der Algorithmus ein geringes Bias aber eine hohe Varianz aufweist.<sup>166, 167</sup> Im Gegensatz dazu tritt sogenanntes Underfitting immer dann auf, wenn das statistische Modell den zugrundeliegenden Trend nicht erfassen kann und einen hohen Bias aber eine geringe Varianz aufweist. Es ist meist das Ergebnis eines zu simplen Modells, welches die Komplexizität der Daten nicht beschreiben kann und ist ebenso wie Overfitting unerwünscht.<sup>166, 168</sup> Abbildung 3-10 visualisiert den Effekt von Overfitting bzw. Underfitting auf einen zu klassifizierenden Datensatz.

## 3.4.3 Multivariate Analysenmethoden

### 3.4.3.1 Multiple Lineare Regression (MLR)

Die multiple lineare Regression ist als Erweiterung der einfachen linearen Regression mit i erklärenden Variablen (Regressor Variablen) zu verstehen, bzw. umgekehrt stellt die einfache lineare Regression einen Spezialfall der multiplen linearen Regression mit i = 1 dar. Die Beziehung zwischen den erklärenden Variablen  $X_k$  (k = 1, 2, ..., i-1) und den abhängigen Variablen Y wird durch folgende Gleichung repräsentiert:

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2 + \ldots + \boldsymbol{\beta}_i \boldsymbol{x}_i + \boldsymbol{e}_i$$

mit  $\beta_0$  als Achsenabschnitt (Mittelwert von Y, wenn alle  $X_k = 0$ ) der Regressionshyperebene, ei als Störgröße (entspricht unbeobachtbaren Zufallsvariablen) und jedes  $\beta_k$  als Steigung zugehörig zu  $X_k$ .  $\beta_k$  werden daher manchmal auch als partielle Regressionskoeffizienten (engl.: *partial regression coefficients*) bezeichnet.<sup>170, 171</sup>

Die Bezeichnung "linear" resultiert aus der Tatsache, dass die abhängigen Variablen eine Linearkombination der Regressionskoeffizienten darstellen. In der MLR erfolgt die Regression nach dem Prinzip der kleinsten Fehlerquadrate zwischen den tatsächlichen und den vorhergesagten Werten der abhängigen Variablen. Auf diese Weise werden diejenigen Regressionskoeffizienten  $\beta$  gesucht, welche die Störgröße e minimieren.<sup>172</sup>

### 3.4.3.2 Hauptkomponentenanalyse (PCA)

Ziel der PCA ist die Datenreduktion und Interpretation eines komplexen Datensatzes durch Projektion der multidimensionalen Daten in einen niederdimensionalen Raum unter Darstellung der größten Varianzen des zugrundeliegenden Datensatzes. Es handelt sich dabei um eine *unsupervised* Methode. Dazu werden die gemessenen Ausgangsdaten (Merkmale oder Variablen) durch Linearkombination in neue, sogenannte latente Variablen (Faktoren oder Hauptkomponenten, kurz PCs) transformiert, welche untereinander unkorreliert und in absteigender Varianz geordnet sind.<sup>173, 174</sup> Diese bilden ein neues Koordinatensystem, welches

darauf ausgerichtet ist, die maximal mögliche Varianz der Originaldaten abzubilden (siehe Abbildung 3-11). Neben der Dimensionsreduzierung wird hierbei eine Erhöhung des Signal-Rausch-Verhältnisses erreicht. Ursache hierfür ist die gleichmäßige Verteilung des Rauschanteils über alle PCs, während die analytische Information in der Regel in den ersten PCs konzentriert ist. Die PCs sind per Definition orthogonal zu der jeweils vorangegangenen und repräsentieren einen bestimmen Umfang der in den Originaldaten enthaltenen Gesamtinformation. Durch die mathematische Transformation der Originaldatenmatrix X in eine Scoresmatrix T und eine Faktorenmatrix P (siehe Abbildung 3-12) können versteckte



Abbildung 3-11: Prinzip der Ermittlung der Hauptkomponenten für eine PCA Berechnung. Variable 1 besitzt dabei eine größere Varianz als Variable 2 und besitzt dementsprechend einen höheren *Loadings* Wert in Bezug auf die PC1, da ihr Beitrag zu dieser Hauptkomponente größer ausfällt.

Muster und Zusammenhänge durch die Darstellung der Objekte im neu berechneten Koordinatensystem der Matrix T (sogenannter *Scores Plot*) identifiziert und visualisiert werden. Orthogonale Rotationstransformation und damit die Hauptkomponentenanalyse lassen sich auf ein Eigenwertproblem zurückführen, wobei die Spalten der Faktorladungsmatrix durch die Eigenvektoren der Kovarianz- bzw. Korrelationsmatrix gebildet werden. Der Betrag der Eigenwerte ist ein direktes Maß für den Varianzanteil eines Faktors an der Gesamtvarianz der Datenmatrix. Häufig verwendete Algorithmen sind NIPALS (*Nonlinear Iterative Partial Least Squares*) oder Singulärwertzerlegung (*Singular Value Decomposition*; SVD).<sup>175</sup>



Abbildung 3-12: Mathematische Matricenfunktion der Hauptkomponentenanalyse. (modifiziert nach Kessler 2007)

Der *Scores Plot* stellt die Position der Objekte im neu berechneten Koordinatensystem (Faktorenraum) dar (z.B. PC1 vs. PC2). Gleichzeitig enthält die Matrix P Informationen über Korrelation zwischen den Originalvariablen (hier: Metabolite) und den Hauptkomponenten und stellt die Einflussnahme der einzelnen Variablen als sogenannte *Loadings* im *Loadings Plot* dar. Je größer der *Loadings* Wert einer Variable in der entsprechenden Hauptkomponente, desto größer ist der Einfluss auf den *Scores Plot* und dementsprechend größer die Varianz dieser Variable im Originaldatensatz.<sup>146, 147, 159</sup> Vereinfacht ausgedrückt für die vorliegende Problemstellung stellt ein Metabolit, dessen Position im *Loadings Plot* weit vom Mittelpunkt entfernt ist eine potente Markersubstanz zur Differenzierung der geographischen Herkunft von Haselnüssen dar (sofern der *Scores Plot* diskrete Unterschiede zwischen den Probenpopulationen beinhaltet). Die Residuenmatrix E enthält die Differenz zwischen ursprünglicher Datenmatrix und den über *Scores* und *Loadings* reproduzierten Daten und damit den Teil der Originaldaten, der varianzbezüglich nur wenig relevant ist. Besteht die Datenmatrix aus N Proben mit jeweils M Variablen, so werden bei der PCA A (A ≤ N; A ≤ M) Hauptkomponenten gebildet, die jeweils N *Scores* und M *Loadings* aufweisen.<sup>176</sup>

Die PCA ist eine der gängigsten Methoden in der multivariaten Statistik, besitzt jedoch den Nachteil, dass sie häufig nicht in der Lage ist, die optimale Struktur der Variablen zu finden, welche für eine Diskriminierung der Probenpopulationen geeignet ist, da sie lediglich nach der größten Varianz des Datensatzes sucht. Diese muss jedoch nicht zwangsläufig durch Gruppenunterschiede in Bezug auf die aktuelle Fragestellung vorgegeben sein, sondern kann durch jede mögliche Eigenschaft hervorgerufen werden. Falls die Sortenunterschiede bei Haselnüssen beispielsweise größer sind als die der geographischen Herkunft, so enthalten die ersten PCs wahrscheinlich keinerlei relevante Informationen in Bezug auf die Diskriminierung der geographischen Herkunft.<sup>159</sup> Ein weiterer Nachteil der PCA ist, dass dem Modell

durch die Orthogonalität der Hauptkomponenten eine starre Struktur auferlegt wird, welche zwar gut zur Extraktion größtmöglicher Informationen aus den Proben geeignet ist, aber natürliche Erscheinungen in den entsprechenden PCs nicht gut separiert, da die meisten natürlich auftretenden Eigenschaften nicht orthogonal sind.<sup>146</sup>

# 3.4.3.3 Partial Least Squares Regression bzw. Partial Least Squares Discriminant Analysis (PLS-R bzw. PLS-DA)

Im Gegensatz dazu stellt die PLS als supervised Methode eine Modifizierung der PCA dar, bei der gezielt diejenigen Faktoren gesucht werden, die für die Datenmatrix X (X =  $T \cdot P^T + E$ , vgl. Abbildung 3-12, Abschnitt 3.4.3.2) und einer zuvor definierten Zielgrößenmatrix Y (analog.  $Y = U \cdot Q^T + F$ ) die maximale Varianz repräsentieren. Eine Variation der ursprünglichen PLS ist die PLS-DA, bei der Y statt numerischer Daten kategorische Informationen (z.B. geographische Herkunft) beinhaltet. In der Praxis wird dies durch die Einführung von Dummy-Variablen zur Definition der Gruppenzugehörigkeiten erreicht. Prinzipiell wird bei der PLS sowohl eine PCA der X-Daten, als auch eine PCA der Y-Daten durchgeführt. Diese sind allerdings nicht unabhängig voneinander, sondern jeweils beeinflusst durch die Ergebnisse der anderen PCA.<sup>176</sup> In der PLS wird die Kovarianz zwischen den unabhängigen Variablen X und denn korrespondierenden abhängigen Variablen Y durch systematische Rotation der PCA Hauptkomponenten maximiert.<sup>177</sup> Die so berechneten Faktoren werden als PLS Komponenten bezeichnet und sind vergleichbar mit den Hauptkomponenten der PCA. Diese Faktoren beschreiben entsprechend das Verhalten der abhängigen Variablen Y und bilden ein neues Koordinatensystem, in das die unabhängigen Variablen X projiziert werden (Scores Plot). Der Vorteil dieser Vorgehensweise ist die Möglichkeit relevante Informationen aus höchst kollinearen und komplexen Daten mit viel Untergrundrauschen zu extrahieren, wie sie bei Metabolomics Experimenten häufig anzutreffen sind.<sup>178</sup>

### 3.4.3.4 Soft Independent Modelling of Class Analogy (SIMCA)

Wie bereits beschrieben, ist der Einsatz der PCA zuweilen wenig effizient zur Differenzierung definierter Gruppen, da die Gruppeninformationen nicht zur Modellerstellung genutzt werden, sondern die generelle Variation des Datensatzes erfasst wird. Mit dem SIMCA Verfahren kann jedoch durch ebendiese Kombination der PCA mit Gruppeninformationen ein Klassifizierungsmodell für unbekannte Proben erstellt werden.<sup>179</sup> Die Bezeichnung "*Soft"* steht dabei dafür, dass keine Hypothese über die Verteilung der Variablen erfolgt, und "*Independent"* für die separate Modellerstellung der Klassen. Dabei wird für jede der X zu



klassifizierenden Gruppen ein individuelles PCA Modell und Hyperfläche daher im ydimensionalen Raum mit den jeweils zugehörigen Proben erstellt. Da die Anzahl der Hauptkomponenten (y) für jedes der unterschiedlichen Modelle variieren kann, muss die optimale Anzahl an Hauptkomponenten individuell bestimmt werden, typischerweise über

Abbildung 3-13: Prinzip der Klassifizierung bei der SIMCA Approximation am Beispiel eines zweidimensionalen Raumes. (modifiziert nach Henrion und Henrion 1994)

Kreuzvalidierung.<sup>180</sup> Dies ist insbesondere bei biologischen Proben sinnvoll, da sich auch die Varianz der verschiedenen Probengruppen stark unterscheiden kann. Das SIMCA Modell definiert somit X Sub-Räume (Klassen Modelle), in die eine unbekannte Probe projiziert wird. Anschließend wird überprüft, ob sie sich mit einer vorgegebenen Genauigkeit innerhalb dieser Hyperfläche befindet. Die Distanz einer Probe zum Klassenmodell wird über die orthogonale Distanz zum Modell-Raum und die Distanz zum Scores-Raum berechnet. Sie wird einer bestimmten Klasse zugeordnet, sofern die quadrierte Distanz des Testobjektes in einem F-Test nicht signifikant von der Zufallsstreuung der Trainingsobjekte des betreffenden PCA Modells abweicht. Eine andere Möglichkeit der Berechnung der Distanzen sind Hotelling's T2 oder Mahalanobis Distanzen, welche auch häufig zur Ausreißerdetektion bei der PCA-Berechnung verwendet werden.<sup>160, 181</sup> Das Klassifizierungsprinzip ist in Abbildung 3-13 dargestellt. Bei sich überlappenden Gruppen sind auch Mehrfachzuordnungen möglich. Eines der wesentlichen Nachteile der SIMCA ist die Tatsache, dass die Berechnung der jeweiligen PCA Modelle darauf ausgelegt ist die maximale Varianz innerhalb der jeweiligen Klasse zu beschreiben. Daher bleibt eine gezielte Suche nach Möglichkeiten zur Separierung der Klassen aus, wodurch die Effizienz dieser Methode limitiert ist. Die SIMCA ist daher oft fehleranfällig bei komplexen Daten, deren Inter-Klassen-Varianz kleiner als die Intra-Klassen-Varianz ist, wie es bei Metabolomics-Daten häufig zu erwarten ist.<sup>182</sup> Einer der Vorteile dieser Methode ist jedoch genau diese unabhängige Behandlung der Lerngruppen, wodurch die gesamte Rechnung sukzessive Gruppe für Gruppe aufgebaut werden kann und leicht zusätzliche Gruppen einbezogen werden können, während das Ergebnis in der linearen Diskriminanzanalyse stets von der Gesamtheit der Probengruppen abhängt.<sup>183</sup>

### 3.4.3.5 Lineare Diskriminanzanalyse (LDA)

Die lineare Diskriminanzanalyse ist ein der PCA verwandtes Verfahren, wobei ebenfalls geeignete latente Variablen als Linearkombinationen zur bestmöglichen Beschreibung des Datensatzes gesucht werden. Die Transformation des Koordinatensystems erfolgt hierbei jedoch nicht entlang der größten Varianz im Datensatz, sondern in der Richtung der bestmöglichen Trennung der zuvor definierten Klassen (siehe Abbildung 3-14). Die LDA basiert auf der Annahme, dass innerhalb der Gruppen Normalverteilung herrscht und die Gruppen identische Kovarianz-Matrizen und damit die gleiche Struktur der Variabilität aufweisen. Typischerweise beinhaltet die herkömmliche Bezeichnung LDA die FISHER'sche Diskriminanzfunktion, bei der eine Hyperebene mit optimaler Projektionsrichtung gesucht wird, welche die Inter-Klassen-Varianz maximiert und die Intra-Klassen-Varianz minimiert.

Anschließend wird diejenige lineare Funktion auf Basis der Originalvariablen gesucht, die die Klassen optimal separieren kann.<sup>184, 185</sup> Die Transformation der Originaldaten (engl.: input space) in geeignete Hyperräume (engl.: *feature space*) kann durch lineare oder nichtlineare Funktionen erfolgen.<sup>185</sup> Zur Unterscheidung von zwei Gruppen ist nur eine Diskriminanzfunktion notwendig, während die Trennung von m Gruppen mehrere Diskriminanzfunktionen erfordert. Anschließend können unbekannte Objekte mithilfe von Klassifikationsfunktionen in die jeweiligen Klassen eingeordnet werden. Für jedes einzuordnende Objekt wird über diese Klassifikationsfunktionen ein Classification



Abbildung 3-14: Vergleich der Prinzipien zur Bestimmung der neuen latenten Variablen im Zuge einer LDA bzw. PCA am Beispiel eines 2-Klassen-Problems (rot und blau) mit linearer Transformation. (modifiziert nach Webb 2003)

*Score* berechnet. Anschließend wird das Objekt der Gruppe mit dem größten Wert zugeordnet. Die Anzahl der Klassifikationsfunktionen entspricht der Gruppenzahl m, während die Anzahl der Diskriminanzfunktionen maximal m-1 beträgt. Weiterhin kann die Zuordnung über die minimale Mahalanobis-Distanz zum Zentroid der Klasse erfolgen.<sup>183, 186, 187</sup> Die Nachteile dieser Methode liegen vor allem darin begründet, dass sie eine Inter-Klassen-Kovarianz-Matrix verwendet, wodurch das Ergebnis nicht immer ideal ist, wenn die Varianzstruktur der unterschiedlichen Klassen sehr unterschiedlich ist. Darüber hinaus muss die Anzahl der Proben in jedem Fall größer als die der Variablen sein, sodass eine inverse

Kovarianz-Matrix erhalten werden kann.<sup>188</sup> Daher kann eine LDA bei Metabolomics-Applikationen häufig nicht direkt angewendet werden, da meist zu wenig Proben vorliegend sind. In diesen Fällen ist eine Dimensionsreduktion der Variablen, z.B. durch eine vorgeschaltete PCA, eine mögliche Lösung um anschließend eine herkömmliche LDA auf die relevanten Hauptkomponenten anzuwenden. Für diese sogenannte PCA-LDA muss die optimale Anzahl der Hauptkomponenten typischerweise via Kreuzvalidierung im Vorfeld bestimmt werden.<sup>159</sup> Die LDA eignet sich insbesondere für lineare und niederdimensionale Daten, während die Resultate bei multidimensionalen Daten meist nur limitiert nutzbar sind.<sup>189</sup>

### 3.4.3.6 Support Vector Machines (SVM)

Die *Support Vector Machine* ist eine effiziente, nicht-parametrische Methode des maschinellen Lernens, welche sowohl bei Klassifizierungs- (SVM-C) als auch Regressionsproblemen (SVM-R) angewendet werden kann.<sup>190</sup> Der wesentliche Vorteil der SVM liegt in der Flexibilität der Wahl der sogenannten Kernel Funktion, welche die Separierung der verschiedenen Klassen erlaubt, und in der Anwendbarkeit dieser Kernel sowohl für lineare als auch nichtlineare Probleme. Bei nichtlinearen Trennproblemen wird der sogenannte Kernel Trick verwendet, welcher die Originaldaten in eine höherdimensionalen Hyperraum transformiert, in dem eine lineare Trennung ermöglicht wird (siehe Abbildung 3-15).<sup>177</sup> Die SVM ist dabei speziell für die Analyse von Stichproben geeignet. Sie verfolgt das Ziel eine optimale Lösung mit den vorhandenen Informationen zu erzeugen, anstatt einen idealen Wert bei unendlicher Anzahl an Proben zu ermitteln.<sup>72</sup> Sie ist dadurch in der Lage selbst bei sehr komplexen Datensätzen mit wenig Proben und starken Überlappungen der einzelnen Klassen ideale Modelle zu berechnen, um zukünftige Probe zuverlässig vorhersagen zu können und eignet sich daher hervorragend für den Metabolomics-Bereich.<sup>191</sup>

Zu den möglichen Kernel Funktionen gehören lineare, polynomiale und sigmoide Funktionen sowie Radial Basis Function (RBF, auch als "Gaussian" bezeichnet).

	(X·Y	Linear )
$K(X, Y) = \langle$	$(\gamma \cdot X \cdot Y + C)^d$	Polynomial
	$\exp(-\gamma \cdot  X - Y ^2)$	Radial Basis (
	$tanh(\gamma \cdot X \cdot Y + C)$	Sigmoid )

mit  $K(X, Y) = \Phi(X) \cdot \Phi(Y)$ , das als Kernel Funktion das Skalarprodukt der Originaldaten nach Transformation ( $\Phi$ ) in einen höherdimensionalen Hyperraum repräsentiert. X stellt dabei die Originaldatenmatrix und Y die zuvor definierte Zielgrößenmatrix (bei Klassifizierung z.B.



Abbildung 3-15: Prinzip des SVM. (A) Anwendung bei linear separierbaren Datensätzen, in denen der Algorithmus in den Originaldaten nach einer optimalen Hyperebene sucht (Gerade bei 2D Darstellung), welche die Spannweite zwischen den beiden Gruppen (rot und blau) maximiert. (B) Anwendung bei nichtlinearen Fällen (hier "polynomial" Kernel), in denen der Algorithmus die Originaldaten (links), welche nicht linear separabel ist (Mitte), in einen höherdimensionalen Hyperraum transformiert. Dort wird eine Hyperebene gesucht, welche die beiden Gruppen optimal voneinander separieren kann. (modifiziert nach Gromski 2015)

geographische Herkunft) dar. Die eigentliche Trennung wird durch die Definierung einer kleinen Anzahl an Proben erreicht, welche die Position der separierenden Hyperebene bestimmen und dieser am nächsten liegen. Diese werden als Stützvektoren (engl.: *support vectors*) bezeichnet. Sie werden durch die SVM so berechnet, dass um die Klassengrenzen herum ein möglichst breiter Bereich (Spannweite, engl.: *margin*) frei von Objekten bleibt (*large margin classifier*). Damit auch in Fällen von nicht perfekt separierbaren Daten eine gute, generalisierte Klassifizierung erreicht werden kann, werden sogenannte Schlupfvariab-

len (engl.: *slacking variable*) eingeführt um das Modell flexibler gestalten zu können. Diese erlauben es dem Modell, dass einzelne Objekte falsch klassifiziert werden, bevor die Komplexizität des Modells zu stark erhöht wird. Dies würde zwar die perfekte Beschreibung der Training-Daten ermöglichen, jedoch fehlte einem solchen Modell die Allgemeingültigkeit auch für zukünftige Proben, sodass diese falsch klassifiziert werden würden. Durch die Einführung der Schlupfvariablen wird darüber hinaus gleichzeitig die Anzahl der benötigten Stützvektoren verringert. Dieses Prinzip wird auch als *soft margin* (dt.: weiche Grenze) bezeichnet.<sup>177, 191, 192</sup> Die optimalen Werte der veränderbaren Parameter  $\gamma$  (Kernel Parameter) und C (*Soft Margin Parameter*) der einzelnen Kernel Funktionen sind abhängig von der Datengrundlage und müssen individuell für jeden Datensatz bestimmt werden, typischerweise

via Kreuzvalidierung um ein Overfitting übermäßige durch Komplexizität des Modells mit bereits beschriebenen Folgen zu vermeiden (vgl. zusätzlich Abschnitt 3.4.2). C reguliert die falsche Klassifizierung der Training-Daten, geringer wobei ein Wert die Entscheidungsoberfläche weich hält, wodurch mehr Schlupfvariablen erlaubt werden (soft margin). Ein hoher Wert zielt darauf ab, alle Training-Daten korrekt ZU klassifizieren, indem es dem Modell die Freiheit zur Selektion von mehr Stützvektoren zur klareren Definierung der Hyperebene gibt (hard margin).  $\gamma$  wiederum definiert die Reichweite des Einflusses jeder



Abbildung 3-16: Darstellung der Effekte für unterschiedliche Einstellungen der SVM Parameter (C und  $\gamma$ ) auf die resultierenden separierten Hyperebenen (projiziert auf die Originaldaten). (A) kleiner bzw. großer Wert des *Soft Margin* Parameters C, und (B) kleiner bzw. großer Wert des Kernel Parameters  $\gamma$ . (modifiziert nach Xu 2006)

einzelnen Training-Probe. Niedrige Werte bedeuten, dass auch weit von der separierenden Hyperebene entfernte Proben einen starken Einfluss auf die Ausrichtung und Form der Hyperebene besitzen, während große Werte nur nahe Proben berücksichtigen (siehe Abbildung 3-16).<sup>146, 190, 191, 193</sup> Während zu große Werte der beiden Parameter zwar zu einer idealen Separierung der Training-Objekte führen, ist das Modell jedoch für unbekannte

Proben wegen der fehlenden Generalisierung für realistische Klassenunterschiede nicht geeignet. Im Gegensatz dazu werden die Grenzen bei zu niedrigen Werten so undefiniert, dass das Modell allgemein sowohl für die Training-Objekte als auch unbekannt Proben keine sinnvolle Klassendifferenzierung ermöglicht.<sup>185</sup>

Im Vergleich zur PLS-DA und LDA wird die SVM nicht durch die Varianz der verschiedenen Probenklassen beeinflusst, sondern beschränkt sich darauf zu prüfen, auf welcher Seite der Stützvektoren eine unbekannt Probe liegt.<sup>193</sup> Dadurch zeichnet sich die SVM im Vergleich zu anderen multivariaten Verfahren durch eine hohe Robustheit sowie Resistenz gegenüber chemischem Rauschen aus.<sup>162</sup> Die Nachteile dieser Methode sind vor allem die fehlende Transparenz der Ergebnisse, die Schwierigkeit der Identifikation von wichtigen, einflussreichen Variablen (Metaboliten) sowie die Komplexizität und der Aufwand der Modellerstellung mit Wahl der optimalen Kernel Funktion bei idealen Parametern für  $\gamma$  und C.<sup>177</sup> Letzteres stellt mittlerweile durch Verwendung geeigneter und bedienerfreundlicher statistischer Software kein Problem mehr dar. Durch (semi-)automatisierte Rastersuchfunktionen (*Grid Search Function*) werden in kürzester Zeit die optimalen Parameter bestimmt, während das Overfitting gleichzeitig durch parallele Kreuzvalidierung geringgehalten werden kann.

Ursprünglich wurde die SVM für Zwei-Klassen-Probleme entwickelt. Allerdings enthalten die meisten Datensätze mehr als zwei Klassen, sodass verschiedene Strategien zur Erweiterung der SVMs entwickelt wurden. Der prinzipielle Ansatz verfolgt die Teilung des Multi-Klassen-Problematik in mehrere binäre Fälle und die anschließende Kombinierung des Outputs aller *"sub-binary classifier"* zu einer finalen Vorhersage der Gruppenzugehörigkeit einer Probe. Die drei Hauptmethoden nach diesem Prinzip werden als "Einer-gegen-Alle" (engl.: *one-against-all*), "Einer-gegen-Einen" (engl.: *one-against-one*) und "DAG-Baum" (engl.: *directed acyclic graph tree*) bezeichnet.<sup>191</sup> Durch deren Anwendung sind mittlerweile auch Multi-Klassen-Probleme zugänglich geworden.

# 4 Zielsetzung der Arbeit

Das grundlegende Ziel der vorliegenden Arbeit war die Entwicklung von LC-MS-basierten Metabolomics-Applikationen zur analytischen Bestimmung der geographischen Herkunft von Haselnüssen. Die bisherige Praxis beschränkt sich auf die Prüfung von Lieferdokumenten, welche durch geschickte Manipulationen gezielt getäuscht werden kann.

Dazu sollten zunächst in einem non-targeted Ansatz mittels hochauflösendem UPLC-ESI-QTOF-MS Metabolite identifiziert werden, die signifikante Unterschiede zwischen wirtschaftlich relevanten Anbauländern aufweisen und sich somit als Markersubstanzen eignen.

Nach einer Strukturaufklärung der relevanten Metabolite sollte eine einfache und schnelle targeted HPLC-ESI-QqQ-MS/MS Applikation entwickelt werden, mit der eine absolute Quantifizierung über interne Standards ermöglicht wird. Die Entwicklung dieser Methode sollte insbesondere in Hinblick auf eine mögliche Implementierung in die Routineanalytik in staatlichen Untersuchungsämtern, Qualitätssicherungsabteilungen in Haselnuss verarbeitenden Unternehmen oder Dienstleistungslaboratorien erfolgen und daher einen besonderen Fokus auf eine möglichst einfache und schnelle Handhabung bei möglichst geringem apparativen Aufwand und Kosten legen. Gleichzeitig sollte die Methode sensitiv und selektiv sein sowie eine robuste und reproduzierbare Quantifizierung der Markersubstanzen ermöglichen.

Um die Eignung dieser Applikation für Routineanwendungen zu beweisen, sollte eine Methodenvalidierung gemäß anerkannter Richtlinien erfolgen.

Anschließend sollten anhand authentischer Haselnussproben und mittels multivariater Auswerteverfahren statistische Klassifizierungs- und Regressionsmodelle entwickelt und validiert werden, die eine zuverlässige Vorhersage des Herkunftslandes von unbekannten Proben ermöglichen.

Abschließend sollte die Stabilität der zuvor identifizierten Markersubstanzen in Haselnuss-Analysenmaterial durch Lagerungsversuche bei unterschiedlichen Bedignungen überprüft werden.

# 5 Ergebnisse und Diskussion

Teile der hier dargestellten Ergebnisse, Abbildungen oder Tabellen wurden bereits in wissenschaftlichen Publikationen veröffentlicht und werden im Folgenden mit "Klockmann 2016" bzw. "Klockmann 2017" zitiert.<sup>194, 195</sup>

Copyright:

Reprinted or adapted with permission from Klockmann, S.; Reiner, E.; Bachmann, R.; Hackl,
T.; Fischer, M. Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS. J. Agric. Food Chem.
2016, 64, 9253-9262. Copyright 2016 American Chemical Society.

Reprinted or adapted with permission from Klockmann, S.; Reiner, E.; Cain, N.; Fischer, M. *Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application*. J. Agric. Food Chem. **2017**, *65*, 1456-1465. Copyright 2017 American Chemical Society.

# 5.1 Methodenentwicklung

Der Workflow eines Metabolomics-Experiments umfasst die Abschnitte Probenvorbereitung, Extraktion und apparative Analytik, welche alle einen unmittelbaren Einfluss auf die Ergebnisse der Messungen haben (vgl. Abschnitt 3.2.2). Ziel der Methodenoptimierung ist es daher, jeden dieser Schritte derart zu gestalten, dass ein Maximum an Informationen bei minimalem Aufwand generiert werden kann, wobei die metabolische Momentaufnahme möglichst unverändert bleiben soll. Ein wichtiger Faktor ist daher das stetige Arbeiten und Lagern von Proben bei niedrigen Temperaturen, um keine unerwünschten biochemischen Reaktionen zu beschleunigen oder thermolabile Metabolite zu zerstören. Der Probenumfang jeder Einzelprobe an Haselnusskernen betrug ca. 200 - 1000 g. Dabei wurden 200 g als absolutes Minimum definiert, das bei eindeutig definierten Proben noch ausreichend statistisch repräsentativ ist. Das Probenmaterial wurde nach vorangegangenem Schockgefrieren mittels Flüssigstickstoff (-196 °C) und unter Zugabe einer äquivalenten Menge Trockeneis (-78 °C) zum Mahlgut zur Kompensation der während des Mahlprozesses freiwerdenden Reibungswärme mittels einer Messermühle homogenisiert. Trotz des hohen Fettgehaltes der Haselnüsse blieb das Homogenisat durch das Gefrieren darüber hinaus pulverförmig, was die Verarbeitung erleichterte. Zur Verbesserung der Lagerfähigkeit wurde

das Homogenisat anschließend gefriergetrocknet. Durch den Wasserentzug werden die meisten Enzyme in ihrer Aktivität inhibiert.<sup>258</sup> Für eine unmittelbare Analyse der Proben ist die Lyophilisation des Mahlguts allerdings nicht unbedingt erforderlich, da geringfügige Unterschiede im Wassergehalt im Zuge der Normalisierung der Datensätze während der Datenvorbereitung mathematisch herausgerechnet werden.<sup>159</sup>

Die Wahl der Extraktionsmethode besitzt einen direkten, maßgeblichen Einfluss sowohl auf das detektierte metabolische Profil als auch auf die Datenqualität und sollte daher eine besondere Aufmerksamkeit erhalten.<sup>196</sup> Duportet et al. (2011) zeigten beispielsweise, dass sich bei Verwendung vier verschiedener Extraktionsmethoden auf eine Probe signifikant unterschiedliche Metabolit-Gehalte ergeben und dadurch gegensätzliche biologische Interpretationen von aktiven Stoffwechselwegen möglich sind. Die ideale Extraktionsmethode für LC-MS-basierte Metabolomics Analysen von biologischen Proben sollte daher (1) unselektiv, (2) einfach und schnell mit einer minimalen Anzahl an Arbeitsschritten und (3) reproduzierbar sein sowie (4) einen Metabolismus-Quenching Schritt (insbesondere Proteinpräzipitation und Kühlung) beinhalten.<sup>69</sup> Bei pflanzlichen Proben sollte darüber hinaus ein geeignetes Zellaufschlussverfahren angewendet werden um Zellwände zu zerstören und die Metabolite aus dem Zellinneren zugänglich zu machen.<sup>198</sup> Da es aufgrund der hohen chemischen Diversität und den damit verbundenen Polaritätsunterschieden keine Extraktionsmethode geben kann, die das gesamte Metabolom in einem Extrakt vereinigen kann, mussten für polare bzw. unpolare Metabolite zwei komplementäre Ansätze entwickelt werden. Die anschließende massenspektrometrische Akquisition unterteilte sich je nach Fragestellung in einen targeted Ansatz (HPLC-ESI-QqQ-MS/MS) und einen non-targeted Ansatz (UPLC-ESI-QTOF-MS).

# 5.1.1 Non-targeted LC-MS

# 5.1.1.1 Methodenentwicklung für die Analyse von unpolaren Metaboliten

Die Extraktion der Haselnüsse erfolgte zunächst nach der im Abschnitt 6.1.3.1 beschriebenen Initialmethode mit dem Probenmaterial NB001 (siehe Tabelle 8-16, Abschnitt 8.2). Die Initialmethode wurde in Anlehnung an übliche Vorgehensweisen in publizierten Metabolomics-Studien sowie bereits etablierte Methoden der Forschungsstelle entworfen.<sup>52, 68, 98, 199,</sup> <sup>200</sup> Ausgehend von dieser Methode wurden im Folgenden diverse Parameter getestet und optimiert.

### 5.1.1.1.1 UPLC-ESI-QTOF-MS-Methode

Die verwendeten LC-MS Methoden zur Detektion von unpolaren Metaboliten (NTNP-pos und NTNP-neg; siehe Abschnitt 6.3.1) wurden am Institut bereits in vorangegangenen Arbeiten für die Analyse von Humanplasma entwickelt und optimiert. Zur umfassenden Charakterisierung des Metaboloms erfolgt die Akquisition sowohl im positiven als auch im negativen Ionenmodus. Ausgehend von diesen Methoden wurde die Anwendbarkeit auf Haselnüsse überprüft. Die Methoden waren sowohl in Hinsicht auf die chromatographische Trennung der Metabolite als auch die massenspektrometrische Detektion als selektiv und sensitiv einzustufen und konnten somit mit geringfügigen Anpassungen (z.B. Verlängerung der Equilibrierungsdauer oder Modifizierung der massenspektrometrischen Parameter) übernommen werden. Ein exemplarisches LC-MS Chromatogramm der Messungen eines unpolaren Haselnussextraktes mit der Methode NTNP-pos ist in Abbildung 5-12 (Abschnitt 5.2.2.1) bzw. mit NTNP-neg in Abbildung 8-1 (Anhang, Abschnitt 8.2) dargestellt.

### 5.1.1.1.2 Extraktionsmittel

Die verwendeten Extraktionsmittel bzw. deren Kombination haben einen unmittelbaren Einfluss auf die Extraktionseffizienz bestimmter Metabolite, bedingt durch deren unterschiedliche Polarität und funktionellen Gruppen. Ziel ist die Extraktion einer größtmöglichen Anzahl an Metaboliten mit hoher Signalintensität bei einfacher Handhabung und maximalen Reproduzierbarkeit. Für die Extraktion von unpolaren Metaboliten haben sich in diversen Studien bereits die Lösungsmittel Chloroform (C), Acetonitril (A), Methanol (M) und Isopropanol (I) etabliert,<sup>72</sup> die im Folgenden in 23 verschiedenen Kombinationen für die Extraktion von Haselnussmetaboliten evaluiert wurden (siehe Tabelle 5-1). Die beiden im Metabolomics-Bereich häufig verwendeten Extraktionsmethoden für Lipide nach BLIGH und DYER<sup>201</sup> sowie FOLCH<sup>202</sup> wurden dabei für die vorliegende Fragestellung nicht in Betracht gezogen, obwohl die hervorragende Extraktionseffizienz dieser Methoden für bestimmte Lipide bereits in zahlreichen Studien belegt wurde. Allerdings beruhen sie auf einer verhältnismäßig aufwändigen Zweiphasenextraktion mit zahlreichen Arbeitsschritten und widersprechen somit dem Ziel einer effizienten und dabei unkomplizierten und schnellen Extraktion. Im Zuge der non-targeted Extraktion liegt der Fokus stärker auf der gleichzeitigen Analyse möglichst vieler Metabolite, wobei nicht die Extraktionseffizienz für die einzelnen Metabolite, sondern vielmehr die Reproduzierbarkeit, die höchste Priorität besitzt.

Bei neun der 23 getesteten Varianten konnten zwar nach der Zentrifugation und Membranfiltration der Überstande zunächst klare Extrakte erhalten werden, aus denen jedoch nach einiger Zeit Feststoffe auskristallisierten. Bei vier Extraktionsmitteln konnte trotz Membranfiltration keine klare Lösung erhalten werden und bei weiteren zwei ergaben sich nach der Zentrifugation zwei Phasen. Zusätzlich trat bei vier weiteren nach der Lagerung (ca. eine Woche) eine Trübung der Lösung auf. Lediglich bei acht Varianten wurden stabile, einphasige und klare Extrakte erhalten. Für die LC-MS Analyse ist es zwingend notwendig, dass Lösungen ohne ungelöste Feststoffe vermessen werden um Störungen der Analyse sowie Verschmutzungen sowie Verstopfungen von Kapillaren und der UPLC-Säule zu vermeiden. Aus diesem Grund wurden für die folgende Bewertung nur die vier klaren Extrakte (CI12, CI11, CIA111, CIA221) mittels der Methoden NTNP-pos und NTNP-neg in Dreifachbestimmung analysiert.

Zwar könnten die vier Extrakte, bei denen eine Trübung erst nach einiger Zeit auftrat, prinzipiell unmittelbar nach der Extraktion analysiert werden, jedoch wurden auch diese Extraktionsvarianten nicht in Betracht gezogen, da die Gefahr der Niederschlagsbildung vor der massenspektrometrischen Analyse zu groß ist, insbesondere bei längeren Messreihen. Darüber hinaus könnten derartige Proben nicht gelagert und zu einem späteren Zeitpunkt erneut analysiert werden. Die Niederschläge und die oberen, öligen Phasen der nicht-

Dogoiahnung		Lösungsn	Ergebnis nach			
bezeichnung	CHCl <sub>3</sub>	IPA	MeOH	ACN	Extraktion	
CM12	333		667		2 Phasen	×
CM21	667		333		Niederschlag	×
CM11	500		500		Niederschlag	x
CA12	333			667	2 Phasen	×
CA21	667			333	Niederschlag	×
CA11	500			500	Niederschlag	
ISO		1000			trüb (nach 7 Tagen)	
CI12	333	667			klar	
CI11	500	500			klar	$\checkmark$
CIM111	333	333	333		Niederschlag	×
CIM211	500	250	250		Niederschlag	×
CIM121	250	500	250		Niederschlag	
CIM112	250	250	500		trüb (nach 7 Tagen)	
CIM221	400	400	200		trüb (nach 7 Tagen)	
CIM212	400	200	400		trüb	
CIM122	200	400	400		trüb (nach 7 Tagen)	
CIA111	333	333		333	klar	
CIA211	500	250		250	trüb	
CIA121	250	500		250	Niederschlag	
CIA112	250	250		500	trüb	
CIA221	400	400		200	0 klar	
CIA212	400	200		400	trüb	×
CIA122	200	400		400 Niederschlag		×

Tabelle 5-1: Lösungsmittelkombinationen zur Extraktion von unpolaren Metaboliten und deren Resultate.

Universität Hamburg, Hamburg School of Food Science, 2017

konformen Proben konnten in Chloroform gelöst und massenspektrometrisch analysiert werden (siehe Abbildung 8-2, Abschnitt 8.2). Sowohl die öligen Phasen als auch die Niederschläge besaßen eine ähnliche Zusammensetzung aus überwiegend sehr unpolare Substanzen wie Triacylglycerolen, deren Löslichkeitsprodukt wahrscheinlich überschritten wurde.

Zur Bewertung der Extraktionseffizienz und Reproduzierbarkeit wurde die Peakfläche ausgewählter Metabolite sowie deren Variationskoeffizienten berechnet. Da zum Zeitpunkt der Methodenentwicklung noch keine potentiellen Markersubstanzen für eine geographische Herkunftsbestimmung identifiziert waren, wurden jeweils 32 repräsentative, zufällig gewählte *Features* (enthält Information über Retentionszeit und *m/z*-Verhältnis eines Metaboliten) für NTNP-pos und NTNP-neg ausgewählt. Die *Features* wurden dabei so gewählt, dass sie sowohl den gesamten chromatographischen Zeitraum als auch den gesamten Massenbereich repräsentieren. Abbildung 5-1 zeigt eine graphische Auftragung der mittleren Signalintensitäten als prozentuale Abweichung vom Mittelwert aller vier Extraktionsmittel mit der Methode NTNP-pos und Abbildung 5-2 für NTNP-neg.

Bei der Messung mit der Methode NTNP-pos zeigte sich, dass CIA221 im Massenbereich 200 - 400 m/z die beste Extraktionseffizienz (größte Signalintensität) erzielte, während CI12 im Bereich von 400 - 1000 m/z gegenüber den anderen drei überlegen war. Unter Berücksichtigung, dass die meisten unpolaren Metabolite größere Massen aufweisen (z.B. Triacylglyceride, Phospholipide und Steroide) und auch mit der hier verwendeten Methode NTNP-pos die







Abbildung 5-2: Prozentuale Abweichung vom Mittelwert der Peakflächen der zufällig ausgewählten *Features*, gemessen in NTNP-neg.

meisten Metabolite über 400 m/z detektiert werden konnten, erscheint die Mischung von Chloroform und Isopropanol im Verhältnis 1:2 dementsprechend als bestes Extraktionsmittel im Kontext der zugrundeliegenden Fragestellung. Im Mittel resultierten bei der Extraktion mit CI12 höhere Signalintensitäten für die 32 *Features* als bei der Verwendung von CI11 (24 % niedriger), CIA111 (33 % niedriger) oder CIA221 (14 % niedriger).

Im Fall der Methode NTNP-neg ergaben sich weniger klare Tendenzen. Allerdings lässt sich auch hier schlussfolgern, dass CIA221 (CI11 hier ebenfalls) im niedrigen Massenbereich in der Lage war, Metabolite effizienter extrahieren zu können, während CI12 bei höheren Massen zusammen mit CI11 bessere Ergebnisse erzielen konnte. Gegenüber den anderen drei Kombinationen konnte CI12 insgesamt gleichwertige Ergebnisse erzielen (7 % niedriger als CI21 und 11 % besser als CIA111).

Neben der Extraktionseffizienz spielt auch die Reproduzierbarkeit bei Metabolomics-Experimenten eine essentielle Rolle, da meist großen Probenpopulationen analysiert werden und statistischen Auswerteverfahren verwendet werden müssen. Bei einer größeren Streuung von Messwerten sinkt die statistische Sicherheit, wodurch die statistisch messbare Signifikanz von tatsächlichen Unterschieden verringert wird. Aus diesem Grund wurden die Variationskoeffizienten (CV) für jeden Metaboliten bestimmt, welche eine Aussage über die Genauigkeit der Messung treffen (kleine Werte = geringe Messunsicherheit). In Tabelle 5-2 sind die mittleren Variationskoeffizienten für jedes Extraktionsmittel für die verschiedenen Methoden dargestellt. Während bei NTNP-neg keine großen Unterschiede zu erkennen waren, lieferte CI12 bei NTNP-pos die mit Abstand höchste Genauigkeit.

Bezeichnung	Ø CV zufällige <i>Features</i> NTNP-pos	Ø CV zufällige <i>Features</i> NTNP-neg	Ø CV Markersubstanzen NTNP-pos
CI11	9,53 %	7,28 %	15,60 %
CI12	5,75 %	8,89 %	7,10 %
CIA111	15,03 %	7,40 %	22,25 %
CIA221	19,27 %	9,03 %	21,15 %

Auf Grundlage dieser Bewertungskriterien wurde CI12 als optimales Extraktionsmittel identifiziert, da es insbesondere bei der Messung mit der Methode NTNP-pos die besten Ergebnisse erzielte, während die Unterschiede bei NTNP-neg generell eher gering ausfielen. Hierbei muss allerdings berücksichtigt werden, dass es sich um zufällig ausgewählte Signale handelt, die nicht zwangsläufig in Bezug auf die Fragestellung der geographischen Herkunft von Relevanz sind. Ob CI12 auch für potentielle Markersubstanzen das optimale Extraktionsmittel ist, konnte zu diesem Zeitpunkt noch nicht ermittelt werden. Allerdings stellt die Bestimmung des optimalen Extraktionsmittels die Grundvoraussetzung für die Suche nach geeigneten Markersubstanzen dar und sollte daher vor Beginn der Suche abgeschlossen sein, damit im Anschluss möglichst viele potentiell verwendbare Metabolite detektiert werden können.

Nachdem die in Abschnitt 5.2.2.1 beschriebenen 20 Markersubstanzen identifiziert wurden, wurden die hier verwendeten Datensätze im Nachhinein neu ausgewertet, um zu überprüfen, ob die Ergebnisse auch repräsentativ für diese Metabolite waren oder ob die Wahl des Extraktionsmittels noch einmal überarbeitet werden musste (siehe Abbildung 5-3). Auch bei



# Abbildung 5-3: Prozentuale Abweichung vom Mittelwert der Peakflächen der Markersubstanzen, gemessen in NTNP-pos.

diesen Substanzen zeigte sich eindeutig, dass CI12 am besten geeignet war und im Vergleich zu den anderen drei Lösungsmittelkombinationen intensivere Signale erzielte (gegenüber CI11 +30 %, CIA111 +51 % und CIA221 +29 %). Bei der Reproduzierbarkeit war CI12 mit 7,10 % den anderen ebenfalls deutlich überlegen (siehe Tabelle 5-2), weshalb für diese Extraktionsmethode nachfolgend keine Überarbeitung notwendig war.

### 5.1.1.1.3 Extraktionsmethode

Für den Extraktionsvorgang an sich wurden ausgehend von der Initialmethode (siehe Abschnitt 6.1.3.1) zusätzlich verschiedene Parameter variiert und miteinander verglichen. Dazu wurden sowohl verschiedene Mahlmethoden (Zentrifugalmühle oder Messermühle), Aufschlussmethoden (Ultraschall, Kugelmühle oder eine Kombination aus beiden), Aufschlussdauern und -intensitäten, Stickstoffkühlung während des Mahlvorgangs, Anzahl der Wiederholungen sowie Inkubationszeiten zwischen den Wiederholungen als auch verschiedene Einwaagen getestet (siehe Tabelle 5-3). Die Proben zur Bewertung der Parameter wurden alle mit der Methode NTNP-pos gemessen, da sich die Ergebnisse dieser Methode bis dato als variabler und dadurch aussagekräftiger erwiesen hatten als die der Methode NTNP-neg. Es muss bei der Auswertung beachtet werden, dass die Proben der Kugelmühle-B nicht im selben Batch mit den restlichen Proben gemessen werden konnten. Da die Vergleichbarkeit zwischen Batches aus verschiedenen Messtagen bei QTOF-Analysen generell begrenzt ist und zusätzlich bauliche Veränderungen an dem Massenspektrometer vorgenommen wurden, sind die beiden Batches in diesem Fall als nicht direkt vergleichbar einzustufen. In einem nachgeschalteten Versuch konnten jedoch keine signifikanten Unterschiede zwischen den jeweils optimalen Methoden beider Batches (ExV01 und ExV26) festgestellt werden, sodass die Ergebnisse indirekt zu einander in Beziehung gesetzt werden konnten.

Zum Vergleich der verschiedenen Methoden wurden einerseits die Peakflächen der 32 zufällig ausgewählten, repräsentativen Metabolite (siehe Abbildung 5-4) als auch die Reproduzierbarkeit anhand der CVs miteinander verglichen. Die Variation der Aufschlussdauer, -wiederholungen und der Inkubationszeit, sowie die Wahl der Aufschluss- oder Mahlmethode konnten keine signifikanten Verbesserungen im Vergleich zur Initialmethode erzielen. Sowohl die Reproduzierbarkeit ( $\emptyset$  CV = 5,35 %) als auch die mittlere Signalintensität (durchschnittlich 13 % höher) der Initialmethode sind gegenüber den anderen Methoden im Vorteil. Einzig im niedrigen Massenbereich von 200 - 400 m/z zeigten sich einige Defizite

Bezeich- nung	Einwaage [mg]	Mahl- methode	Aufschluss- methode	Inten- sität	Dauer [min]	Inkubations- zeit [min]	WDH	N2- Küh- lung
			Init	ialmethode	e.			
ExV01	50	MM	KM-A	30 Hz	5	5	3	nein
ExV02	-	_		-	-	-	2	-
ExV03	-	_	-	-	_	-	1	-
ExV04	-	-	-	-	3	3	_	-
ExV05	-	-	-	-	1	1	-	-
ExV06	-	-	US	-	-	-	-	-
ExV07	-	_	US	-	3	3	5	-
ExV08	-	-	US	-	1	1	15	-
ExV09	_	-	US+KM-A	-	-	-	-	-
ExV10	-	-	US+KM-A	-	3	3	5	-
ExV11	-	-	US+KM-A	-	1	1	15	-
ExV12	_	-	-	-	-	10	-	-
ExV13	-	-	-	-	-	15	-	-
ExV14	-	ZM	-	-	-	-	-	-
ExV15	30	-	-	-	-	-	-	-
ExV16	40	-	-	-	-	-	-	-
ExV17	60	-	-	-	-	-	-	-
ExV18	70	-	-	-	-	-	-	-
ExV19	-	-	KM-B	3 m/s	-	-	1	-
ExV20	-	-	KM-B	3 m/s	-	-	2	-
ExV21	-	-	KM-B	3 m/s	-	-	3	-
ExV22	-	-	KM-B	3 m/s	-	-	1	ja
ExV23	-	_	KM-B	3 m/s	-	-	2	ja
ExV24	-	_	KM-B	3 m/s	-	-	3	ja
ExV25	-	_	KM-B	3 m/s	1	-	1	-
ExV26	-	-	KM-B	3 m/s	3	-	1	-
ExV27	-	-	KM-B	3 m/s	7	-	1	-
ExV28	-	-	KM-B	3 m/s	10	-	1	-
ExV29	-	-	KM-B	2 m/s	-	-	1	-
ExV30	-	-	KM-B	4 m/s	-	-	1	-
ExV31	-	-	KM-B	5 m/s	-	-	1	-

Takalla 5 2. V	Vanilanta Fritrialitiana		A house a house a second	Jan Initialmash	de ala de a a affiliat
<b>Expense 5-5</b> :	varuerte fixtraktions	narameter: nur	Abweichungen von (	der initialmetho	ae sina angetunrt.
	and the barren with the barren of the barren	, and a meeter , man	is werenangen von		ae sina angerani o

MM: Messermühle; ZM: Zentrifugalmühle KM-A: Kugelmühle (Quiagen); KM-B: Kugelmühle (BeadRuptor); US: Ultraschall

der Initialmethode. In diesem Bereich boten der Aufschluss mittels Ultraschall sowie eine Erhöhung der Inkubationsdauer zwischen den Wiederholungen größere Peakflächen. Möglicherweise ist dies darauf zurückzuführen, dass größere Moleküle vom Ultraschall leichter fragmentiert werden, während kleinere demgegenüber stabiler sind. Darüber hinaus könnte die Verlängerung der Inkubationsdauer den Phasenübergang von kleineren, polareren Molekülen in die eher unpolare Lösung begünstigen. Da jedoch, wie unter Abschnitt 5.1.1.1.2 bereits beschrieben, bei der Extraktion von unpolaren Metaboliten größere Moleküle bevorzugt extrahiert werden sollten, überwiegen die Vorteile der Initialmethode.

Bei der Extraktion mittels Kugelmühle-B führte eine Erhöhung der Aufschlussintensität auf 4 bzw. 5 m/s zum Flüssigkeitsverlust durch partielle Öffnung der Reaktionsgefäße aufgrund des



Abbildung 5-4: Prozentuale Abweichung vom Mittelwert der Einwaagen-normierten Peakflächen der Proben ExV01, ExV15, ExV16, ExV17 und ExV18 bei den zufällig ausgewählten *Features*, gemessen in NTNP-pos.

hohen Energieeintrages. Daher wurden diese Proben für die folgende Auswertung nicht herangezogen. Die Methode ExV26 erwies sich im Fall der Kugelmühle-B als am besten geeignet, da sie im Mittel 6 % höhere Signale für die 32 Metabolite erzielte als die anderen. Der mittlere CV war mit 9,58 % jedoch etwas über dem Durchschnitt von 8,60 % (ExV19 -ExV31). Im Gegensatz zur Kugelmühle-A, deren Prinzip auf rein eindimensionalen Bewegungen beruhte, werden bei der Kugelmühle-B dreidimensionale Bewegungen genutzt um die Homogenisierungseffizienz zu steigern, während der Temperatureintrag in die Proben gleichzeitig verringert werden soll.<sup>206</sup> Daraus resultierend führte die mehrfache Wiederholung des Mahlvorgangs bei der Kugelmühle-B weder zu einer Verbesserung der Reproduzierbarkeit noch der Extraktionseffizienz, sodass eine einfache Extraktion in diesem Fall ausreichend war. Insbesondere in Hinblick auf die Anwendung von *high-throughput* (dt.: Hochdurchsatz) Analysen im Metabolomics-Bereich ist eine Verkürzung der Probenvorbereitungsdauer ein essentieller Schritt um bei einer hohen Probenanzahl Zeit sparen zu können. Die Methode ExV26 (mit Kugelmühle-B) wurde dementsprechend der Initialmethode ExV01 (mit Kugelmühle-A) vorgezogen, da sie die gleiche Extraktionseffizienz in kürzerer Zeit (3 min im Vergleich zu 25 min der Initialmethode) erreichen konnte und daher als Standardmethode für alle folgenden Analysen festgelegt.

Bei der Wahl der Einwaage (bzw. des Lösungsmittel-zu-Probe-Verhältnisses) fiel auf, dass sich kein linearer Zusammenhang ergab. Für die Auswertung wurden die Peakflächen der Metabolite der jeweiligen Einwaage der Probe in Relation gesetzt, sodass die Extraktionseffizienz als Einwaagen-normierte Signalintensität widergegeben wurde. Abbildung 5-4 zeigt diesen Zusammenhang als graphische Darstellung der prozentualen Abweichung vom Mittelwert. Hierbei zeigte sich, dass eine größere Einwaage dazu führte, dass die Extraktionseffizienz insbesondere im Massenbereich > 800 m/z deutlich abgesenkt wurde, während kleinere Einwaagen in diesem Bereich begünstigend wirkten. Im niedrigeren Massenbereich ließen sich bei schwankenden Tendenzen keine derart eindeutigen Zusammenhänge feststellen. Über den gesamten Massenbereich betrachtet sind Einwaagen von 30 bis 50 mg prinzipiell als sinnvoll anzusehen, da keine klaren Unterschiede festzustellen waren. Dies spricht dafür, dass die Extraktionseffizienz bis zu einer Einwaage von 50 mg quantitativ verlief und nicht weiter gesteigert werden konnte. Unter Berücksichtigung der Tatsache, dass bei kleineren Einwaagen der relative Fehler ansteigt und sich dadurch die Reproduzierbarkeit verschlechtert, wurde die Einwaage folglich auf 50 mg festgesetzt. Diese Hypothese spiegelt sich auch in den Messergebnissen wider, bei denen der mittlere CV bei einer Einwaage von 50 mg nur 5,35 % betrug, wohingegen dieser bei 40 mg mit 16,58 % und 30 mg mit 14,53 % deutlich erhöht war.

Eine Kühlung der Proben während des Mahlvorgangs mittels Flüssigstickstoff führte zu keinem Vorteil in Bezug auf Signalintensität oder Reproduzierbarkeit. Generell ist eine Kühlung der Proben während der Extraktion, wie in diesem Fall durch eine integrierte Flüssigstickstoff-Kühlung, im Kontext von Metabolomics-Experimenten als sinnvoll einzustufen, da der Energieeintrag während des Mahlvorgangs unerwünschte chemische Reaktionen beschleunigen und temperaturlabile Metabolite zerstören kann.<sup>43</sup> Im Gegenzug dazu wird die Extraktionsrate hauptsächlich durch die Kinetik von Phasenübergängen beeinflusst, welche maßgeblich von der Temperatur abhängig sind und sich bei niedrigeren Temperaturen verringern.<sup>197</sup> Für den vorliegenden Fall überwogen die negativen Auswirkungen der Kühlung den positiven, wodurch sich die Signalintensität um 2 - 7 % verringerte. Daher erfolgte im Folgenden keine Kühlung der Probengefäße während des Mahlvorgangs in der Kugelmühle.

# 5.1.1.1.4 Membranfiltration und Stabilität der Extrakte

Da sich bei der Ermittlung des optimalen Extraktionsmittels bereits gezeigt hatte, dass einige der Lösungsmittelkombinationen keine stabilen Extrakte ergaben (vgl. Abschnitt 5.1.1.1.2), wurden für den CI12 Extrakt ergänzend unterschiedliche Bedingungen für den Zeitraum nach der Extraktion bis zur Analyse getestet, um sicher zu stellen, dass durch unterschiedliche
Lagerungstemperaturen oder –dauern nicht ebenfalls nachträglich eine Kristallisation von unpolaren Lipiden auftritt. Zusätzlich sollte dadurch überprüft werden, ob eine Membranfiltration der Überstände mittels Spritzenvorsatzfilter (Rotilabo, PTFE, 0,45 µm, 0,15 mm, Carl Roth) durch geeignete Schritte umgangen werden kann, da diese einen zusätzlichen zeitlichen und finanziellen Aufwand darstellt und zudem unerwünschte Kontaminationen der Extrakte verursachen kann. Aus diesem Grund wurden die Extrakte nach der Zentrifugation entweder direkt in ein Glasvial überführt oder zuvor über Nacht bei unterschiedlichen Temperaturen (4 °C, -20 °C oder -80 °C) inkubiert, erneut zentrifugiert und dann überführt. In gleicher Weise wurden Extrakte erstellt, die vor der Überführung in das Glasvial membranfiltriert wurden. Anschließend wurden die so erstellten Extrakte zunächst direkt auf auskristallisierte Feststoffe geprüft und zusätzlich nach einer 14-tägigen Lagerung bei 4 °C, -20 °C bzw. -80 °C. Die Ergebnisse dieser Versuchsreihe sind in Tabelle 5-4 dargestellt.

Tabelle 5-4: Arbeitsschritte nach der ersten Zentrifugation und deren Auswirkung auf die Stabilität und Nutzbarkeit der unpolaren Extrakte.

Inkubation über Nacht vor 2. Zentrifugation	Membran- filtration	Lagertemperatur des Extrakts im Vial		embran- Lagerten Itration Extrak		Ergebnis direkt nach Extraktion	Ergebnis nach 14 Tagen	
4 °C −20 °C −80 °C		4 °C	-20 °C	-80 °C				
		+			trüb	Niederschlag	x	
			+		trüb	Niederschlag	x	
				+	trüb	Niederschlag	×	
	+	+			klar	klar	<ul><li>✓</li></ul>	
	+		+		klar	klar	<ul><li>✓</li></ul>	
	+			+	klar	klar	<ul><li>✓</li></ul>	
+		+			trüb	trüb	×	
+			+		trüb	trüb	×	
+				+	trüb	trüb	×	
+	+	+			klar	klar	<ul><li>✓</li></ul>	
+	+		+		klar	klar	<ul><li>✓</li></ul>	
+	+			+	klar	klar	<ul><li>✓</li></ul>	
+		+			trüb	trüb	×	
+			+		trüb	trüb	×	
+				+	trüb	trüb	×	
+	+	+			klar	klar	<ul><li>✓</li></ul>	
+	+		+		klar	klar	$\checkmark$	
+	+			+	klar	klar	$\checkmark$	
+		+			trüb	trüb	×	
+			+		trüb	trüb	×	
+				+	trüb	trüb	x	
+	+	+			klar	klar	$\checkmark$	
+	+		+		klar	klar	✓	
+	+			+	klar	klar	$\checkmark$	

Aus den Ergebnissen geht eindeutig hervor, dass eine Membranfiltration der Überstände zwingend erforderlich ist, da ohne diese keine klaren Extrakte gewonnen werden konnten. Die Membranfiltration ist allerdings als kritisch anzusehen, da durch diesen Schritt möglicherweise Metabolite durch Adsorption am Filtrationsmaterial verloren gehen oder Stoffe aus dem Material herausgelöst werden können. Aus diesem Grund wurden die Chromatogramme eines Extraktes mit und ohne Membranfiltration massenspektrometrisch analysiert (siehe Abbildung 8-3, Abschnitt 8.2). Der Vergleich beider Chromatogramme zeigt, dass durch die Extraktionsmittel Chloroform und Isopropanol verschiedene Additive herausgelöst wurden, insbesondere Oleamid (282,279 m/z), Linoleamid (280,263 m/z) und Oleyl-Palmitamid (506,530 m/z), die in vielen Kunststoffprodukten als Gleitmittel (engl.: slip agents) eingesetzt werden um Flüssigkeitsanhaftungen zu vermeiden. Daneben konnten weitere Verunreinigungen detektiert werden, bei denen es sich wahrscheinlich um verschiedene weitere Fettsäureamide sowie Dimere dieser handelte (z.B. 254,248 m/z Palmitoleinamid, 256,263 m/z Palmitamid, 284,294 m/z Stearamid, 338,342 m/z Erucamid, 504,514 m/z Linoyl-Palmitamid, 508,544 m/z Stearyl-Palmitamid). Auch bei Spritzenvorsatzfiltern anderer Hersteller (Hahnemühle, Macherey&Nagel) resultierte ein vergleichbares Bild. Da dies auch bei verschiedenen Membranmaterialien (PTFE und RC) zutraf und alle Hersteller angeben, dass die Membranen intensiven Test unterzogen wurden, in denen die Lösungsmittelbeständigkeit bestätigt wurde, muss davon ausgegangen werden, dass die Additive aus dem Plastik der Membranhalterung stammten, die ebenfalls in Kontakt mit dem Extraktionsmittel gelangten. Weil die Membranfiltration jedoch für die vorliegende Problemstellung unabdingbar ist, um klare, analysierbare Extrakte zu erhalten, wurde diese Problematik bei den folgenden Analysen berücksichtigt. Alle potentielle Markersubstanzen wurden daher im Folgenden dahingehend untersucht, ob diese vollständig oder teilweise als Kontamination durch die Membranfiltration anzusehen sind (Signalerhöhung in einem repräsentativen, membranfiltrierten Extrakt im Vergleich zum entsprechenden, unfiltrierten Extrakt) oder tatsächlich als valide Schlüsselmetabolite zur Herkunftsdiskriminierung von Haselnüssen verwendet werden konnten. Die Lagerungstemperatur hatte keinen messbaren Einfluss auf die Stabilität der Extrakte in Hinsicht auf eine Kristallisation von Feststoffen. Eine ausführliche Untersuchung der Lagerstabilität der Extrakte bei unterschiedlichen Temperaturen in Bezug auf die Konzentration der einzelnen Metabolite ist in Abschnitt 5.6 dargestellt.

Zur Verhinderung von unerwünschten Oxidationsreaktionen der lipidhaltigen Extrakte, die sich insbesondere durch den hohen Anteil an ungesättigten Fettsäuren in Haselnüssen ergeben, wurde dem Extraktionsmittel zusätzlich 0,01 % Butylhydroxytoluol (BHT) als Antioxidans zugegeben. Diese Praxis ist bereits in zahlreichen Lipidomics Studien weit verbreitet.<sup>203-205</sup> Es muss dabei lediglich beachtet werden, dass die detektierbaren Signale des BHT (Rt: 4,2 min; M+H: 221,190 m/z; M+NH4: 238,2165 m/z; M-H: 219,1743 m/z) bei der Auswertung exkludiert werden.

#### 5.1.1.2 Methodenentwicklung für die Analyse von polaren Metaboliten

Analog zur Methodenentwicklung für unpolare Metabolite erfolgte in diesem Fall die Extraktion der Haselnüsse zunächst nach der im Abschnitt 6.1.3.1 beschriebenen Initialmethode mit dem Probenmaterial NB001 (siehe Tabelle 8-16, Abschnitt 8.2). Die Initialmethode wurde in Anlehnung an übliche Vorgehensweisen in Metabolomics-Studien sowie bereits etablierte Methoden der Forschungsstelle kreiert.<sup>102, 115, 198, 207, 208</sup> Anhand dieser Methode wurden im Folgenden diverse Parameter getestet und optimiert.

#### 5.1.1.2.1 UPLC-ESI-QTOF-MS-Methode

Eine initiale Methode zur chromatographischen Trennung von polaren Metaboliten wurde am Institut bereits in vorangegangenen Arbeiten entwickelt. Ausgehend von dieser Methode wurden die in Tabelle 5-5 dargestellten Parameter variiert und optimiert. Bei der verwendeten UPLC-Säule handelt es sich um eine Silica Hydrid Phase, die sich durch die Anwendbarkeit der sogenannten *aqueous normal phase chromatography* (ANP; dt.: wässrige Normalphasenchromatographie) auszeichnet, wodurch die Elution mit herkömmlichen MS gängigen Lösungsmitteln (Wasser, Methanol, Acetonitril, Isopropanol oder Ethanol) ermöglicht wird. Diese Methode wurde der im Metabolomics Bereich meistens genutzten Anwendung einer HILIC-Phase (engl.: *hydrophilic interaction liquid chromatography*) vorgezogen, da diese einige wesentliche Nachteile besitzt (vgl. Abschnitt 3.3.1.2). Als Hauptargument ist hierbei die verhältnismäßig schlechte Reproduzierbarkeit zu nennen.<sup>110, 114</sup>

Tabelle 5-5: Parameter zur Optimierung des flüssigchromatographischen Systems der non-targeted Methode für polare Metabolite.

Parameter	Optionen
	ACN, ACN/EtOH 19/1, ACN/EtOH 9/1, ACN/EtOH 4/1,
Eluent B ( $je + 5 \text{ mM AF}$ )	ACN/EtOH 7/3, ACN/IPA 19/1, ACN/IPA 10/1, ACN/IPA 9/1,
	ACN/IPA 4/1, ACN/IPA 7/3 (v/v)
Gradient	6 verschiedene (siehe Tabelle 8-6, Abschnitt 8.2)
Fließgeschwindigkeit	300 μL/min, 400 μL/min, 500 μL/min
Säulenofentemperatur	6 °C, 15 °C, 25 °C, 30 °C, 35 °C, 45 °C

Das grundlegende Ziel der Methodenentwicklung war eine möglichst gute Auftrennung der polaren Metabolite zur Minimierung der Ionensuppressionseffekte und zur Trennung von isobaren Molekülen zur Generierung einer größtmöglichen Selektivität und Sensitivität. Eine bestmögliche Trennung der polaren Metabolite wurde durch die Verwendung von ACN/IPA 10/1 + 5 mM AF als Eluent B und Gradient-nt6 (siehe Tabelle 8-6, Abschnitt 8.2) bei einer Fließgeschwindigkeit von 400  $\mu$ L/min und Säulenofentemperatur von 30 °C erreicht. Die massenspektrometrische Detektion erfolgte wie auch bei den unpolaren Metaboliten sowohl im positiven (NTP-pos) als auch negativen Ionenmodus (NTP-neg) um die Detektion eines breiten Spektrums an verschiedenen Metaboliten zu ermöglichen. Exemplarische LC-MS Chromatogramme der optimierten LC-Methoden für die Messung mittels NTP-pos und NTP-neg sind in Abbildung 8-4 und Abbildung 8-5 und (Anhang, Abschnitt 8.2) dargestellt.

### 5.1.1.2.2 Extraktionsmittel, Stabilität der Extrakte und Membranfiltration

In ähnlicher Vorgehensweise zur Methodenentwicklung der Extraktion von unpolaren Metaboliten (vgl. Abschnitt 5.1.1.1.2 und Abschnitt 5.1.1.1.4) wurde die Methode für polare Metabolite optimiert. Es wurden verschiedene Lösungsmittelkombinationen in Dreifachbestimmung getestet, wobei unterschiedliche Mischungsverhältnisse von Wasser und Methanol, bzw. Wasser und Acetonitril als polare Lösungsmittel verwendet wurden, die sich in diversen Metabolomics-Ansätzen bereits als geeignet erwiesen.<sup>208-210</sup> Die Extrakte wurden nach der Zentrifugation entweder direkt entnommen oder zusätzlich membranfiltriert. Ausgehend von den Ergebnissen der unpolaren Extraktion wurde auch hier die Bewertung der Extrakte direkt nach der Extraktion sowie nach 14 Tage durchgeführt, um sicherzustellen, dass keine nachträgliche Kristallisation auftrat. Die Ergebnisse sind in Tabelle 5-6 dargestellt.

Zehn der 12 Extrakte waren nicht analysierbar, da sie entweder Trübungen oder Niederschläge aufwiesen und teilweise zusätzlich eine feste oder ölige zweite Phase oberhalb des polaren Extraktes bildeten. Auch hier wurden die Niederschläge sowie die festen oder flüssigen Überstände in Chloroform gelöst und mit der Methode NTNP-pos analysiert, wobei ebenfalls ein dominanter Anteil an Triacylglycerolen detektiert werden konnte. Daher scheint der hohe Lipidgehalt der Haselnüsse auch bei der polaren Extraktion problematisch zu sein. Einzig eine Kombination aus H<sub>2</sub>O/MeOH (1/2; v/v) lieferte klare, stabile Extrakte ohne störende Lipidschicht als Überstand und wurde daher als geeignetes Extraktionsmittel festgelegt. Dabei waren sowohl die Extrakte mit als auch ohne Membranfiltration klar, sodass aufgrund der bereits diskutierten Aspekte (vgl. Abschnitt 5.1.1.1.4) auf eine Membranfiltration verzichtet wurde.

Lösı	ungsmittel	[µL]	Membran-	Ergebnis direkt nach Extraktion		Ergebnis nach 14	
H <sub>2</sub> O	МеОН	ACN	filtration	Überstand über Lösung	Lösung	Tagen	
500	500			schleimig/ölig	trübe	trübe	×
500	500			schleimig/ölig	trübe	trübe	×
333	667			-	klar	klar	✓
333	667		+	-	klar	klar	$\checkmark$
667	333		+	fest/weiß	trübe	trübe	x
667	333		+	fest/weiß	trübe	trübe	×
500		500		ölig	klar mit Partikeln	klar mit Partikeln	×
500		500		ölig	trübe	trübe	x
333		667		fest/weiß	klar	trübe	×
333		667	+	fest/weiß	klar	trübe	x
667		333	+	-	trübe	trübe	×
667		333	+	-	trübe	trübe	×

Tabelle 5-6: Extraktionsbedingungen und deren Auswirkung auf die Stabilität und Nutzbarkeit der polaren Extrakte.

### 5.1.1.2.3 Extraktionsmethode

Die Extraktionsmethode ExV26 wurde ausgehend von den Ergebnissen der unpolaren Extraktion auch für die Extraktion von polaren Metaboliten übernommen. Allerdings stellen Proteine, die im Zuge der Extraktion mit isoliert werden können, ein Problem bei der Analyse von komplexen Lebensmittelproben dar. Bei Aufrechterhaltung der enzymatischen Aktivität einiger Proteine in den Extrakten werden unerwünschte chemische Reaktionen katalysiert und dadurch das Ergebnis der Untersuchung verfälscht. Darüber hinaus stören Proteine die chromatographische und massenspektrometrische Detektion durch Verstopfung von Säulen und Kapillaren oder durch Verursachung von Ionensuppressionseffekten.<sup>196, 211</sup> Aus diesem Grund müssen Proteine vor der Analyse abgetrennt werden. Im einfachsten Fall erfolgt dies (wie z.B. bei der Methode für unpolare Metabolite) durch irreversible Denaturierung durch Zugabe von organischen Lösungsmitteln. Da das verwendete Extraktionsmittel einen hohen Wasseranteil von 33 % enthält, erfolgte zur Proteinpräzipitation zunächst nur die Zugabe von 667 µL Methanol zur Probe. Erst nach 1 min Homogenisieren mittels Kugelmühle bei 3 m/s wurden 333 µL Wasser hinzugegeben und danach erneut für 2 min bei 3 m/s homogenisiert.

# 5.1.2 Targeted LC-MS

# 5.1.2.1 Extraktion

Da es sich bei den identifizierten Markersubstanzen (siehe Tabelle 5-11, Abschnitt 5.2.2.1) ausschließlich um unpolare Metabolite handelte, war die entwickelte Extraktionsvorschrift zur non-targeted Analyse des unpolaren Metaboloms (ExV26) auch für die targeted Anwendung als geeignet anzusehen. Die Überprüfung der Eignung der entwickelten Extraktionsmethode für die 20 Markersubstanzen durch nachträgliche Neu-Auswertung der non-targeted LC-MS Messungen (siehe Abschnitt 5.1.1.1.2) bestätigte diese Hypothese bereits, sodass keine weitere Anpassung vorgenommen wurde.

### 5.1.2.2 Flüssigchromatographische Trennung

Zur Entwicklung einer HPLC-Methode zur Trennung der Marker- und internen Standardsubstanzen wurden verschiedene analytische *Reversed Phase* Trennsäulen mit unterschiedlichen Dimensionen und Eigenschaften getestet (siehe Tabelle 5-7). Eine vollständige Basislinientrennung ist bei der Verwendung des massenspektrometrischen Detektors nicht zwangsläufig notwendig, da überlappende Peaks bei verschiedenen m/z getrennt voneinander detektiert werden. Aufgrund der auftretenden Ionensuppressionseffekte bei größeren Mengen an koeluierenden Substanzen ist eine grundlegende Auftrennung der komplexen Lebensmittelprobe dennoch notwendig. Der Fokus der Methodenentwicklung lag unter Berücksichtigung einer möglichen Applikabilität in Routineanwendungen maßgeblich auf einer möglichst kurzen Analysenzeit bei hoher Reproduzierbarkeit und gleichzeitig hinreichender Trennung.

Säule	Länge [mm]	Innendurch- messer [mm]	Partikelgröße [µm]	Partikelart
Agilent Zorbax Eclipse XDB-C18	150	4,6	5	vollporös
Agilent Zorbax SB-C18	150	4,6	5	vollporös
Agilent Poroshell 120 EC-C18	50	4,6	2,7	teilporös
Hamilton PRP-C18	150	2,1	5	vollporös
Macherey-Nagel Nucleodur 100-3 CN-RP	125	2,0	3	vollporös
Macherey-Nagel Nucleoshell RP 18	100	2,0	2,7	teilporös
Macherey-Nagel Nucleodur C18 HTec	125	2,0	5	vollporös

Tabelle 5-7: Spezifikationen der getesteten HPLC-Säulen.

Die Agilent Poroshell 120 EC-C18 erfüllte diese Kriterien am besten. Der teilporöse Charakter der stationären Phase (Core-Shell Technologie) führt zusätzlich zur Verbesserung der Trennleistung bei gleichzeitiger Verringerung des Rückdrucks.<sup>107</sup> Letzteres ermöglicht die

Parameter	Optionen
Eluent B	IPA, IPA/ACN 3/1, IPA/MeOH 3/1 (v/v)
Additiv zu Elutionsmitteln	5 mM AF, 10 mM AF, 15 mM AF, 20 mM AF
Gradient	16 verschiedene (siehe Tabelle 8-7, Abschnitt 8.2)
Fließgegehwindigtreit	350 μL/min, 400 μL/min, 450 μL/min, 500 μL/min, 550 μL/min,
Filebgeschwindigken	600 µL/min
Säulenofentemperatur	30 °C, 35 °C, 44 °C

Tabelle 5-8: Parameter zur Optimierung des flüssigchromatographischen Systems der targeted Methode.

Verwendung kleinerer Partikelgrößen, die ebenfalls zur Erhöhung der Trennleistung beitragen. Aus diesem Grund war eine ausreichende Trennung der Analyten trotz der verhältnismäßig geringen Länge der Säule (50 mm) möglich. Gleichzeitig konnten die verbleibenden Parameter (Eluenten, Additive, Gradient, Fließgeschwindigkeit und Säulenofentemperatur; siehe Tabelle 5-8) derart modifiziert werden, dass die Analysendauer unter Berücksichtigung des maximalen Rückdrucks (600 bar) auf 13,5 min verkürzt werden konnte. Durch diese in Hinsicht auf die Komplexität der Matrix extrem kurze Analysendauer werden Messungen mit hohem Probendurchsatz ermöglicht und gleichzeitig die laufenden Kosten z.B. für Lösungsmittel verringert. Ein exemplarisches Chromatogramm der optimierten Methode ist in Abbildung 5-5 dargestellt.



Abbildung 5-5: Beispielchromatogramm der Messung eines repräsentativen Extraktes mit den Signalen der 20 Markersubstanzen und fünf internen Standards unter Verwendung der optimierten targeted LC-MS Methode.

#### 5.1.2.3 Massenspektrometrische Detektion

Für die Entwicklung einer sensitiven und selektiven Methode zur Quantifizierung der Markersubstanzen wurden die ionenspezifischen, idealen massenspektrometrischen Einstellungen mit Hilfe der "Automatic Compound Optimization" Funktion für jede Substanz bestimmt. Zur Detektion wurde das Multiple Reaction Monitoring (MRM) eingesetzt, bei dem Precursor-Ionen zunächst massenselektiv im ersten Quadrupol isoliert werden und die im zweiten Quadrupol erzeugten Fragmentionen anschließend im dritten Quadrupol detektiert werden. Durch dieses Verfahren werden sowohl die Selektivität als auch Sensitivität drastisch erhöht.<sup>133, 134</sup> Für jede zu analysierende Substanz wurden mit Hilfe von Standardsubstanzen spezifische Fragmentierungsmuster identifiziert und die elektrischen Spannungen der Ionenoptiken und Quadrupole für bestimmte Massenübergänge optimiert. Bei den Metaboliten, für die keine Standards erhältlich waren, wurden auf Grundlage strukturell sehr ähnlicher Stoffe wahrscheinliche Fragmente sowie dazugehörige geeignete Spannungs-Werte extrapoliert. Aus einer Vielzahl unterschiedlicher Einstellungen wurden anschließend nach wiederholter massenspektrometrischer Analyse eines repräsentativen Haselnussextraktes die idealen Werte bestimmt. So konnte auch für diese Metabolite eine selektive und sensitive Detektion gewährleistet werden. Um eine ausreichende Empfindlichkeit zu erzielen sowie gleichzeitig den jeweiligen Analyten ein Minimum an 10 Datenpunkten einzuräumen, wurde die MRM-Methode in zwei Perioden aufgeteilt. Dadurch besteht die Möglichkeit die Dwell-Time für jeden Massenübergang zu erhöhen, wodurch die Sensitivität (Signal-zu-Rausch-Verhältnis) verbessert werden kann.<sup>134</sup>

### 5.1.2.4 Validierung

Zur Überprüfung und zum Nachweis der Eignung und Zuverlässigkeit der targeted HPLC-ESI-QqQ-MS/MS Methode erfolgte eine Methodenvalidierung gemäß anerkannter Richtlinien der FDA und der DIN 32645 anhand der kommerziell erhältlichen Markersubstanzen sowohl in Abwesenheit (Grundkalibrierung) als auch in Anwesenheit (Matrixkalibrierung) einer Haselnussmatrix.<sup>212, 213</sup> Da für die Triacylglycerole keine Standardsubstanzen erhältlich waren, wurde die Validierung zur Demonstration der prinzipiellen Anwendbarkeit anhand zwei strukturell vergleichbarer Triacylglycerole (TG(16:0/18:1/18:1) und TG(16:0/18:1/18:2)) durchgeführt. Die Ergebnisse der Methodenvalidierung für die Grundund Matrixkalibrierung sind in Abschnitt 8.2 Tabelle 8-8 und Tabelle 8-9 zusammengefasst. Im Rahmen der Grundkalibrierung zeigte sich, dass die entwickelte Methode für sämtliche untersuchten Metabolite empfindlich, robust und über einen breiten linearen Arbeitsbereich anwendbar war. Für die meisten Metabolite erstreckte sich der lineare Bereich über mehrere Dekaden mit Grenzen von 0,04 bis 92,08 µM. Gleiches galt auch für die internen Standards. Die erzielte Linearität aller Substanzen im ermittelten Arbeitsbereich ist durchweg als sehr gut einzustufen ( $R^2 > 0.99$ ). Die Nachweisgrenzen der Markersubstanzen befanden sich im unteren nanomolaren Bereich mit Werten bis zu 1,36 nM. Einzig für y-Tocopherol und die Triacylglycerole lagen die Nachweisgrenzen mit 81 - 978 nM deutlich höher. Für die internen Standards hingegen wurden noch niedrigere Werte im oberen pikomolaren Bereich erhalten (190 - 410 pM), was vermutlich auf ihre spezielle Struktur zurückzuführen ist. Durch die Deuterierung und die natürlich nur selten vorkommende, ungeradzahlige C15:0 Fettsäure ist eine Überlagerung der Signale mit natürlichen Störsignalen (z.B. durch Fettrückstände auf Materialien) unwahrscheinlich, wodurch sich das Signal-zu-Rausch-Verhältnis verbessert und damit die Sensitivität steigt. Beim y-Tocopherol hingegen wurden bereits im Rahmen der Methodenentwicklung schlechte Ionisierungs- und Fragmentierungseigenschaften festgestellt, die sich in einer verminderten Sensitivität für die Massenübergänge während des MRM widerspiegelten. Die verhältnismäßig schlechten Nachweisgrenzen der beiden Triacylglycerole sind insbesondere durch hohe Blindwerte zu erklären, die im Zuge der Leerwertmethode für die Berechnung erhalten wurden. Im Gegensatz zu den anderen Metaboliten waren bei diesen Stoffen stets ausgeprägte Signale detektierbar, die durch Verunreinigungen der Lösungsmittel verursacht werden könnten oder aber durch Carry-Over Effekte durch starke Adsorption z.B. an der Injektions-Nadel des Autosamplers. Für Letzteres spricht, dass die Intensität der Signale bei wiederholter Messung der Blindwerte stetig sank. Nach gründlicher Reinigung aller Gerätschaften konnte der Blindwert auf ein Minimum reduziert werden. Allerdings resultierten bei erneuter Injektion eines Haselnussextraktes erneut erhöhte Blindwerte. Trotz eines zusätzlich eingefügten Wasch-Schrittes der Injektionsnadel vor jeder Injektion mit Chloroform/Isopropanol (1/2, v/v) konnte die Adsorption - insbesondere der Triacylglycerole – zwar verringert, jedoch nicht komplett verhindert werden.

Als weitere Parameter wurden gemäß den Richtlinien für bioanalytische Methodenvalidierung der FDA die Richtigkeit (engl.: *accuracy*) und die Präzision (engl.: *precision*) der Kalibrierfunktionen berechnet. Die Werte sollten jeweils 15 % nicht überschreiten, außer beim *Lower Limit of Quantification* (LLOQ;  $\triangleq$  Bestimmungsgrenze), für den 20 % nicht überschritten werden sollten. Die beiden Parameter wurden jeweils beim LLOQ und dem erwarteten Konzentrationsbereich (Mitte des LA und obere Grenze des LA) berechnet (siehe Abschnitt 8.2 Tabelle 8-10). Die Werte für Richtigkeit und Präzision im erwarteten Konzentrationsbereich waren bei 94 % der Signale unterhalb der empfohlenen 15 %, während bei keinem der Substanzen 20 % überstiegen wurden. Beim LLOQ, für den höchstens 20 % empfohlen werden, sind ebenfalls alle Werte unterhalb unter dieser Grenze und 82 % der Werte sogar noch < 15 %.

Im Rahmen der Matrixkalibrierung wiesen 71 % der Analyten eine sehr gute Linearität  $(R^2 > 0.99)$  auf, während 12 % mit  $0.98 < R^2 < 0.99$  noch immer akzeptable Werte erzielten. Eine beispielhafte Darstellung der resultierenden Regressionsgeraden der Grund- bzw. Matrixkalibrierung für PC(16:0/18:2) ist in Abbildung 5-6 dargestellt. Für die beiden Triacylglycerole konnten im verwendeten Konzentrationsbereich keine sinnvolle, lineare Regression erstellt werden. Bei den TGs sind stärkere Matrixeffekte durch den hohen natürlichen Triacylglycerol-Gehalt in Haselnüssen zu erwarten, wodurch das Signal-zu-Rausch-Verhältnis durch sich überlagernde Isotopensignale verschlechtert wird und gleichzeitig bereits der Sättigungsbereich des Detektors erreicht worden sein könnte.

Darüber hinaus ist durch die verhältnismäßig kurze chromatographische Trennzeit mit stärkeren Ionensuppressionseffekten durch Koelution diverser Metabolite zu rechnen. Ein genereller, störender Einfluss der Matrix im chromatographischen Zeitfenster der Triacylglycerole kann dadurch ausgeschlossen werden, dass der interne Standard der TG-Stoffklasse auch unter der Anwesenheit der unverdünnten Matrix eine sehr gute Linearität. eine niedrige Nachweisgrenze und sehr gute Richtigkeits- und Präzisions-Werte unter 15 % zeigte, obwohl hohe Konzentrationen an koeluierenden TGs zu erwarten sind. Die Nachweisgrenzen der Metabolite waren nur leicht erhöht und die Werte der Richtigkeit und Präzision (ohne die beiden TGs) waren nahezu unverändert im Vergleich zur Grundkalibrierung (kein Wert unter 20 %; 95 % unter 15 % im erwarteten Konzentrationsbereich



Abbildung 5-6: Beispielhafte Regressionsgeraden von PC(16:0/18:2) (A) der Grundkalibrierung und (B) der Matrixkalibrierung.

und 80 % unter 15 % beim LLOQ), was einen äußerst geringen Einfluss der Haselnussmatrix auf die unverdünnt messbaren Metabolite demonstriert. Die entwickelte Methode ist dementsprechend sowohl mit als auch ohne Zusatz der Matrix als reproduzierbar, empfindlich und präzise für die Quantifizierung der verwendeten Markersubstanzen zu bewerten.

Für die verbleibenden Markersubstanzen wurde eine Verdünnungsreihe einer repräsentativen Extrakt-Mischung gemessen um eine Abschätzung über den linearen Bereich treffen zu können. Darüber hinaus konnten dadurch die benötigten Verdünnungsfaktoren bestimmt werden, anhand dessen die enthaltenen Metabolite in ihrem in Haselnüssen natürlich vorkommenden Gehalt in einem linearen Bereich quantifiziert werden können. Die entsprechenden linearen Bereiche der Extrakt-Verdünnungen für alle 20 Markersubstanzen sowie Informationen zu den jeweiligen Regressionsgeraden sind in Abschnitt 8.2 Tabelle 8-11 zusammengefasst. Während für die Messung der Triacylglycerole TG(14:0/16:0/18:1), TG(16:0/16:1/18:1), TG(17:1/18:1/18:2) und TG(18:2/18:2/18:3) eine Verdünnung des Extraktes von 1:100 notwendig war, konnten die restlichen Metabolite im unverdünnten Extrakt quantifiziert werden. Einige der Metabolite konnten sowohl im unverdünnten als auch im 1:100 verdünnten Extrakt gemessen werden. Aufgrund der sich erhöhenden Messunsicherheit bei niedrigeren Konzentrationen bzw. kleineren Signalintensitäten (= steigender CV) wurde die Auswertung für diese Metabolite im unverdünnten Zustand durchgeführt. Die Ergebnisse der Extrakt-Verdünnungsreihe lassen vermuten, dass die fehlende Linearität für die beiden TGs im Zuge der Methodenvalidierung darauf zurückzuführen ist, dass die in der Matrix natürlich enthaltenen TGs in zu hoher Konzentration vertreten sind, sodass im unverdünnten Extrakt bereits der Sättigungsbereich erreicht wurde. Nach der Analyse des 1:100 verdünnten Extraktes enthält der unverdünnte Extrakt rechnerisch ca. 8 µM TG(14:0/16:0/18:1), 223 µM TG(16:0/16:1/18:1), 50 µM TG(17:1/18:1/18:2) und 67 µM TG(18:2/18:2/18:3). Ausgehend vom linearen Arbeitsbereich der beiden TGs in der Grundkalibrierung endet dieser jedoch bei 20,06 bzw. 21,21 µM. Drei der vier Metabolite würden diesen also bereits überschreiten. Zwar ist der lineare Arbeitsbereich substanzspezifisch und kann nicht direkt verglichen werden, dennoch indiziert dieser Vergleich, dass der Sättigungsbereich im unverdünnten Extrakt höchstwahrscheinlich bereits erreicht wurde.

Im Zuge der Analyse der Extrakt-Verdünnungsreihe stellte sich zusätzlich heraus, dass der Gehalt an  $\gamma$ -Tocopherol selbst im unverdünnten Extrakt unterhalb der Bestimmungsgrenze (LLOQ) lag. Auch in den darauffolgenden Analysen im Zuge der Probenmessungen (siehe Abschnitt 5.3.2) lagen die Signalintensitäten entweder unterhalb der Nachweisgrenze oder

unterhalb des LLOQ, weshalb eine Quantifizierung von y-Tocopherol anhand der vorliegenden Methode nicht möglich war. Ausgehend von der Nachweisgrenze (0,08 µM) und dem LLOQ (15,00 µM) sowie unter Berücksichtigung der zugrundeliegenden Extraktion und LC-MS Methode ist ein  $\gamma$ -Tocopherol-Gehalt von 0,69 mg/kg theoretisch nachweisbar bzw. 201,61 mg/kg quantifizierbar. Dabei wurden in zahlreichen wissenschaftlichen Studien y-Tocopherol-Gehalte von 2 - 240 mg/kg extrahiertem Haselnussöl - abhängig von Sorte, Herkunft und Analysenmethode - bestimmt, oder es war nicht detektierbar.<sup>24, 214-218</sup> Andere Studien errechneten Gehalte zwischen 0,0038 und 20,8 mg/kg Haselnuss oder es war ebenfalls nicht nachweisbar.<sup>219-221</sup> Demnach wäre eine Anpassung der Extraktionsmethode (z.B. durch Veränderung des Extraktionsmittels) zur Steigerung der Extraktionseffizienz von γ-Tocopherol zwar eine Möglichkeit um die Ausbeute zu steigern, allerdings läge Konzentration selbst bei quantitativer Extraktion aufgrund des niedrigen natürlichen Gehaltes wahrscheinlich nicht oberhalb des LLOQ. Aufgrund der Polarität des Tocopherols müsste das derzeitige Extraktionsmittel dahingehend angepasst werden, dass ein höherer Isopropanol-Anteil oder zusätzlich ein Methanol-Anteil verwendet würde, wodurch wiederum die Extraktion der sehr unpolaren DGs und TGs verringert werden würde. Da aber DG(16:0/16:1) und TG(15:0/16:0/18:1) bereits am unteren Limit des linearen Bereichs lagen, ist eine derartige Änderung der Extraktionsmethode nicht sinnvoll. Eine Alternative wäre eine zusätzliche, separate Extraktion und Messung, speziell zur Quantifizierung des  $\gamma$ -Tocopherols oder die Einführung eines Anreicherungsschrittes während der Extraktion. Weil diese beiden Möglichkeiten jedoch mit zusätzlichem materiellen und zeitlichen Aufwand verbunden sind, welcher im Kontext von high-throughput Analysen nicht erwünscht ist, und sie eventuell zusätzlich die Reproduzierbarkeit verringern würden, blieb die Extraktionsmethode unverändert und das y-Tocopherol wurde für die folgenden targeted LC-MS Analysen nicht weiter betrachtet.

Abschließend wurden für alle Markersubstanzen Korrekturfaktoren berechnet, welche das substanzspezifische Detektionsvermögen der unterschiedlichen Metabolite gegenüber dem jeweiligen zur Quantifizierung herangezogenen internen Standard der entsprechenden Stoffklasse berücksichtigen. Da diese Berechnung nur für die Metabolite mit vorhandenen Standardsubstanzen erfolgen konnte, wurden die Korrekturfaktoren für die restlichen Metabolite anhand der jeweils korrespondierenden Stoffklassen extrapoliert. Dadurch wurde sichergestellt, dass die berechneten Konzentrationen möglichst exakt dem wahren Werten entsprechen. Dennoch würde eine Berechnung des Korrekturfaktors durch Standardsubstan-

zen zu noch genaueren Ergebnis führen, sodass dies nach Möglichkeit noch nachgeholt werden sollte. Entsprechende Ergebnisse dieser Arbeit ließen sich dann leicht mit Hilfe des tatsächlichen Korrekturfaktors nachbessern.

Zusammenfassend lässt sich rückschließen, dass eine Analysenmethode entwickelt werden konnte, die eine robuste Quantifizierung von 19 Markersubstanzen bei einer ausgesprochen kurzen Probenvorbereitungs- und Analysenzeit ermöglicht. Die reine Gerätezeit beträgt nur 25:50 min (2,3 min Homogenisierung, 10 min Zentrifugation und 13,5 min LC-MS-Messung). Die Gesamt-Analysenzeit beläuft sich somit unter Berücksichtigung aller notwendigen Arbeitsschritte (Einwiegen, Filtrieren, Abfüllen, etc.) insgesamt auf ca. 30 - 40 min je Probe. Die Applikation ist daher sowohl in Hinsicht auf die benötigte apparative Ausstattung als auch die benötigte Analysenzeit für eine Implementierung in die Routineanalytik geeignet.

# 5.2 Geographische Herkunftsdiskriminierung mittels non-targeted LC-MS

# 5.2.1 Vergleich der Methoden NTNP-pos, NTNP-neg, NTP-pos und NTP-neg

Zur Überprüfung der Eignung der vier verschiedenen LC-MS Methoden für die Diskriminierung der geographischen Herkunft von Haselnüssen wurden zunächst 86 authentische Haselnussproben aus dem Erntejahr 2014 (60 Frankreich, 12 Deutschland, 5 Italien, 5 Georgien und 4 Türkei) mit den Methoden NTNP-pos, NTNP-neg, NTP-pos und NTP-neg (siehe Abschnitt 6.3.1 bzw. 6.3.2) analysiert. Authentisch bedeutete in diesem Zusammenhang, dass die geographische Herkunft der Probe zweifelsfrei durch Projektpartner garantiert werden konnte oder die Haselnüsse direkt von Bauern bezogen wurden, während es sich bei den nicht-authentischen Proben überwiegend um Handelsware für die Süßwarenindustrie oder aus dem Einzelhandel handelte, bei denen die deklarierte Herkunft nicht eindeutig nachverfolgt werden konnte.

Wegen eventuell auftretender instrumenteller Drifts erfolgte die Messung der Proben eines Batchs jeweils in randomisierter Reihenfolge, um sicherzustellen, dass Unterschiede zwischen Probengruppen nicht durch den Messablauf bedingt werden. Zur Equilibrierung des massenspektrometrischen Systems wurden jeweils 5 Proben zur Qualitätskontrolle (engl. *quality control*; QC) vor Beginn der eigentlichen Messung injiziert, welche im Anschluss nicht für die Auswertung verwendet wurden. Ohne diese Prozedur können insbesondere bei QTOF-Geräten Verschiebungen der Retentionszeit, Ionenmasse und/oder Signalintensität der ersten Proben zu Störungen der Auswertung führen.<sup>222</sup> Zur Überprüfung der Stabilität der Messung erfolgte zusätzlich die Injektion eines Blindwertes gefolgt vom QC nach jeder 10. Injektion. *Time Alignment*, Normalisierung und Peak-Gruppierung (engl.: *peak grouping*) wurden für jeden Batch separat durchgeführt. Die daraus resultierenden Bucket Tables wurden zu einer Datenmatrix vereint und anschließend mittels Centering und Scaling weiter prozessiert. Die Parameter des Feature Detection Algorithmus wurden dabei eher großzügig gewählt um falsch-negative Ergebnisse durch nicht erfolgte Detektion von weniger abundanten Metaboliten zu vermeiden. In der Annahme, dass sich das Metabolom von Haselnüssen aus unterschiedlichen Herkunftsländern nicht signifikant in der Anwesenheit/Abwesenheit von einzelnen Metaboliten unterscheidet, sondern vielmehr aufgrund von Konzentrationsunterschieden, wurden die Parameter der Peak-Gruppierung im Gegensatz dazu zur Vermeidung von falsch-positiven Ergebnissen sehr streng gesetzt. Diese könnten leicht durch Signale des Grundrauschens verursacht werden, welche aufgrund der statistischen Häufung anschließend zufällig in mehreren Proben anzutreffen sind. Durch das Testen von weicheren Einstellungen konnte diese Hypothese bestätigt werden, da sich keine Features fanden, die nur bzw. nicht in bestimmten Ländern vorhanden waren. Da die einzustellenden Parameter der Peak-Gruppierung einen erheblichen Einfluss auf die Resultate der statistischen Auswertung ausüben, müssen diese mit besonderem Bedacht gewählt werden. Andernfalls können leicht Fehlinterpretationen der Datensätze resultieren. Darüber hinaus ist zu prüfen, ob es sich bei relevanten Features um echte, diskrete und reproduzierbare Signale von Metaboliten handelt (inkl. Isotopenmuster) oder lediglich um Störsignale/Grundrauschen. Die aus den Datensätzen resultierenden PCA-Modelle der vier LC-MS Methoden sind in Abbildung 5-7 dargestellt.

Im Allgemeinen führten die Ergebnisse der unpolaren Messungen zu besseren Resultaten hinsichtlich der Herkunftsunterscheidung als die der polaren Messungen, wobei der positive Ionenmodus in beiden Fällen dem negativen überlegen war. Das Gleiche traf auch sowohl auf die Anzahl der detektierten Metabolite als auch die Robustheit zu. Während die unpolaren Methoden ein enges *Clustering* aller QC Proben über die gesamte Messdauer aufwiesen, mussten bei der Methode NTP-pos vor Beginn der eigentlichen Messung zusätzlich 10 QC-Proben zur Equilibrierung des Systems injiziert werden. Bei der Methode NTP-neg hingegen wurde über die gesamte Dauer der Messreihe ein gerichteter Drift innerhalb der QCs festgestellt, weshalb die Ergebnisse dieser Messung nur bedingt interpretierbar sind.



Abbildung 5-7: PCA *Scores Plots* der 86 authentischen Proben aus 2014 von (A) NTNP-pos, PC1 vs. PC3; (B) NTNP-neg, PC1 vs. PC2; (C) NTP-pos, PC1 vs. PC2; und (D) NTP-neg, PC1 vs. PC2; unter Verwendung jeweils aller detektierten *Features*. (modifiziert nach Klockmann 2016)

Zur Bewertung der Reproduzierbarkeit wurde der Variationskoeffizient für jeden Metaboliten in allen QC-Proben berechnet. Nach Maßgabe der allgemein anerkannten Richtlinien der FDA für bioanalytische Methodenvalidierungen sollte die Genauigkeit von Messungen, ausgedrückt als Variationskoeffizient (CV), 15 % nicht übersteigen.<sup>212</sup> Für die unpolaren Methoden erfüllten mehr als 95 % der Metabolite diese Anforderungen, während dies nur für 24 % bzw. 31 % der Metabolite in den polaren Methoden der Fall war. Nichtsdestotrotz fanden sich in jeder der Methoden potentielle Markersubstanzen für eine Herkunftsdiskriminierung in steigender Anzahl von NTP-neg (1) über NTP-pos (3) und NTNP-neg (39) zu NTNP-pos (43). Die Validierungsparameter für die vier Methoden sind in Tabelle 5-9 zusammengefasst. In diesem Zusammenhang wurde ein Metabolit als potentielle Markersubstanz definiert, wenn alle der folgenden drei Kriterien erfüllt wurden:

- p-value der ANOVA zum Vergleich von mehr als zwei Mittelwerten unter Verwendung aller Proben < 0,01</li>
- Anzahl der statistisch signifikanten Unterschiede (p-value des T-Tests zum Vergleich zweier Mittelwerte < 0,01) zwischen den bis zu 10 möglichen Länderpaarungen > 5
- CV < 15%

I C MS Mothodo	Anzahl detektierter	mittlerer CV	Anteil	potentielle
LC-MB Methode	Features	[%]	CV < 15 % [%]	Markersubstanzen
NTNP-pos	137	5,57	99	43
NTNP-neg	50	7,04	98	39
NTP-pos	149	23,76	24	3
NTP-neg	25	21,53	31	1

Tabelle 5-9: Validierungsparameter für alle vier non-targeted LC-MS Methoden. (modifiziert nach Klockmann 2016)

CV: Variationskoeffizient

Dabei wurden zur Vermeidung von falsch-positiven Ergebnissen ausschließlich p-values verwendet, die gleichzeitig *family-wise error rates* (FWER) und *false discovery rates* (FDR) unter 0,01 besaßen.<sup>223, 224</sup> Während die ANOVA Berechnungen die Gesamtvarianz und somit die Fähigkeit eines Metabolits zur Diskriminierung bei der Betrachtung der Gesamtprobenpopulation beschreibt, werden durch die Berücksichtigung der Anzahl der statistisch signifikanten Unterschiede zwischen den Länderpaarungen mittel T-Test darüber hinaus nur Metabolite



Abbildung 5-8: PCA Scores Plot der 86 authentischen Proben aus 2014 von (A) NTP-pos; und (B) NTNP-pos; jeweils unter Verwendung aller Features; farbliche Markierung der Sorten. (modifiziert nach Klockmann 2016)

als Markersubstanzen ausgewählt, die zwischen mehreren Ländern gute diskriminative Eigenschaften aufweisen. Dadurch wird gewährleistet, dass nur Metabolite ausgewählt werden, die nicht nur ein bestimmtes Land besonders gut charakterisieren können, sondern sich möglichst für die Unterscheidung des gesamten Länderspektrums eignen.

Bei der Betrachtung von Proben, bei denen von einer Sorte mehr als eine Probe vorlag, zeigte sich nur in den polaren Messungen ein geringer Einfluss der Sorte (siehe Abbildung 5-8). Dort lagen Proben derselben Sorte teilweise leicht gehäuft im *Scores Plot* der PCA vor, aber dennoch überlappend mit den meisten anderen Proben. Im Gegensatz dazu waren in beiden unpolaren Methoden bei zufälliger Verteilung der Proben keine Sorten-Cluster zu erkennen.

Um den Sorteneinfluss in den unpolaren Methoden weitergehend zu evaluieren, wurde ergänzend eine PLS-DA berechnet. Dabei handelt es sich um eine *supervised* Methode, deren Ziel die Verbesserung der Unterscheidung von Probengruppen durch Maximierung der Kovarianz zwischen den Variablen (hier Signalintensitäten der Metabolite) und den zuvor definierten Eigenschaften (hier Herkunftsländer) ist.<sup>35</sup> Demnach würden sortencharakteristische Informationen im Datensatz durch diese multivariate Methode gezielt herausgearbeitet werden, falls vorhanden. Die Tatsache, dass die Ergebnisse der PLS-DA im Vergleich zur PCA jedoch keine großen Unterschiede aufwies, deutet darauf hin, dass die meiste Varianz im Datensatz des unpolaren Metaboloms durch geographische Herkunft resultiert und wahrscheinlich nur in geringem Umfang auf die Sorte zurückzuführen ist.

Aus den genannten Gründen wurde NTNP-pos als am besten geeignete Methode zur Herkunftsunterscheidung von Haselnüssen bestimmt und für alle folgenden Analysen verwendet. Sie besaß die besten Klassifizierungsresultate bei der höchsten Anzahl an potentiellen Markersubstanzen und größter Reproduzierbarkeit und wurde zusätzlich nicht



Abbildung 5-9: PCA *Scores Plot* der 86 authentischen Proben aus 2014 als Kombination von NTNP-pos, NTNP-neg, NTP-pos und NTP-neg; die Distanzen zwischen den Ländergruppen sind deutlich größer als bei NTNP-pos allein.

oder nur in geringem Maße durch die beeinflusst Sorte wird Eine Kombination aus mehreren Methoden zur Verbesserung der diskriminativen Fähigkeiten wäre zwar denkbar und führt nachweislich zu besseren Resultaten (siehe Abbildung 5-9), ist aber gleichzeitig extrem zeitaufwendig und unpraktikabel. Bei sehr komplexen Diskriminierungsproblemen könnte diese Vorgehensweise jedoch zielführend sein. Im vorliegenden Fall erscheint der zusätzliche Nutzen unverhältnismäßig klein im Vergleich

zum finanziellen, materiellen und zeitlichen Mehraufwand. Darüber hinaus überschneiden sich einige der potentiellen Markersubstanzen durch multiple Detektion, da sie entweder sowohl im positiven als auch negativen Ionenmodus messbar sind und/oder sowohl mit der unpolaren als auch polaren Methode erfasst werden. Ein Beispiel hierfür ist PC(18:2/18:2), welches aufgrund seiner chemischen Struktur (Phosphocholin-Rest als polare, zwitterionische Kopfgruppe) mit allen vier Methoden erfasst werden konnte.

# 5.2.2 Analyse der authentischen Haselnussproben von 2014 und 2015 unter Anwendung von NTNP-pos

Nach der Bestimmung der optimalen LC-MS-Methode zur Diskriminierung der geographischen Herkunft von Haselnüssen wurden sämtliche Haselnussproben der Erntejahre 2014 und 2015 (207 authentisch und 44 nicht-authentisch; siehe Tabelle 5-10) erneut mit der Methode NTNP-pos analysiert. Aufgrund der großen Probenanzahl bei einer Analysenzeit von 26 min pro Probe konnte aus Zeitgründen nur eine Einfachbestimmung durchgeführt werden. Diese Vorgehensweise ist allerdings gängige Praxis in non-targeted basierten Metabolomics-Studien mittels UPLC-QTOF-MS Geräten aufgrund des großen finanziellen, apparativen und zeitlichen Aufwands.<sup>102, 151, 209, 259</sup> Die Proben aus 2014 und 2015 wurden dabei aus zeitlichen Gründen in separaten Batches gemessen. Auch hier erfolgte aus bereits erläuterten Gründen

(vgl. Abschnitt 5.2.1) die Akquisition jeweils in randomisierter Reihenfolge unter wiederholter Messung eines QCs und eines Blindwertes nach allen 10 Injektionen und initialer Messung von 5 QCs ohne diese für die spätere Auswertung zu verwenden. Die Prozessierung erfolgte analog zu Abschnitt 5.2.1 mit dem einzigen Unterschied, dass das *Scaling* und *Centering* der beiden Batches jeweils separat durchgeführt werden musste. Wurden *Scaling* und *Centering* der beiden Datensätze gemeinsam durchgeführt, so zeigte sich in der resultierenden PCA eine deutliche Differenz zwischen Proben aus 2014 und 2015,

Tabelle5-10:Anzahlderauthentischenundnicht-authentischenHaselnussprobenfür die non-targetedMessungen.

	authentisch	nicht- authentisch
Aserbaidschan*	3	2
Deutschland	35	-
Frankreich	115	-
Georgien	10	6
Italien	22	9
Kanada*	1	-
Spanien*	6	1
Türkei	14	25
$\mathrm{USA}^*$	1	1

\*aufgrund zu geringer Probenanzahl nicht für die statistische Auswertung verwendet

während ein derartiger Bias bei separater Durchführung nicht beobachtet werden konnte. Dies ist auf die gerätespezifischen Varianzen des empfindlichen QTOF-MS bei Inter-Batch Messungen zurückzuführen, welche insbesondere bei größeren zeitlichen Abständen die Vergleichbarkeit der Datensätze häufig limitiert.<sup>225</sup>

Für die folgende statistische Auswertung wurden allerdings nicht alle Proben verwendet. Die authentische Probe FRA017 erwies sich in Zuge der statistischen Auswertung als starker Ausreißer, da sowohl die Residual Variance als auch die Leverage der PCA über den jeweiligen kritischen Limits lagen (*critical F residuals limit* = 0,49; *critical leverage* 



Abbildung 5-10: Herkunft der authentischen Haselnussproben zur statistischen Auswertung. (modifiziert nach Klockmann 2016) <sup>1</sup>Für die Auswertung der non-targeted Messungen nicht-verwendet

*limit* = 0,2), und wurde daher nicht weiter berücksichtigt. Eine Einsicht in den entsprechenden Datensatz zeigte einen anormalen, starken Retentionszeitenshift über die gesamte Messung, wodurch die Peak-Gruppierung bei dieser Probe fehlerhaft wurde. Im Allgemeinen ist die Entfernung von Ausreißern bei Metabolomics-Daten ein kritischer Schritt, weil Proben möglicherweise nur aufgrund ihrer natürlichen Beschaffenheit zu "biologischen" Ausreißern werden, obwohl die Messung ordnungsgemäß erfolgte. Solche Proben dürfen nicht ohne weiteres ausgeschlossen werden, sondern spiegeln vielmehr die tatsächliche Variation des Metaboloms wider, welche durch einen Ausschluss verfälscht dargestellt werden würde. Die Detektion und Interpretation von Ausreißern bei Metabolomics-Messungen müssen daher stets individuell betrachtet und mit großer Sorgfalt durchgeführt werden.

Des Weiteren wurden 11 authentische Proben der Länder Spanien (6), Aserbaidschan (3), Kanada (1) und USA (1) von der statistischen Auswertung ausgeschlossen, weil die jeweilige Gesamtprobenmenge für eine valide Statistik nicht ausreichte. Damit eine Probengruppe für die multivariate Datenanalyse verwendet werden konnte, wurde ein Minimum von 10 Datensätzen als unterste Grenze festgelegt. Probenpopulationen unterhalb dieser Menge wurden als nicht ausreichend repräsentativ für entsprechende statistische Modelle und Bewertungen eingeschätzt, da biologische Variationen nicht ausreichend berücksichtigt werden können und somit ein verzerrtes oder falsches Bild des jeweiligen Landes resultieren könnte. Außerdem würden systematische Fehler bei Einfachbestimmungen – wie hier durchführt – unbemerkt bleiben und die Ergebnisse bei einer solch geringen Probenanzahl zu stark verzerren. Abbildung 5-10 zeigt die regionale Herkunft der zur Auswertung verwendeten Haselnussproben mit der jeweils entsprechenden Anzahl.

Eine nachträgliche Messung von zusätzlichen Proben bspw. für die Ergänzung von weiteren Ländern ist in diesem Fall nicht sinnvoll, da die bereits beschriebene Problematik bei Inter-Batch Messungen zu einer mangelnden Vergleichbarkeit der Daten führt. Durch die große Anzahl der gemessenen Proben aus 2014 bzw. 2015 konnte das separate *Scaling* und *Centering* der beiden Batches zu dennoch vergleichbaren Daten führen, da die Gesamtvariation innerhalb der Proben aus 2014 bzw. 2015 in etwa gleich ist und dadurch der unterschiedliche *Threshold* durch das *Scaling* eliminiert werden konnte. Bei der Messung von einzelnen oder wenigen Proben ist dies allerdings nicht möglich, sodass für derartige Fälle der komplette Batch erneut gemessen werden müsste.

Des Weiteren konnten vier nicht-authentische Proben der Länder Aserbaidschan, Spanien und USA nicht verwendet werden, da die entsprechenden authentischen Proben dieser Länder nicht für die Modellerstellung ausreichten.

Insgesamt bildeten somit 195 authentische und 40 nicht-authentische Proben die Grundlage für die anschließende statistische Auswertung. Abbildung 5-11 zeigt einen PCA *Scores Plot* 



Abbildung 5-11: PCA *Scores Plot* der 195 authentischen Haselnussproben aus 2014 und 2015, gemessen mit NTNPpos; unter Verwendung aller 225 *Features*. (modifiziert nach Klockmann 2016)

der entsprechenden authentischen Haselnussproben der fünf Herkunftsländer und demonstriert, dass kein erkennbareres Bias in Bezug auf das Erntejahr existierte. Somit schienen die Unterschiede zwischen den Erntejahren (insbesondere die klimatischen Verhältnisse) – zumindest in Bezug auf die hier detektierten 225 Features - keinen oder nur einen sehr geringen Einfluss auf die Verteilung der Metabolit-Gehalte zu haben, was als Grundlage für eine sinnvolle und reproduzierbare Herkunftsanalytik unabdingbar ist.

Die Auftrennung der Proben im *Scores Plot* war jedoch widererwarten weniger stark ausgeprägt als bei der initialen Methodenentwicklung. Insbesondere bei den georgischen, türkischen und italienischen Proben waren die Überlagerungen so groß, dass keine eindeutigen Cluster zu erkennen waren. Aber auch die Überlappung der deutschen und französischen Proben war viel stärker ausgeprägt. Aus diesem Grund und um eine gezielte targeted Analyse mit einem Triple Quadrupol-MS zu ermöglichen wurden daraufhin diejenigen Metabolite bestimmt, die sich besonders für eine Herkunftsunterscheidung eignen und somit den größten Einfluss auf die entsprechenden Modelle haben.

### 5.2.2.1 Bestimmung der Markersubstanzen

Die Auswahl von Metaboliten mit hohen diskriminativen Eigenschaften in Bezug auf die geographische Herkunft der Haselnüsse erfolgte in Anlehnung an die in Abschnitt 5.2.1 beschriebenen Kriterien. Demnach wurde ein Metabolit als Markersubstanz ausgewählt, wenn alle der folgenden Kriterien erfüllt wurden:

- p-value der ANOVA zum Vergleich von mehr als zwei Mittelwerten unter Verwendung aller Proben < 0,01 (2014 und 2015 separat) <u>und</u> Anzahl der statistisch signifikanten Unterschiede (p-value des T-Tests zum Vergleich zweier Mittelwerte < 0,01) zwischen den bis zu 10 möglichen Länderpaarungen ≥ 3 (2014 und 2015 separat); von diesen insgesamt vier möglichen Fällen mind. drei erfüllt
- CV < 15 % (2014 und 2015 gemeinsam)

Auch hier wurden nur p-values mit FWER und FDR < 0,01 akzeptiert. Die Kriterien wurden leicht modifiziert um den Anforderungen von möglichen Erntejahr-Unterschieden gerecht zu werden und Metabolite auszuschließen, die allein in einem der beiden Erntejahre Unterschiede zwischen den Ländern aufwiesen. Dadurch wird gewährleistet, dass die ausgewählten Markersubstanzen unabhängig vom Erntejahr gute diskriminative Eigenschaften besitzen und sich somit auch für folgende Ernteperioden eignen.

Darüber hinaus wurden vier Metabolite ausgewählt, die lediglich zwei der vier Fälle des ersten Kriteriums erfüllen, weil sie sich besonders gut für die Diskriminierung zwischen der Türkei, Italien und Georgien eigneten, dessen Proben durch die restlichen Markersubstanzen nur unzureichend unterschieden werden konnten. Sie erreichten zwar bei den ANOVA Berechnungen für beide Erntejahre gute p-values, konnten jedoch nur schlecht zwischen anderen Länderpaarungen (z.B. Deutschland-Frankreich oder Italien-Frankreich) unterschei-

den. Um die Gesamtheit der beteiligten Länder zu erfassen haben sie sich dennoch als notwendig für die Erstellung von geeigneten Diskriminierungsmodellen erwiesen.

Drei weitere *Features* erfüllten zwar die Kriterien der Markersubstanzen, wurden aber dennoch nicht verwendet, da sich bei genauerer Betrachtung herausstellte, dass es sich nicht wie zunächst vermutet um Palmitoyl- (m/z 313,274/R<sub>t</sub> 14,19 min), Oleoyl- (m/z 339,290/ R<sub>t</sub> 14,24 min) und Linoleoyl-Glycidylester (m/z 337,274/R<sub>t</sub> 12,99 min) handelte. Vielmehr entstanden diese Signale wahrscheinlich durch *In-Source* Fragmentierung der koeluierenden Markersubstanzen DG(16:0/18:1) (R<sub>t</sub> 14,17 min), DG(18:1/18:1) (R<sub>t</sub> 14,26 min) und DG(18:2/18:2) (R<sub>t</sub> 12,99 min) als [acyl + 74]<sup>+</sup> Fragmentionen der korrespondierenden Diacylglycerole (vgl. dazu Abschnitt 5.2.2.2). Zum einen stimmte die Signal-Intensitätsverteilung für die verschiedenen Länder mit denen der entsprechenden DGs überein, zum anderen wären die tatsächlichen Retentionszeiten für Glycidylester aufgrund der chemischen Eigenschaften viel geringer. Darüber hinaus wurde diese Art Signale häufiger detektiert, stets wenn abundante Lipide mit entsprechenden Fettsäureresten eluierten.

Insgesamt konnten somit 20 Metabolite aus den 225 detektierten *Features* mit hervorragenden Fähigkeiten zur Diskriminierung der geographischen Herkunft der Haselnüsse bestimmte werden (siehe Tabelle 5-11). Die p-values für die Erntejahre 2014 und 2015 sowie die Variationskoeffizienten der Markersubstanzen sind in Abschnitt 8.2 Tabelle 8-12 dargestellt. In Abbildung 5-12 ist ein exemplarisches LC-MS Chromatogramm eines Haselnussextraktes mit der Methode NTNP-pos dargestellt, in dem die Positionen der 20 Markersubstanzen erkennbar sind.



Abbildung 5-12: Beispielchromatogramm eines repräsentativen Haselnussextraktes mit Markierungen der Signale der 20 Markersubstanzen, gemessen mit NTNP-pos.

Matabalit	Summonformal	mittlana m/-	Ion	Rt	<b>MS/MS Fragment-Ionen</b>
wietabolit	Summentormet	mittiere m/z	1011	[min]	[m/z]
DG(16:0/16:1)	C35H66O5	589,480	+Na	13,46	313; 311; 239; 237
DG(16:0/18:1) <sup>b</sup>	C37H70O5	612,557	$+NH_4$	14,19	577; 339; 313; 265; 239
DG(18:2/18:2) <sup>b</sup>	C <sub>39</sub> H <sub>68</sub> O <sub>5</sub>	634,542	$+NH_4$	12,99	599; 337; 263
DG(18:1/18:1) <sup>b</sup>	C <sub>39</sub> H <sub>72</sub> O <sub>5</sub>	638,574	$+NH_4$	14,26	603; 339 ;265
PC(18:2/0:0) b	C <sub>26</sub> H <sub>50</sub> NO <sub>7</sub> P	520,340	+H	2,63	502; 337; 184
PC(16:0/18:3)	$C_{42}H_{78}NO_8P$	756,554	+H	11,07	573; 339; 313; 184
PC(16:0/18:2) <sup>b</sup>	$C_{42}H_{80}NO_8P$	758,571	+H	11,84	575; 337; 313; 184
PC(18:2/18:2) <sup>b</sup>	$C_{44}H_{80}NO_8P$	782,571	+H	11,18	599; 337; 184
PC(18:1/18:2)	$C_{44}H_{82}NO_8P$	784,588	+H	11,94	601; 339; 337; 184
PE(16:0/18:2) <sup>b</sup>	$C_{39}H_{74}NO_8P$	716,523	+H	12,08	575; 337; 313; 263; 239
PE(18:2/18:2) <sup>b</sup>	$C_{41}H_{74}NO_8P$	740,523	+H	11,44	599; 337; 263
PE(18:1/18:1)	$C_{41}H_{78}NO_8P$	744,554	+H	12,91	603; 339; 265
γ-Tocopherol <sup>a,b</sup>	$C_{28}H_{48}O_2$	417,372	+H	11,19	191; 151
TG(2:0/18:2/18:2)	$C_{41}H_{70}O_6$	676,551	$+NH_4$	13,41	599; 379; 263
TG(2:0/18:1/18:2)	C41H72O6	678,567	$+NH_4$	14,02	601; 381; 379; 265; 263
TC(14.0/16.0/18.1) a	C. H. O.	822 755	±NH.	17 25	577; 549; 523; 339; 313;
10(14.0/10.0/18.1)	C511196O6	822,755	+11114	17,55	285; 265; 239; 211
TG(15:0/16:0/18:1) a	CarHagOr	836 771	+NH.	17 50	577; 563; 537; 339; 313;
10(15.0/10.0/10.1)	052119806	850,771	11114	17,50	299; 265; 239; 225
TG(16·0/16·1/18·1) a	CraHagOr	848 771	+NH.	17 33	577; 575; 549; 339; 313;
10(10.0/10.1/10.1)	053119806	040,771	11114	17,55	311; 265; 239; 237
$TG(17\cdot1/18\cdot1/18\cdot2)$	CreeH100Oc	886 787	+NH4	17.12	601; 589; 587; 339; 337;
10(17.1/10.1/10.2)	000110000	000,707	11114	17,12	325; 265, 263; 251
TG(18.2/18.2/18.3)	$C_{57}H_{04}O_{4}$	894 756	+NH4	16 27	599; 597; 337; 335; 263;
10(10.2/10.2/10.5)	03/119000	074,750	11114	10,27	261

Tabelle 5-11: Identifizierte Markersubstanzen. (modifiziert nach Klockmann 2016)

R<sub>t</sub>: Retentionszeit

<sup>a</sup> nicht alle Kriterien erfüllt

<sup>b</sup> Identität durch Referenzstandards bestätigt

Bei den Markersubstanzen handelt es sich neben  $\gamma$ -Tocopherol um Vertreter der Phosphatidylcholine, Phosphatidylethanolamine, Diacylglycerole und Triacylglycerole mit unterschiedlichen Fettsäureresten (siehe Abbildung 5-13). Besonders hervorzuheben sind dabei TG(15:0/16:0/18:1) und TG(17:1/18:1/18:2), die mit der Pentadecansäure (C15:0) bzw. Heptadecensäure (17:1) ungewöhnliche, ungeradzahlige Fettsäuren als Seitenketten besitzen. Allerdings ist bereits seit einiger Zeit bekannt, dass unverzweigte, ungeradzahlige Fettsäuren nicht nur in Mikroorganismen, sondern auch in nahezu jedem Organismus - selbst im Pflanzen- und Tierreich als Minor-Komponenten - zu finden sind.<sup>226</sup> Dementsprechend ist deren natürliches Vorhandensein in Haselnüssen als plausibel anzusehen und nicht als verschiedengradige Verunreinigung der Proben mit Mikroorganismen zu verstehen, weshalb diese beiden TGs als Markersubstanzen weiter genutzt wurden. So konnten C15:0 und C17:1 bereits in verschiedenen pflanzlichen Ölen (auch Haselnussöl) detektiert werden.<sup>24, 27, 227</sup>



Abbildung 5-13: Strukturformeln der Stoffklassen der identifizierten Markersubstanzen. (modifiziert nach Klockmann (2016)

Holčapek et al. (2003) wiesen sogar TG(17:1/18:1/18:2) in Pistazien-, Mandel-, Sojabohnenund Sonnenblumenöl in Spuren nach, während sie diesen Metabolit in Haselnussöl jedoch nicht nachweisen konnten.

Eine erneute Berechnung der PCA, dieses Mal unter Verwendung der zuvor bestimmten 20 Markersubstanzen, zeigt Vergleich im zur Verwendung aller 225 Features eine deutliche Verbesserung (siehe Abbildung 5-14). Insbesondere die Auftrennung der türkischen, italienischen und georgischen Proben konnte signifikant verbessert werden, sodass die Proben dieser Länder nun fast vollständig getrennt vorlagen. Die Überlappung der deutschen und französischen Proben konnte leicht verringert werden, wobei diese beiden Länder noch immer die größte Überschneidung besaßen. Somit ist die Variablenreduktion zumindest in Hinblick auf die Erstellung von PCA-Modellen in diesem Fall als vorteilhaft anzusehen. Die erklärte Gesamtvarianz von PC1 und PC2 konnte dadurch von 37 % auf 73 % gesteigert werden, was im Zusammenhang mit der verbesserten Auftrennung





Abbildung 5-14: PCA Scores Plot der 195 authentischen Haselnussproben aus 2014 und 2015, gemessen mit NTNP-pos, (A) PC1 vs. PC2; und (B) PC1 vs. PC2 vs. PC3; unter Verwendung der 20 Markersubstanzen. (modifiziert nach Klockmann 2016)

dafür spricht, dass viele der 225 *Features* auch andere Informationen als die geographische Herkunft beinhalten (z.B. metabolischer Zustand bei Ernte, Sorte, Lagerung o.ä.). Sie vergrößern somit zwar die Varianz innerhalb der Probenpopulationen, aber nicht zugunsten einer validen geographischen Herkunftsdiskriminierung.

#### 5.2.2.2 Strukturaufklärung der Markersubstanzen

Zur Aufklärung der Identität der Markersubstanzen wurden die mittels hochauflösendem QTOF gemessenen exakten Massen und Isotopenmuster zur Vorhersage der wahrscheinlichsten Summenformeln verwendet. Die Signale entsprachen dabei nicht zwangsläufig dem Pseudomolekülion  $[M+H]^+$ , da der *Molecular Feature* Algorithmus zur Peakdetektion Signale von verschiedenen Addukten eines Metabolits einem einzigen *Feature* zuordnete und jeweils nur das intensivste Signal zur Auswertung verwendet wurde (hier entweder  $[M+H]^+$ ,  $[M+Na]^+$  oder  $M+NH_4]^+$ ). Die Elementzusammensetzung des Addukts musste dementsprechend von der vorhergesagten Summenformel subtrahiert werden um die Summenformel des Neutralmoleküls zu erhalten, welche anschließend zusammen mit den Informationen der akquirierten MS/MS-Spektren (Kollisionsenergie = 40 eV) für eine Datenbank-basierte Vorhersage der wahrscheinlichsten Strukturformeln mittels *MetFrag* verwendet wurde. Jede Strukturformel wurde zusätzlich manuell anhand des entsprechenden Fragmentspektrums auf deren Plausibilität überprüft.

Für alle Fette und Phospholipide konnten die charakteristischen  $[Acyl]^+$  und  $[Acyl + 74]^+$ Ionen identifiziert werden. Die Fragmentspektren der Triacylglycerole enthielten darüber hinaus  $[(M + NH_4) - NH_3 - FA]^+$  Ionen, während bei den Diacylglycerolen  $[(M + NH_4) - NH_3 - H_2O]^+$  Ionen gefunden werden konnten. Die Phosphatidylethanolamine zeigten  $[(M + H) - Phosphoethanolamin]^+$  Ionen, wohingegen die Fragmentspektren der Phosphatidylcholine von *m/z* 184 dominiert wurden, welches aus dem Verlust eines Phosphocholin-Ions resultiert.<sup>229, 230</sup> Anhand dieser charakteristischen Fragment-Ionen lässt sich die Identität der Fette und Phospholipide sicher bestätigen, da sich für jede Fettsäurezusammensetzung spezifische Fragment-Ionenmassen detektieren lassen. Am Beispiel von TG(15:0/16:0/18:1) ist dieses charakteristische Fragmentierungsmuster in Abbildung 5-15 exemplarisch aufgeschlüsselt und veranschaulicht. Das Prinzip lässt sich jedoch auch auf die anderen Markersubstanzen übertragen (ggfs. unter Berücksichtigung der jeweils Stoffklassenspezifischen Fragmente von DG, PE oder PC).



Abbildung 5-15: MS/MS-Spektrum von TG(15:0/16:0/18:1) bei 40 eV sowie exemplarische Erläuterungen zu aussagekräftigen, charakteristischen Fragment-Ionen; blau: C15:0-Seitenkette, rot: C18:1-Seitenkette, grün: C16:0-Seitenkette.

 $\gamma$ -Tocopherol zeigte die für  $\beta$ - und  $\gamma$ -Tocole charakteristischen Fragmente bei m/z 191 und  $151.^{231}$  Die Identitäten von neun Metaboliten inklusive  $\gamma$ -Tocopherol konnten zusätzlich durch den Vergleich von Retentionszeit, exakter Masse, Isotopenmuster und Fragmentspektrum mit kommerziell erhältlichen Standardsubstanzen eindeutig bestätigt werden, während für die restlichen Metabolite keine Standards erhältlich waren. Obwohl in diesem Zuge auch die beiden eher ungewöhnlichen TGs mit den ungeradzahligen Fettsäuren C15:0 bzw. C17:1 nicht bestätigt werden konnten, wurde deren Identität zweifelsfrei durch die bereits beschriebenen charakteristischen Fragmentierungsmuster (vgl. Abbildung 5-15) in den aufgezeichneten MS/MS-Spektren belegt, da alle theoretischen Fragment-Ionen detektiert werden konnten. Nichtsdestotrotz sollte die Identität gemäß den Empfehlungen der *Chemical* 

*Analysis Working Group* (CAWG) im Rahmen der *Metabolomics Standards Initiative* (MSI) zusätzlich durch den Vergleich mit authentischen Standards bestätigt werden, um einen ausreichenden Beweis zu erbringen.<sup>232</sup> Für die fehlenden Markersubstanzen sollte dies daher bspw. mittels individuell synthetisierter Standards ergänzt werden.

#### 5.2.2.3 Metabolit-Verteilung in unterschiedlichen Ländern

Bei der Betrachtung der Metabolit-Gehalte der 20 Markersubstanzen in den fünf verschiedenen Ländern war die Tendenz markant, dass bei allen Metaboliten, die mindestens eine mehrfach ungesättigte Fettsäure enthielten, sinkende Gehalte von Deutschland über Frankreich und Italien hin zu Georgien und der Türkei zu finden waren, während die Verteilung bei allen Metaboliten, die keine mehrfach ungesättigte Fettsäure besaßen, genau gegenteilig war (siehe Abbildung 5-16). Dies ist insofern zu erwarten, als bereits bekannt ist, dass der Grad der Sättigung der Fettsäuren in Pflanzen mit der mittleren Umgebungstemperatur des Standortes korreliert. Im Allgemeinen gilt: je niedriger die Temperaturen, desto hoher die Anzahl an Doppelbindungen. Dies ist notwendig, damit eine ausreichende Fluidität der Membranen und Fettdepots bei niedrigen Temperaturen für die physiologischen Prozesse gewährleistet werden kann.<sup>260</sup> Es muss jedoch beachtet werden, dass die detektierten Markersubstanzen nur eine kleine Auswahl von hunderten möglichen Metaboliten mit Fettsäureseitenketten darstellen und dass nicht alle Metabolite diesem Verteilungsmuster folgen müssen, obwohl die Summen der jeweiligen Fettsäuren aller Lipide dennoch diesem Temperaturgradienten folgen können.

Bei der Betrachtung der durchschnittlichen Monatstemperaturen in den jeweiligen Haselnussanbaugebieten der fünf Länder lässt sich ein zur generellen Verteilung der Metabolit-Gehalte passendes Bild erkennen (siehe Abbildung 5-17). Während Deutschland, gefolgt von Frankreich, die niedrigsten Temperaturen besitzt, sind die Unterschiede zwischen Italien, Georgien und der Türkei weniger stark ausgeprägt, obwohl diese wiederum deutlich über den Temperaturen von Frankreich liegen. Dasselbe Bild spiegelt sich in den Metabolit-Verteilungen wider, wobei die Unterschiede zwischen Deutschland und Frankreich untereinander bzw. zu den restlichen drei Ländern meist größer sind und die Reihenfolge der türkischen, italienischen und georgischen Proben je nach Metabolit variiert.



Abbildung 5-16: Metabolite-Gehalte (Centered und Scaled) der Markersubstanzen mittels non-targeted LC-MS als relative Gehalte, die (A) mindestens eine mehrfach ungesättigte Fettsäure enthalten; und (B) nur einfach ungesättigte Fettsäuren enthalten sowie γ-Tocopherol; DE: Deutschland, FR: Frankreich, IT: Italien, GE: Georgien, TR: Türkei. (aus Klockmann 2016)

Aufgrund der großen Ähnlichkeit der Proben aus Italien, Georgien und der Türkei wurden wie bereits erwähnt vier Metabolite (TG(14:0/16:0/18:1), TG(15:0/16:0/18:1), TG(16:0/16:1/18:1) und  $\gamma$ -Tocopherol) ergänzend als Markersubstanzen ausgewählt, obwohl sie die Auswahlkriterien nicht vollständig erfüllten (vgl. 5.2.2.1). Es ist jedoch ersichtlich, dass bei diesen vier



Abbildung 5-17: Monatliche Durchschnittstemperaturen in den jeweils wirtschaftlich relevanten Regionen der Herkunftsländer; für jedes Land wurde ein Mittelwert aus drei Städten in oder in unmittelbarer Nähe der jeweiligen Regionen gebildet. (Deutschland: München, Ulm, Nürnberg; Frankreich: Nantes, Toulouse, Bordeaux; Italien: Rom, Turin, Neapel; Türkei: Samsun, Trabzon, Zonguldak; Georgien: Osurgeti, Kutaissi, Zugdidi)<sup>233, 234</sup>

Metaboliten die Unterschiede zwischen diesen drei Ländern im Gegensatz zu den restlichen Markersubstanzen besonders ausgeprägt waren. Insbesondere  $\gamma$ -Tocopherol wies mit einem mehr als zweifach höheren Gehalt in der Türkei gegenüber Georgien und einem drei- bis vierfach höheren Gehalt gegenüber den anderen Ländern besonders hohe diskriminative Eigenschaften auf.

Obwohl die Haselnussproben aus Spanien, Aserbaidschan, den USA und Kanada aufgrund der zu geringen Probenanzahl nicht in die eigentliche Auswertung mit einbezogen wurden, zeigt die Betrachtung der (nicht-repräsentativen) Messergebnisse dieser Proben dennoch ein zu den bisherigen Beobachtungen passendes Bild. Während die Haselnussproben aus Spanien mit ähnlichem klimatischen Verhältnissen zu Italien, Georgien oder der Türkei von den jeweiligen Metabolit-Gehalten zwischen diesen Ländern anzusiedeln sind, ähneln sowohl die klimatischen Bedingungen als auch die Metabolit-Gehalte der Haselnüsse aus den USA den französischen Proben und die aus Kanada den deutschen. Einzig die Proben aus Aserbaidschan zeigten ein variableres Verhalten, sodass sich die Metabolit-Gehalte je nach Markersubstanz teilweise mit eher mit den wärmeren Ländern und teilweise eher mit den kälteren überschnitten. Zur Veranschaulichung dieser Beobachtungen ist eine PCA mit authentischen Proben aus allen analysierten Ländern in Abbildung 5-18 dargestellt. In dieser Darstellung ist darüber hinaus zu erkennen, dass sich z.B. eine der spanischen Proben sehr stark von den andern unterschied durch deren Position bei den türkischen Proben, während die restlichen fünf Proben im Überschneidungsbereich von Italien und Frankreich lagen. Aufgrund der Einfachbestimmung ist es unmöglich festzustellen, ob dieser große Unterschied durch die natürliche biologische Varianz verursacht wurde oder es sich um einen analytischen Ausreißer aufgrund von Messfehlern handelt. Dies verdeutlicht die Wichtigkeit eines ausreichenden Probenumfangs für statistische Bewertungen und begründet die Festlegung einer Mindestprobenanzahl von 10. Eine nachträgliche Wiederholung einzelner Messungen ist aus bereits genannten Gründen in diesem Fall ebenfalls als nicht sinnvoll anzusehen (vgl. Abschnitt 5.2.2).



Abbildung 5-18: PCA *Scores Plot* der 195 authentischen Haselnussproben aus 2014 und 2015 inklusive der Proben aus Aserbaidschan, Spanien, Kanada und den USA, gemessen mit NTNP-pos; unter Verwendung der 20 Markersubstanzen.

# 5.3 Geographische Herkunftsdiskriminierung mittels targeted LC-MS

# 5.3.1 Motivation und Zielsetzung

Obwohl die Ergebnisse der non-targeted UPLC-ESI-QTOF-MS Analysen bereits erfolgreich verliefen und die entwickelte Methode in der Lage ist, Unterschiede der verschiedenen Länder sowohl unter Anwendung des echten non-targeted Ansatzes mit allen 225 *Features* als auch nur unter Verwendung der 20 identifizierten Markersubstanzen darzustellen und zu diskriminieren, bietet die Entwicklung einer targeted HPLC-ESI-QqQ-MS/MS Methode zur absoluten Quantifizierung der Markersubstanzen einige Vorteile demgegenüber und stellt daher im Rahmen eines *Down-Stream* Prozesses eine konsequente Weiterentwicklung der Metabolomics-basierten Herkunftsbestimmung von Haselnüssen dar. Im direkten Vergleich der beiden apparativen Methoden zeichnen sich Triple Quadrupol Geräte insbesondere durch einen größeren dynamischen, linearen Bereich, eine höhere Präzision, weniger Matrix-Effekte und eine bessere Robustheit aus. Im Speziellen unter Anwendung des MRM bieten sie durch eine hohe Sensitivität, Selektivität und Spezifität hervorragende Quantifizierungseigenschaften für die Analyse von komplexen Lebensmittelproben wie bspw. Haselnüsse.<sup>136, 235, 236</sup> Diese Scan-Technik hat die Fähigkeit zur simultanen Analyse einer Vielzahl an Verbindungen in komplexen Matrices bei gleichzeitiger Verringerung des chemischen Untergrunds und

Verbesserung der Selektivität und Sensitivität durch die Akquisition von spezifischen Massenübergängen des Precursor-Ions zu einem oder mehreren Fragment-Ionen.<sup>237, 238</sup>

Neben diesen analytischen Vorteilen war die Motivation des Wechsels vom non-targeted UPLC-QTOF-MS zum targeted HPLC-QqQ-MS/MS unter anderem die vergleichsweise geringen Anschaffungs- und Betriebskosten sowie die weite Verbreitung dieser Geräte in modernen Laborationen der industriellen Qualitätssicherung, staatlichen Lebensmittelüberwachung und des Handelsanalytik-Dienstleistungsgewerbes, wodurch gute Voraussetzungen zur späteren Implementierung der entwickelten Methoden in eine Routineanalytik bestehen. Durch die gute Vergleichbarkeit und Robustheit von Analysenergebnissen auch bei Durchführung in unterschiedlichen Laboren und mit unterschiedlichen Geräten ist der Aufbau einer umfassenden und konstant wachsenden Datenbank denkbar, durch die jährliche Veränderungen des Metaboloms besser verstanden werden sowie die statistischen Modelle ständig verbessert werden könnten. Ohne die konsequente Fortführung der Analysen, die unmöglich von einem Labor alleine getragen werden können, würden außergewöhnliche Phänomene wie extrem kalte und regnerische oder heiße und trockene Sommer aufgrund von dadruch ausgelösten metabolischen Veränderungen zu Fehlinterpretationen der Ergebnisse führen. Darüber hinaus ist eine nachträgliche Messung von Proben der bislang nicht repräsentierten Länder oder Wiederholungsmessungen von Ausreißerproben hier im Gegensatz zum non-targeted Ansatz (siehe Abschnitt 5.2.2) möglich. Das instrumentelle Bias wird durch die Verwendung des internen Standards als Quantifizierungsreferenz rechnerisch eliminiert, sodass auch nachträgliche Messungen vergleichbar mit vorherigen Messungen sind. Im Gegensatz dazu erscheint die Umsetzbarkeit des non-targeted Konzepts in eine Routineanalytik derzeit als nicht zielführend, insbesondere durch die geringere Reproduzierbarkeit und dadurch niedrige Vergleichbarkeit sowie die verhältnismäßig hohen Kosten.

Für eine potentielle Implementierung in die Routineanalytik lag der Fokus bei der Methodenentwicklung wie bereits unter Abschnitt 5.1.2.2 erläutert auf einer robusten, *high-throughput* Methode mit möglichst kurzer Analysenzeit. Die mit Hilfe dieser Methode erstellten "Mini-Fingerprints" der Markersubstanzen sind für jedes Land charakteristisch und können somit zur Unterscheidung verwendet werden.

# 5.3.2 Quantifizierung der Markersubstanzen

Insgesamt wurden 265 Haselnussproben der Erntejahre 2014 und 2015 (207 authentisch und 58 nicht-authentisch) mit der targeted LC-MS Methode analysiert (siehe Tabelle 5-12) und 19

Markersubstanzen mittels internen Standards absolut quantifiziert. Es wurden sowohl jeweils unverdünnte als auch 1:100 verdünnte Extrakte analysiert, damit jede Markersubihrem jeweiligen stanz in linearen Arbeitsbereich quantifiziert werden konnte. Durch die Entwicklung einer Methode mit lediglich 13,5 min Analysendauer konnte eine Dreifachbestimmung für die targeted Messungen durchgeführt werden. Auch hier erfolgte aus bereits erläuterten Gründen (vgl. Abschnitt 5.2.1) die Akquisition in randomisierter Reihenfolge (2014 und 2015 gemeinsam) unter wiederholter Messung

	authentisch	nicht- authentisch
Aserbaidschan <sup>a</sup>	3	2
Deutschland	35	-
Frankreich	115	1
Georgien	10	6
Italien	22	10 <sup>b</sup>
Kanada <sup>a</sup>	1	-
Spanien	6	1
Türkei	14	37 <sup>b</sup>
USA <sup>a</sup>	1	1

Tabelle 5-12: Anzahl der authentischen und nicht-authentischen Haselnussproben für die targeted Messungen.

<sup>a</sup>wegen zu geringer Probenanzahl nicht für die statistische Auswertung verwendet <sup>b</sup>davon geröstet: Italien = 1, Türkei = 4

eines QCs und eines Blindwertes nach jeder 10. Injektion. Die Prozessierung der Daten erfolgte analog zu Abschnitt 5.2.1, wobei das *Scaling* und *Centering* für sämtliche Proben zusammen durchgeführt wurde.

Unter der gleichen Begründung wie bei den non-targeted Messungen wurden auch hier nicht alle der gemessenen Proben aufgrund der zu geringen Probenanzahl (siehe Abschnitt 5.2.2) für die spätere statistische Auswertung verwendet (in diesem Fall Aserbaidschan, Kanada und die USA). Ein kurzer Ausblick über das zu erwartende Verhalten dieser Länder ist unter Abschnitt 5.7 diskutiert. Obwohl für Spanien ebenfalls nur sechs Proben vorlagen, wurden diese Proben im Gegensatz zur non-targeted Analyse dennoch verwendet. Durch die Anwendung der Dreifachbestimmung konnte die analytische Messunsicherheit soweit gesenkt werden, dass systematische Messfehler nahezu ausgeschlossen werden können. Die Problematik der Unklarheit, ob Proben als potentielle analytische oder biologische Ausreißer anzusehen sind wie im Abschnitt 5.2.2.3 war dadurch nicht möglich. Daher wurde Spanien trotz der verhältnismäßig geringen Probenzahl dennoch in die Methoden aufgenommen. Für die restlichen drei Länder war die Probenanzahl mit nur drei bzw. je einer Probe allerdings dennoch zu gering für eine sinnvolle Integration in statistische Analysen. Die regionale Herkunft der zur Auswertung verwendeten Haselnussproben für die targeted Messungen ist bereits unter Abschnitt 5.2.2 in Abbildung 5-10 dargestellt. Insgesamt bildeten somit 202 authentische sowie 55 nicht-authentische Proben die Grundlage für die anschließende Auswertung, wobei bei es sich bei den letzteren neben 50 ungerösteten auch um fünf geröstete Haselnussproben handelte, die zu Testzwecken mit analysiert wurden.

Die Quantifizierung erfolgte wie bereits beschrieben über interne Standards, wobei jeder Stoffklasse (Lysophosphatidylcholine, Phosphatidylcholine, Phosphatidylethanolamine, Diacylglycerole, Triacylglycerole) jeweils ein interner Standard (PC(18:1d7/0:0), PC(15:0/18:1d7), PE(15:0/18:1d7), DG(15:0/18:1d7) und TG(15:0/18:1d7/15:0)) zugeordnet wurde. Durch die berechneten Korrekturfaktoren für jeden Metaboliten konnten somit die absoluten Metabolit-Gehalte in jeder Haselnussprobe bestimmt werden. Lediglich  $\gamma$ -Tocopherol konnte - erwartungsgemäß nach den Ergebnissen im Rahmen der Methodenvalidierung (siehe Abschnitt 5.1.2.4) - nicht quantifiziert werden, da dessen Gehalte stets entweder unterhalb der Nachweis- oder Bestimmungsgrenze lagen. Die berechneten Metabolit-Gehalte der authentischen Haselnussproben sind in Tabelle 5-13 mit ihren jeweiligen Konfidenzintervallen dargestellt sowie in Abbildung 5-19 graphisch mit ihrer jeweiligen Spannbreite als Box-Whiskers-Plots veranschaulicht.

Die Präzision der Messung, ausgedrückt als gemittelter CV der drei Einzelwerte der Dreifachbestimmung für jede der 202 authentischen Proben, welche gemäß den Anforderungen der FDA nicht mehr als 15 % betragen darf, lag bei allen Metaboliten (1,85 - 5,18 %) in einem sehr guten Bereich, sodass die Messung als äußerst präzise und reproduzierbar einzustufen ist (siehe Tabelle 5-14). Wie erwartet sind die Verteilungen der Metabolit-Gehalte vergleichbar mit den Resultaten der non-targeted Messungen. Es lässt sich die selbe Gesetzmäßigkeit für Metabolite mit bzw. ohne mehrfach ungesättigte Fettsäuren feststellen. Auch hier befinden sich die Konzentrationen der Markersubstanzen in den spanischen Proben in der Region der italienischen, türkischen und georgischen Proben. Mit einer Spannbreite vom Faktor 6 existieren bei PC(18:2/18:2) die größten Unterschiede zwischen niedrigstem und höchstem Wert. Im Gegensatz zu bisherigen Erkenntnissen sind die Unterschiede der Metabolit-Gehalte bei TG(15:0/16:0/18:1) und TG(16:0/16:1/18:2) sehr viel weniger stark ausgeprägt. Die Gehalte reichen im Allgemeinen von 1,03 mg/kg (DG(16:0/16:1) bis zu 4350,25 mg/kg (DG(18:1/18:1) bei relativen Konfidenzintervallen ( $\alpha = 0,05$ ) zwischen

	PC(18:2/0:0)	DG(16:0/16:1)	DG(16:0/18:1)	DG(18:2/18:2)
Deutschland	$5,00 \pm 0,28$	$1,03 \pm 0,05$	$470,10 \pm 25,02$	$1633,95 \pm 73,10$
Frankreich	$3,86 \pm 0,09$	$1,34 \pm 0,03$	$690,\!28 \pm 13,\!93$	$1236,34 \pm 29,41$
Georgien	$2,26 \pm 0,19$	$1,98 \pm 0,10$	$917,22 \pm 33,51$	$500,45 \pm 36,56$
Italien	$2,\!48 \pm 0,\!12$	$1,59 \pm 0,07$	$854,77 \pm 29,74$	$780,03 \pm 45,66$
Spanien	$1,87 \pm 0,23$	$1,87 \pm 0,27$	$992,95 \pm 117,71$	$966,65 \pm 134,04$
Türkei	$1,76 \pm 0,07$	$1,78 \pm 0,08$	$953,\!89 \pm 48,\!02$	$535,37 \pm 33,34$
	DG(18:1/18:1)	TG(2:0/18:2/18:2)	TG(2:0/18:1/18:2)	PE(16:0/18:2)
Deutschland	$2052,64 \pm 86,24$	$394,27 \pm 23,51$	$1295,90 \pm 57,96$	$127, 12 \pm 5, 44$
Frankreich	$3044,35 \pm 50,27$	$220,57 \pm 9,76$	$893,\!69 \pm 29,\!03$	$92,87 \pm 1,64$
Georgien	$3911,33 \pm 108,52$	$78,99 \pm 13,25$	$423,13 \pm 55,02$	$43,22 \pm 4,89$
Italien	$3779,80 \pm 73,22$	$71,79 \pm 9,10$	$398,00 \pm 35,82$	$58,98 \pm 2,53$
Spanien	$4295,52 \pm 234,07$	$154,89 \pm 29,79$	$711,24 \pm 110,87$	$53,\!48 \pm 8,\!77$
Türkei	$4350,25 \pm 267,33$	$103,22 \pm 15,61$	$505,25 \pm 53,06$	$41,14 \pm 2,41$
	PE(18:2/18:2)	PE(18:1/18:1)	PC(16:0/18:3)	PC(16:0/18:2)
Deutschland	$143,\!63\pm 6,\!12$	$491,01 \pm 18,34$	$4,95 \pm 0,24$	$112,98 \pm 4,33$
Frankreich	$86,09 \pm 2,45$	$602,82 \pm 10,68$	$3,74 \pm 0,08$	$85,90 \pm 1,47$
Georgien	$23,11 \pm 3,25$	$621,24 \pm 30,50$	$1,64 \pm 0,20$	$49,87 \pm 5,16$
Italien	$39,88 \pm 2,45$	$650,24 \pm 20,02$	$2,20 \pm 0,11$	$60,17 \pm 2,55$
Spanien	$32,69 \pm 5,06$	$569,65 \pm 31,05$	$1,82 \pm 0,39$	$53,85 \pm 8,92$
Türkei	$22,62 \pm 2,65$	$693,27 \pm 31,34$	$1,41 \pm 0,14$	$40,24 \pm 2,41$
	PC(18:2/18:2)	PC(18:1/18:2)	TG(14:0/16:0/18:1)	TG(15:0/16:0/18:1)
Deutschland	$177,01 \pm 7,90$	$258,52 \pm 5,78$	$99,02 \pm 4,66$	$18,62 \pm 0,99$
Frankreich	$110,08 \pm 2,97$	$230,97 \pm 2,84$	$132,87 \pm 3,86$	$20,90 \pm 0,40$
Georgien	$34,54 \pm 5,09$	$129,44 \pm 8,58$	$192,40 \pm 18,76$	$13,31 \pm 0,66$
Italien	$54,46 \pm 2,85$	$168,63 \pm 4,33$	$182,41 \pm 12,35$	$15,84 \pm 0,98$
Spanien	$45,21 \pm 8,58$	$144,35 \pm 14,13$	$208,90 \pm 24,71$	$14,76 \pm 1,28$
Türkei	$29,06 \pm 3,29$	$120,27 \pm 5,92$	$119,70 \pm 7,08$	$15,86 \pm 0,80$
	TG(16:0/16:1/18:1)	TG(17:1/18:1/18:2)	TG(18:2/18:2/18:3)	γ-Tocopherol
Deutschland	$3658,28 \pm 150,83$	$572,76 \pm 26,86$	$1082,30 \pm 64,11$	< LLOQ
Frankreich	$3505,22 \pm 78,31$	$439,32 \pm 10,00$	$730,32 \pm 18,78$	< LLOQ
Georgien	$3708,43 \pm 434,26$	$292,88 \pm 24,51$	$459,13 \pm 34,98$	< LLOQ
Italien	$3198,\!68 \pm 160,\!06$	$262,15 \pm 15,82$	$460,93 \pm 30,57$	< LLOQ
Spanien	$4233,26 \pm 538,53$	$338,35 \pm 47,09$	$688,94 \pm 104,53$	< LLOQ
Türkei	$2533,94 \pm 124,89$	$346,11 \pm 25,45$	$692,57 \pm 56,16$	< LLOQ

Tabelle 5-13: Metabolit-Gehalte der 202 authentischen Haselnussproben mit Konfidenzintervallen $(\alpha = 0,05)$  in mg/kg. (modifiziert nach Klockmann 2017)

1,22 % und 21,45 %. Die mittleren, relativen Konfidenzintervalle für jeden der Markersubstanzen für alle Herkunftsländer reichen von 3,69 % für DG(18:1/18:1) bis 12,37 % für TG(2:0/18:2/18:2) (siehe Tabelle 5-14). Sie sind als repräsentativ für die Variabilität Metabolite innerhalb der jeweiligen Herkunftsländer zu verstehen. Dementsprechend eignen sich Metabolite mit einer geringeren Variabilität besser für die vorliegende Fragestellung, da berechnete Unterschiede diskreter sind und daraus eine größere statistische Sicherheit



#### Ergebnisse und Diskussion

Abbildung 5-19: Absolute Metabolite-Gehalte der Markersubstanzen für alle 202 authentischen Haselnussproben in mg/kg, gemessen mittels targeted LC-MS; DE: Deutschland, FR: Frankreich, IT: Italien, GE: Georgien, TR: Türkei, ES: Spanien. (modifiziert nach Klockmann 2017)

resultiert. Bei der Berechnung eines PCA-Modells (siehe Abbildung 5-20) konnte eine erklärte Gesamtvarianz von PC1 und PC2 von 73 % erreicht werden, womit im Vergleich zum non-targeted Modell ein identischer Wert erreicht werden konnte, obwohl  $\gamma$ -Tocopherol als einer der einflussreichsten Markersubstanzen nicht mit einbezogen werden konnte, darüber hinaus mit Spanien jedoch ein sechstes Land ergänzt wurde.



Abbildung 5-20: PCA *Scores Plot* der 202 authentischen Haselnussproben in Dreifachbestimmung aus 2014 und 2015 der targeted LC-MS, (A) PC1 vs. PC2; und (B) PC1 vs. PC3. (modifiziert nach Klockmann 2017)

Im Abbild des *Scores Plots* unterschieden sich die beiden PCA-Modelle der targeted (siehe Abbildung 5-20) und non-targeted (siehe Abbildung 5-14) Analyse vor allem in Bezug auf die Streuung der Proben innerhalb eines Landes. Dies könnte möglicherweise lediglich auf die größere Anzahl an Messdaten durch die Dreifachbestimmung zurückzuführen sein. In der generellen Verteilung ähnelten sich beide Modelle jedoch sehr stark, wobei sich besonders die spanischen, italienischen, türkischen und georgischen Proben überlappen, aber dennoch diskrete Cluster zu erkennen waren. Die deutschen und französischen Proben überschnitten sich ebenfalls, wenn auch weniger stark ausgeprägt. Bei der Betrachtung der Erntejahre


Abbildung 5-21: P-values der T-Test Berechnungen anhand der Metabolit-Gehalte der targeted LC-MS Bestimmung zwischen Länderpaarungen für jeden Metaboliten. (aus Klockmann 2017)

keine fanden sich auch hier gerichteten Unterschiede, sodass einer Erntejahrvon unabhängigen Verteilung Metabolite der auszugehen ist.

Zur Bewertung der Signifikanz von Konzentrationsunterschieden zwischen den Ländern wurde für jede Länderpaarung und jeden Metaboliten ein T-Test berechnet. P-values zwischen 0,001 und 0,05 indizieren einen Beweis gegen die Nullhypothese (Metabolit-Gehalte der beiden Länder sind gleich) und bedeuten, dass sich diese beiden Gruppen somit signifikant unterscheiden, während p-values < 0,001 einen

sehr starken Beweis für einen hochsignifikanten Unterschied liefern. Ein p-value > 0,05 steht für eine Bestätigung der Nullhypothese.<sup>239</sup> Die Anzahl der Metabolite mit signifikanten bzw.

Metabolit	Mittlere Präzision der	Relatives	p-value (ANOVA)
	Dreifachbestimmungen	Konfidenzintervall	
	[%]	[%]	
PC(18:2/0:0)	2,84	6,11	3,78.10-63
DG(16:0/16:1)	5,18	5,99	4,40.10-34
DG(16:0/18:1)	2,27	5,23	2,89·10 <sup>-42</sup>
DG(18:2/18:2)	1,92	6,68	3,87.10-64
DG(18:1/18:1)	1,85	3,69	1,72.10-53
TG(2:0/18:2/18:2)	2,70	12,4	1,08.10-51
TG(2:0/18:1/18:2)	2,07	9,30	5,72·10 <sup>-50</sup>
PE(16:0/18:2)	2,71	7,32	1,09.10-61
PE(18:2/18:2)	3,21	9,08	1,14.10-72
PE(18:1/18:1)	2,37	3,91	5,32·10 <sup>-22</sup>
PC(16:0/18:3)	2,94	9,34	1,80.10-56
PC(16:0/18:2)	1,94	7,11	8,19·10 <sup>-57</sup>
PC(18:2/18:2)	2,25	9,57	1,16.10-68
PC(18:1/18:2)	2,01	4,56	8,85·10 <sup>-63</sup>
TG(14:0/16:0/18:1)	4,54	6,98	$1,71 \cdot 10^{-28}$
TG(15:0/16:0/18:1)	2,51	5,36	9,66·10 <sup>-37</sup>
TG(16:0/16:1/18:1)	4,20	6,79	1,05.10-22
TG(17:1/18:1/18:2)	4,06	7,11	1,81.10-39
TG(18:2/18:2/18:3)	5,00	7,67	5,76.10-37

Tabelle 5-14: Mittlere Präzision, p-values der ANOVA Berechnungen und relative Konfidenzintervalle der Markersubstanzen mittels targeted LC-MS.

Universität Hamburg, Hamburg School of Food Science, 2017

Unterschieden hochsignifikanten zwischen den Länderpaarungen liefern eine Aussage über die diskriminativen Eigenschaften des Mini-Fingerprints bzw. im Umkehrschluss über die Ähnlichkeit der jeweiligen Länder. Obwohl die Konzentrationsbereiche bei einigen Metaboliten starke Überschneidungen bei den betroffenen Länderpaaren aufwiesen, besaßen Türkei-Georgien, Türkei-Frankreich, Italien-Georgien und Italien-Spanien je 10 - 14 Markersubstanzen mit hochsignifikanten Unterschieden (siehe Abbildung 5-21). Bei den restlichen Länderpaarungen lag die Anzahl derartiger hochsignifikanter Metabolite bei > 17 bei gleichzeitig nicht-signifikanten maximal zwei Unterschieden. Insbesondere bei der Unterscheidung von Georgien und Spanien könnten Probleme auftreten, da bei dieser Kombination acht Metabolite keine signifikanten Unterschiede aufwiesen. Allerdings könnten die sieben hochsignifikant unterschiedlichen Metabolite dennoch für eine effiziente Diskriminierung ausreichen.



Abbildung 5-22: Metabolit-Gehalte (Centered und Scaled) der 202 authentischen Proben mit Konfidenzintervallen ( $\alpha = 0,05$ ) in den verschiedenen Ländern der targeted LC-MS Messung. (modifiziert nach Klockmann 2017)



Abbildung 5-23: Spinnendiagramme der Metabolit-Gehalte (Centered und Scaled) der authentischen bzw. nicht-authentischen Haselnussproben auf Basis der targeted LC-MS. (modifiziert nach Klockmann 2017) Für Deutschland lagen keine nichtauthentischen Haselnussproben vor.

Bei der Berechnung der ANOVA mit den vorliegenden Proben zeigte PE(18:2/18:2) von den 19 Markersubstanzen die größten diskriminativen Eigenschaften mit dem niedrigsten p-value, gefolgt von PC(18:2/18:2) und DG(18:2/18:2) (siehe Tabelle 5-14). Auch bei den T-Test Berechnungen gehörten diese Metabolite zu den potentesten Markersubstanzen mit den größten Anzahlen an hochsignifikanten Unterschieden zwischen sämtlichen Länderpaarungen.

dynamischen Wegen des großen Konzentrationsbereiches der verschiedenen Markersubstanzen mussten die Metabolit-Gehalte auch in diesem Fall vor der multivariaten Statistik einem Scaling Schritt unterworfen werden (wie z.B. auch für die PCA) um die Zahlenwerte auf ein einheitliches, vergleichbares Niveau zu bringen. Die Scaling Parameter wurden für ieden Metaboliten auf Basis der authentischen Proben separat berechnet, wobei erneut das Interquartile Scaling verwendet wurde. Diese Parameter können anschließend zur Prozessierung auf die nicht-authentischen Proben bzw. weitere oder zukünftige Proben angewendet werden, sodass nachfolgende Daten schnell

und einfach in die vorhandenen statistischen Modelle integriert werden können. Darüber hinaus ermöglichen die skalierten Werte eine vernünftige visuelle Darstellung der Daten ohne Diskriminierung von niedrigen Metabolit-Gehalten. Die daraus erstellten Balkendiagramme stellen einen für jedes Herkunftsland charakteristischen Mini-Fingerprint (oder auch "Barcode") dar und ermöglichen einen unmittelbaren optischen Vergleich (siehe Abbildung 5-22). Da sich die absoluten Metabolit-Gehalte der französischen Proben in Relation zu den anderen Ländern stets im mittleren Bereich befanden, streuen dessen skalierte Werte um Null. Wie bereits durch die vorhergehenden Auswertungsmethoden vermutet, sind sich die Mini-Fingerprints der spanischen und georgischen Proben sehr ähnlich, jedoch können die beiden Länder dennoch optisch noch differenziert werden. Die restlichen Länder können durch diese Art der Darstellung leicht voneinander unterschieden werden.

Zusätzlich zu den authentischen Haselnussproben wurden 50 nicht-authentische Proben in gleicher Weise analysiert und die 19 Markersubstanzen quantifiziert. Wegen der aktuellen Marktsituation handelte es sich bei den Proben hauptsächlich um türkische Proben (33), während von den italienischen neun, von den georgischen sechs und von den französischen und spanischen jeweils eine erhalten werden konnten. Bei 77 % der Proben überschnitten sich die Konfidenzintervalle mit denen der authentischen Proben. Ohne die spanische und französische Probe überschnitten sich die Konfidenzintervalle sogar in 93 % der Fälle. Beim Vergleich der gescaleten Werte der authentischen und nicht-authentischen Proben in Form von Spinnendiagrammen können diese Beobachtungen visuell bestätigt werden (siehe Abbildung 5-23). Die Formen der Diagramme spiegeln die Ähnlichkeiten bzw. Unterschiede der illustrierten Probengruppen wider, wobei bis auf die eine französische Probe die Formen der nicht-authentischen und authentischen Proben größtenteils übereinstimmen. Demensprechend unterschieden sich die Metabolome kaum. Analog zu den Ergebnissen der Mini-Fingerprints können sich die verschiedenen Länder untereinander (ähnlich den Mini-Fingerprints) einfach visuell voneinander differenzieren lassen, wobei Spanien und Georgien die größten Ähnlichkeiten bei dennoch charakteristischen Unterschieden aufweisen.

### 5.3.3 Vergleich mit Literaturwerten

Bislang existieren nur wenige Daten über die Gehalte an konkreten TGs in Haselnüssen. Während sich diverse Studien ausschließlich mit der Bestimmung der prozentualen Anteile oder Gehalte der verschiedenen Fettsäuren als Summenparameter aller in der Probe vorhanden Lipide beschäftigen (für Quellen siehe Tabelle 3-2, Abschnitt 3.1), existieren nur wenige Veröffentlichungen, die sich mit diskreten TGs beschäftigen. Diese Studien geben deren Menge lediglich als prozentualen Anteil am Gesamt-Triacylglycerol-Gehalt an,<sup>25, 26, 54, 216, 240</sup> wobei von den identifizierten Markersubstanzen bisher nur Daten zu TG(18:2/18:2/18:3) mit einem Anteil von 0,1 - 3,2 % publiziert sind.<sup>228, 241-243</sup> Darüber hinaus

waren die Metabolite TG(16:0/16:1/18:1), TG(14:0/16:0/18:1) und TG(17:1/18:1/18:2) zwar Inhalt einiger Studien, konnten in den untersuchten Haselnussproben allerdings nicht nachgewiesen werden.<sup>228, 241</sup> Bei der Betrachtung der in dieser Arbeit analysierten Gehalte fällt auf, dass für die beiden letzteren Metabolite zwar niedrigere Gehalte als für TG(18:2/18:2/18:3) bestimmt wurden und diese somit möglicherweise unterhalb der Nachweisgrenzen der in den Studien verwendeten Methoden liegen könnten, die Konzentrationsspanne von TG(16:0/16:1/18:1) mit 2534 - 4233 mg/kg jedoch deutlich über der von TG(18:2/18:2/18:3) mit 460 - 1082 mg/kg liegt und daher eigentlich hätte detektiert werden sollen. Da sich jedoch in dieser Arbeit im Zuge der Validierung bereits herausgestellt hatte, dass die Nachweisgrenzen je nach Metabolit teilweise erheblich variieren, könnte dieser Effekt möglicherweise ebenfalls zu einer ausbleibenden Detektion in diesen Studien geführt haben. Zu Gehalten an diskreten DGs, PCs oder PEs sind darüber hinaus bislang noch keine Daten verfügbar.

### 5.4 Vorhersagemodelle und Klassifizierung nicht-authentischer Proben

Anhand der authentischen Haselnussproben der beteiligten Länder wurden mehrere statistische Vorhersagemodelle mit unterschiedlichen multivariaten Verfahren getestet und bewertet. Diese Modelle basierten auf den häufig angewandten Methoden *Soft Independent Modelling of Class Analogy* (SIMCA), lineare Diskriminanzanalyse auf Grundlage von Hauptkomponenten (PCA-LDA), *Support Vector Machine Classification* (SVM-C) sowie einem selbst entwickelten Verfahren basierend auf den Konfidenzintervallen ausgewählter Markersubstanzen, im Folgenden als *Metabolite Confidence Intervall Modell* (MCIM) bezeichnet. Zur Bewertung der Effizienz der Modelle wurden einerseits die bereits zur Methodenerstellung genutzten, authentischen Proben (Training-Datensatz) sowie andererseits die 40 nicht-authentischen Proben (Vorhersage-Datensatz) mit den jeweiligen Modellen klassifiziert und der Anteil korrekt bzw. falsch vorhergesagter Proben berechnet. (Modellge-nauigkeit bzw. Vorhersagegenauigkeit)

### 5.4.1 Non-targeted LC-MS

Im Rahmen der non-targeted LC-MS Analyse wurden 195 authentischen Haselnussproben der Länder Deutschland, Frankreich, Georgien, Italien und Türkei zur Modellerstellung (SIMCA, PCA-LDA, SVM-C und MCIM) genutzt und anschließend 40 nicht-authentische Proben vorhergesagt (siehe Abschnitt 5.2.2). Die unterschiedlichen Modelle wurden sowohl auf Basis aller 225 *Features* als auch nur der 20 Markersubstanzen berechnet. Die jeweils besten Ergebnisse der Modellberechnungen sind in Tabelle 5-15 zusammengefasst.

Vorhersagemodell	Probenset	richtig	falsch	nicht klassifiziert
	Training	97,4%	2,6%	-
rCA-LDA	Vorhersage	70,0%	30,0%	-
	Training	97,4%	2,6%	0,0%
rca-lda +Sinica	Vorhersage	60,0%	10,0%	30.0%
MCIMb	Training	72,8%	22,1%	5.1%
MCIIVI	Vorhersage	77,5%	22,5%	0,0%
	Training	72,8%	17,4%	9.7%
	Vorhersage	70,0%	7,5%	22.5%
	Training	100,0%	0,0%	-
5 V IVI	Vorhersage	90,0%	10,0%	-
SVMa+SIMC Ab	Training	100,0%	0,0%	0,0%
	Vorhersage	80,0%	0,0%	20.0%

Tabelle 5-15: Modell- und Vorhersagegenauigkeiten für die verschiedenen Vorhersagemodelle der nontargeted LC-MS Messungen. (modifiziert nach Klockmann 2016)

<sup>a</sup> alle 225 *Features* zur Berechnung verwendet

<sup>b</sup> 20 Markersubstanzen zur Berechnung verwendet

Für die PCA-LDA wurden verschiedene Algorithmen (Linear, Quadratic und Mahalanobis) sowie verschiedene Anzahlen an Hauptkomponenten (3 - 11) getestet, wobei Mahalanobis mit 9 Komponenten die besten Ergebnisse erzielen konnte. Die Durchführung einer herkömmlichen LDA auf Basis des originalen Datensatzes lieferte im Vergleich zur PCA-LDA keine besseren Resultate. Dies liegt vor allem darin begründet, dass die LDA für multidimensionale Daten nur begrenzt geeignet ist, weshalb die Dimensionsreduktion mittels vorgeschalteter PCA und der Verwendung der daraus berechneten Hauptkomponenten statt der Originalvariablen zu besseren Resultaten führt.<sup>159</sup> Auf diese Weise konnten 97,4 % der Training-Daten korrekt klassifiziert werden. Von den nicht-authentischen Proben konnten 70,0 % richtig vorhergesagt werden.

Beim SVM-C wurden verschiedene SVM Typen (Nu-SVC und C-SVC), Kernel Typen (Linear, Polynomial 2. bis 4. Grades, Radial Basis Function und Sigmoid) mit jeweils entsprechenden Nu-, Gamma- und Offset-Werten für Nu-SVC bzw. C-, Gamma- und Offset-Werten für C-SVC getestet. Die optimalen Werte für Nu (0-1), C (0,01-100), Gamma (0,01-100) und Offset (0-4) wurden semi-automatisch über die sogenannte *Grid Search Funktion* bestimmt. Diese berechnet die Validierungs- und Modellgenauigkeit jeweils für verschiedene Werte der entsprechenden Parameter. Die optimalen Parameter für jeden SVM- und Kernel

Typ wurden individuell ausgewählt. Daraus ergab sich, dass die besten Ergebnisse durch das C-SVC mit linearem Kernel Typ bei einem Offset von 0 und einem C-Wert von 0,01 erhalten werden konnten. (Gamma-Wert bei linearem Kernel Typ nicht benötigt) Anhand der idealen Parameter wurde eine Modellgenauigkeit von 100 % bei einer Vorhersagegenauigkeit von 90 % erhalten. Zusätzlich war es bei der Berechnung der SVM möglich die Validierungsgenauigkeit via *Full Cross Validation* zu berechnen, welche mit 97,4 % auf eine gute Robustheit des entwickelten Modells schlussfolgern lässt.

Sowohl für PCA-LDA als auch SVM konnten bei Verwendung der 225 *Features* bessere Ergebnisse erzielt werden, als bei der Reduktion auf die 20 Markersubstanzen (Modellgenauigkeit: 90,3 % für SVM und 88,2 % für PCA-LDA bei 20 Markersubstanzen im Gegensatz zu 100 % bzw. 97,4 % bei 225 *Features*)

Aufgrund dieser Tatsache wurde in Hinsicht auf einen folgenden Vereinfachungsprozess hin zu targeted Ansätzen zusätzlich ein einfaches, selbst erstelltes Modell entwickelt (MCIM), welches nur die 20 Markersubstanzen als Diskriminanzmetabolite nutzt. Die Funktionsweise des MCIM ist in Abbildung 5-24 veranschaulicht. Im ersten Schritt dieses Modells wurden jeweils drei Diskriminanzmetabolite für jedes Länderpaar verwendet, welche die größten signifikanten Unterschiede zwischen diesen beiden Gruppen besaßen (siehe Tabelle 5-16). Die Auswahl der entsprechenden Diskriminanzmetabolite erfolgte auf Basis der jeweils drei niedrigsten p-values der T-Test Berechnungen (siehe Abschnitt 5.3.2). Die Mittelwerte aus unterem Konfidenzintervall ( $\alpha = 0.05$ ) des Landes mit dem höheren Gehalt und oberem Konfidenzintervall ( $\alpha = 0.05$ ) des Landes mit dem niedrigeren Gehalt dienten als Diskriminanzlimit für die jeweiligen Metabolite und Länderpaare. Unbekannte Proben mit einem Metabolit-Gehalt über dem jeweiligen Limit wurden demnach der oberen Gruppe zugeordnet und umgekehrt. Damit eine Probe für ein bestimmtes Land vorhergesagt werden konnte, mussten jeweils zwei oder drei der Metabolite dem jeweiligen Land zugeordnet werden in allen der vier möglichen Länderpaare (z.B. DE vs. FR, DE vs. GE, DE vs. IT, DE vs. TR). Dementsprechend erfolgte die Klassifizierung einer Probe entweder zu einem bestimmten Land oder sie blieb aus, wenn keines der Länder vier Zuordnungen erhielt.



Abbildung 5-24: Erläuterung des MCIM-Modells am Beispiel Türkei. (A) Länderzuordnung bei Vergleich zweier Länder anhand der drei jeweils diskriminativen Metabolite am Beispiel Türkei vs. Georgien; und (B) Klassifizierung einer Probe zu "Türkei" durch 4x Länderzuordnung.

Tabelle	5-16:	Verwendete	Markersubstanzen	für	das	<b>MMIC-Modell</b>	der	non-targeted	Messungen.
(modifiz	ziert na	ach Klockmai	ın 2016)						

	TR	GE	IT	FR	DE
		TG(14:0/16:0/18:1)	γ-Tocopherol	γ-Tocopherol	DG(16:0/18:1)
TR	х	TG(16:0/16:1/18:1)	PC(18:2/0:0)	DG(16:0/16:1)	PE(18:2/18:2)
		TG(18:2/18:2/18:3)	TG(16:0/16:1/18:1)	PE(18:2/18:2)	PC(18:2/18:2)
	TG(14:0/16:0/18:1)		γ-Tocopherol	DG(18:2/18:2)	PE(18:2/18:2)
GE	TG(16:0/16:1/18:1)	Х	PC(18:2/0:0)	PE(18:2/18:2)	PC(16:0/18:3)
	TG(18:2/18:2/18:3)		DG(18:2/18:2)	PC(18:2/18:2)	PC(18:2/18:2)
	γ-Tocopherol	γ-Tocopherol		TG(2:0/18:2/18:2)	DG(16:0/16:1)
IT	PC(18:2/0:0)	PC(18:2/0:0)	Х	TG(14:0/16:0/18:1)	PE(18:2/18:2)
	TG(16:0/16:1/18:1)	DG(18:2/18:2)		TG(17:1/18:1/18:2)	PC(18:2/18:2)
	γ-Tocopherol	DG(18:2/18:2)	TG(2:0/18:2/18:2)		DG(18:1/18:1)
FR	DG(16:0/16:1)	PE(18:2/18:2)	TG(14:0/16:0/18:1)	х	PE(18:2/18:2)
	PE(18:2/18:2)	PC(18:2/18:2)	TG(17:1/18:1/18:2)		TG(15:0/16:0/18:1)
	DG(16:0/18:1)	PE(18:2/18:2)	DG(16:0/16:1)	DG(18:1/18:1)	
DE	PE(18:2/18:2)	PC(16:0/18:3)	PE(18:2/18:2)	PE(18:2/18:2)	х
	PC(18:2/18:2)	PC(18:2/18:2)	PC(18:2/18:2)	TG(15:0/16:0/18:1)	

Für die Proben, die auf diese Weise keinem Land zugeordnet werden konnten (nur sechs der Training-Proben und keine der Vorhersage Proben), wurden im zweiten Schritt für jedes Land die Anzahl der 20 Markersubstanzen gezählt, dessen Gehalte innerhalb der jeweiligen Konfidenzintervalle ( $\alpha = 0.05$ ) lagen. Aufgrund der häufig überlappenden Konfidenzintervalle waren hierbei Mehrfachzuordnungen möglich. Aus diesem Grund wurde eine Probe nur dem jeweiligen Land mit den meisten Übereinstimmungen zugeordnet, wenn dessen Anzahl mindestens um drei größer war als die des nächst niedrigeren. Andernfalls verblieb diese Probe nicht klassifiziert. Dieser zweite Schritt wurde hinzugefügt, da einzelne Metabolit-Gehalte (auch bei den zur Diskriminierung genutzten Markersubstanzen) variieren können, obwohl der Rest des Metaboloms zum Gruppenmittelwert passt. Dieses verhältnismäßige einfache Modell wurde entwickelt, um zu testen, ob eine präzise Vorhersage der Haselnüsse auch ohne komplexe Algorithmen und kostenpflichtige Software zu realisieren ist. Mit Hilfe dieses Modells wurden 72,8 % der Training-Daten korrekt klassifiziert, während 22,1 % falsch und weitere 5,1 % nicht eindeutig zugeordnet wurden. Von den nicht-authentischen Proben konnten 77,5 % korrekt vorhergesagt werden, während 22,5 % falsch klassifiziert wurden. Im Gegensatz zu den etablierten Methoden (PCA-LDA und SVM-C) konnte dieses vergleichsweise einfache Modell somit einigermaßen gute Resultate erzielen, liegt jedoch deutlich hinter der Leistungsfähigkeit der komplexeren Modelle.

Für das SIMCA-Modell wurde für jedes Herkunftsland eine separate PCA berechnet und die optimale Anzahl an Hauptkomponenten zum Erreichen einer minimalen, erklärten Gesamtvarianz von 95 % automatisch bestimmt. Das Signifikanzlimit wurde auf 1 % festgelegt um eine entsprechende statistische Sicherheit generieren zu können. Im Gegensatz zu den anderen Vorhersagemethoden sind bei der SIMCA Mehrfachklassifizierungen möglich. Die Verwendung von 20 Markersubstanzen zur Modellerstellung erwies sich hierbei als geeigneter zur Klassifizierung der Proben. Ein entsprechendes Bild ergab sich bereits bei der Berechnung einer herkömmlichen PCA, dessen *Scores Plot* unter Verwendung der 20 Metabolit anstatt aller 225 ebenfalls ein deutlich besseres Resultat erzielen konnte. Nur 13,8 % der nicht-authentischen Proben konnten mittels SIMCA eindeutig korrekt klassifiziert werden, während immerhin 25,7 % des Training-Datensatzes eindeutig korrekt klassifiziert wurde. Dies liegt vor darin begründet, dass Mehrfachzuordnungen möglich sind, welche wegen der großen Ähnlichkeit vieler Haselnussproben aus unterschiedlichen Ländern und der daraus resultierenden Überlappung vieler Länder bei 168 der 195 Haselnussproben auftraten und keine eindeutige Zuordnung ermöglichten. Daher ist die Verwendung des SIMCA-

Modells zur Klassifizierung der Haselnussproben alleine nicht sinnvoll. Allerdings wurde nur eine einzige Probe falsch negativ klassifiziert, wodurch sich die Option der Verwendung als "Kontrollmodell" ergab. Wenn bei PCA-LDA, SVM-C und MCIM nur vorhergesagte Länder akzeptiert wurden, die auch bei der SIMCA dem jeweiligen Land zuordnet wurden, konnte die Falsch-Vorhersagerate der Training- und Vorhersage-Datensätze in allen Fällen reduziert werden. In den meisten Fällen wurden falsch-positive Resultate zu "nicht klassifizierten" (keine Übereinstimmung zwischen PCA-LDA, SVM bzw. MCIM mit SIMCA) geändert. Einige der bislang korrekt vorhergesagten Proben konnten dadurch allerdings ebenfalls keiner Gruppe eindeutig zugeordnet werden. Nichtsdestotrotz überwiegt der Vorteil der Reduzierung von falsch vorhergesagten Proben, da ungerechtfertigte Reklamationen zu Folgeschäden führen könnten, während nicht eindeutig klassifizierte Proben erneut oder anderweitig analysiert werden könnten. Daher erscheint die Kombination aus SIMCA mit einem anderen Vorhersagemodell im Kontext der non-targeted-basierten Herkunftsbestimmung von Haselnüssen als sinnvolle Maßnahme zur Verbesserung des statistischen Outputs.

Im Vergleich der drei verschiedenen multivariaten Verfahren zur Erstellung der Vorhersagemodelle für Haselnussproben erwies sich die SVM als beste Methode mit einer Vorhersagerate von 90 % bei gleichzeitig 10 % Falsch-Vorhersagen (bzw. 80 % mit SIMCA bei 0 % Falsch-Vorhersagen; Modellgenauigkeit unverändert 100 %). Während die PCA-LDA sowohl mit als auch ohne SIMCA zwar eine deutlich bessere Modellgenauigkeit erzielen konnte als das MCIM, liegen die Defizite dieser Methode insbesondere in der Vorhersagekapazität der nicht-authentischen Proben, welche noch unter der des MCIM rangiert. Somit liegt die Vermutung nahe, dass es bei der Erstellung des Modells zum Overfitting kam. Eine Kreuzvalidierung zur Abschätzung dieses Effekts konnte mit der verwendeten Software allerdings nicht durchgeführt werden. Eine weiterführende Diskussion zur Effizienz der verschiedenen Modelle erfolgte im Anschluss an die Ergebnisse der targeted-basierten Vorhersagemodelle (siehe Abschnitt 5.4.2).

# 5.4.2 Targeted LC-MS

Im Rahmen der targeted LC-MS Messungen wurden 202 authentischen Haselnussproben der Länder Deutschland, Frankreich, Georgien, Italien, Türkei und zusätzlich Spanien zur Modellerstellung (SIMCA, PCA-LDA, SVM und MCIM) genutzt und anschließend 50 nichtauthentische Proben vorhergesagt (siehe Abschnitt 5.3.2). Die unterschiedlichen Vorhersagemodelle wurden auf Basis der 19 erfolgreich quantifizierten Markersubstanzen berechnet. Dabei erfolgte die Berechnung der Modell- und Vorhersagegenauigkeiten sowohl für jede der Einzelbestimmungen individuell als auch für jede Dreifachbestimmung zusammenfasst. Dazu erfolgte die Klassifizierung einer Probe zu einem bestimmten Land nur dann, wenn mindestens zwei der drei Messungen der Dreifachbestimmung übereinstimmende Ergebnisse aufwiesen. Andernfalls wurde die Probe als "nicht klassifiziert" eingestuft. Die Ergebnisse der Modellberechnungen sind in Tabelle 5-17 zusammengefasst.

			Einzelw	erte	Dreifachbestimmung		
Vorhersagemodell	Probenset	richtig	falsch	nicht klassifiziert	richtig	falsch	nicht klassifiziert
LDA	Training	98,3%	1,7%	-	98,5%	1,5%	0,0%
LDA	Vorhersage	24,0%	76,0%	-	23,5%	76,5%	0,0%
	Training	98,0%	1,7%	0,3%	98,5%	1,5%	0,0%
LDA+5IIVICA	Vorhersage	20,0%	52,0%	28,0%	18,0%	52,0%	30,0%
MCIM	Training	73,8%	24,4%	1,8%	73,3%	24,2%	2,5%
IVICIIVI	Vorhersage	55,3%	39,3%	5,3%	58,0%	38,0%	4,0%
MCIM+SIMCA	Training	73,6%	11,1%	15,3%	73,3%	11,4%	15,3%
WICHWI + SHVICA	Vorhersage	35,3%	14,7%	50,0%	36,0%	14,0%	50,0%
SVM	Training	100,0%	0,0%	-	100,0%	0,0%	0,0%
5 V IVI	Vorhersage	80,0%	20,0%	-	80,0%	20,0%	0,0%
SVM+SIMCA	Training	99,7%	0,0%	0,3%	100,0%	0,0%	0,0%
	Vorhersage	37,3%	16,7%	46,0%	34,0%	18,0%	48,0%

Tabelle 5-17: Modell- und Vorhersagegenauigkeiten für die verschiedenen Vorhersagemodelle der targeted LC-MS Messungen.

Die Ergebnisse der Einzelwertbestimmung bzw. der Dreifachbestimmung weichen – wenn auch nur geringfügig - voneinander ab, wobei sie bei der Gesamtbetrachtung einer Probe teilweise zu Gunsten und teilweise zu Ungunsten des Anteils an korrekt klassifizierten Proben verschoben sind. Daher ist eine Dreifachbestimmung unbedingt vorzuziehen, da bei nur einfach gemessenen Proben nicht ausgeschlossen werden kann, dass diese fälschlicherweise falsch klassifiziert wurden.

Für die PCA-LDA wurde der Quadratic Algorithmus bei 17 Hauptkomponenten als ideale Parameter bestimmt, wodurch eine annähernd identische Modellgenauigkeit von 98,3 % erhalten werden konnte im Vergleich zur non-targeted Messung (97,4 %). Allerdings ist dieses Modell (noch weniger als bei non-targeted) nicht in der Lage neue Proben gut vorherzusagen. Mit einer Vorhersagegenauigkeit von lediglich 24,0 % liegt auch hier die Vermutung nahe, dass die PCA-LDA für den vorliegenden Datensatz zu starkem Overfitting neigt, wodurch zwar der Training-Datensatz sehr gut reproduziert werden kann, neue Proben allerdings nur sehr schlecht in dieses Modell passen. Die Tatsache, dass sich dieser Effekt im Vergleich zur non-targeted Messung noch einmal verstärkt hat, könnte zum einen an der Reduktion auf 19 Markersubstanzen liegen und zum anderen an der Verdreifachung der Probendaten ohne die Probenmenge an sich zu vergrößern, wodurch jede einzelne Probe überrepräsentiert werden könnte. Dadurch könnte ein Zusammenhang vorgetäuscht werden, welcher das Modell der PCA-LDA überoptimistisch im Auffinden von vermeintlichen Zusammenhängen agieren lässt, sich jedoch nicht replizieren lässt in Bezug auf neue Proben.<sup>167</sup> Es konnte allerdings wie bereits unter Abschnitt 5.4.1 beschrieben keine Kreuzvalidierung zur Beurteilung dieses Effekts durchgeführt werden.

Ein aus der Dreifachbestimmung resultierendes Overfitting konnte für das SVM nicht beobachtet werden, da für das optimale Modell (SVM Typ: C-SVC, Kernel Typ: Polynomial 4. Grades, Offset: 3, C: 1, Gamma: 0,1) sowohl für die Modellgenauigkeit (targeted und non-targeted: 100 als auch für die Vorhersagegenauigkeit (targeted: 80 %; non-targeted: 90 %) %) ein identisches Resultat bzw. nur ein geringfügig niedrigerer Wert erhalten werden konnte als bei der non-targeted Messung - dort allerdings unter Verwendung der 225 *Features*. Im Vergleich zur non-targeted Messung bei Verwendung der 20 Markersubstanzen (90,3 % Modellgenauigkeit) sind die Resultate der targeted Messung bei Verwendung der 19 Markersubstanzen sogar noch besser. Ein Versuch, die SVM anhand von Hauptkomponenten einer vorher durchgeführten PCA zu berechnen um durch die Dimensionsreduzierung bessere Ergebnisse zu erzielen (ähnlich der PCA-LDA), war nicht zweckmäßig, da identische Werte für die Modell-, Validierungs- und Vorhersagegenauigkeit erhalten wurden. Dies verdeutlicht die Fähigkeiten der SVM bei komplexen Datensätzen, welche durch deren Algorithmus bereits die optimalen Informationen aus den vorhandenen Daten extrahieren kann, sodass eine vorherige Dimensionsreduktion nicht notwendig ist.

Die Berechnung der Validierungsgenauigkeit erfolgte über randomisierte Segmente, wobei unterschiedliche Segment-Größen getestet wurden (siehe Abbildung 5-25). Dabei stellte sich heraus, dass die Validierungsgenauigkeit selbst bei der Verwendung von lediglich zwei Segmenten (je weniger Segmente, desto schlechter die Validierungsgenauigkeit, vgl. Abschnitt 3.4.2) 97 % erreichte, was eine extreme Robustheit des zugrundeliegenden Modells beweist. Aus diesem Grund kann ein Overfitting des SVM-Modells ausgeschlossen werden, sodass die Anzahl der Replikate allein nicht ausschlaggebend für Overfitting-Effekte beim PCA-LDA zu sein scheint. Vielmehr ist das PCA-LDA Verfahren an sich möglicherweise nicht ausreichend potent um die verwendeten Daten hinreichend zu beschreiben bzw. zu diskriminieren. Im Gegensatz dazu liegen die Vorteile der SVM vor allem in deren



Abbildung 5-25: Validierungsgenauigkeit der Kreuzvalidierung des targeted SVM-C-Modells bei unterschiedlicher Anzahl an Segmenten.

Fähigkeiten bei komplexen biologischen Matrices begründet, wodurch dieses Verfahren anderen, klassischen multivariaten Methoden häufig überlegen ist. Durch die Transformation der Daten in höherdimensionale Räume ist der Algorithmus in der Lage mit Hilfe der unterschiedlichen Kernel Funktionen optimal separierende Hyperebenen zu definieren, die die verschiedenen Gruppen selbst in komplexen Fällen bei überlappenden und/oder nicht-linearen Originaldaten ideal voneinander trennen können und somit eine Diskriminierung ermöglichen.<sup>146, 191</sup>

Da sich sowohl für die PCA-LDA als auch die SVM die Parameter zur Berechnung des optimalen Modells für die targeted bzw. non-targeted Analyse unterscheiden, ist zu empfehlen, dass für alle folgenden Modellberechnungen stets individuell die idealen Parameter bestimmt werden.

Auch für die targeted Messung wurde neben den herkömmlichen multivariaten Verfahren das MCIM angewendet, obwohl es in diesem Fall keinen Vorteil durch Variablenreduktion beinhaltet. Nichtsdestotrotz bietet es die Möglichkeit die Probenvorhersage ohne komplizierte Algorithmen und kostenpflichtige Software durchzuführen. Da im Gegensatz zur non-targeted Messung Spanien Teil der Untersuchungen war und  $\gamma$ -Tocopherol nicht quantifiziert werden konnte, musste das Modell entsprechend modifiziert werden. Dazu wurden die Unterschiede der verschiedenen Länderpaare auf Basis der absoluten Metabolit-Gehalte mittels T-Test auf deren Signifikanz überprüft und jeweils die drei Metabolite mit den niedrigsten p-values

verwendet (siehe Abschnitt 8.2 Tabelle 8-14). Die restliche Vorgehensweise blieb unverändert.

Für die Modellgenauigkeit wurden dabei mit 73,3 % ein ähnlicher Wert erhalten wie bei der non-targeted Analyse (72,8%). Allerdings liegt die Vorhersagegenauigkeit – analog zu den Beobachtungen bei der PCA-LDA – mit 58,0 % deutlich unter der bisherigen Rate von 77,5 %. Möglicherweise wirkt sich hier ein ähnlicher Overfitting-Effekt auf das Modell aus. Dies könnte vor allem darin begründet sein, dass die Größe der Konfidenzintervalle abhängig von der Probenmenge ist. Da sich die Probenanzahl durch die Dreifachbestimmung vergrößert, die Metabolit-Gehalte aber annähernd gleichbleiben, steigt die vermeintliche statistische Sicherheit. Im Umkehrschluss verkleinern sich dadurch die Konfidenzintervalle. Da sich dies bei einer unterschiedlichen Anzahl an Proben (hier z.B. sechs spanische aber 22 italienische Proben) unterschiedlich stark auswirkt, wird das Modell dadurch negativ beeinflusst. Darüber hinaus besitzt Spanien aufgrund der geringen Probenanzahl sehr große Konfidenzintervalle, wodurch viele falsch positive Ergebnisse verursacht werden, wenn die Klassifizierung im ersten Schritt nicht eindeutig war und im zweiten Schritt die Anzahl an Metaboliten mit in den jeweiligen Konfidenzintervallen liegenden Gehalten entscheidend ist. Daher sind 57 % der Vorhersagen, die nicht im ersten Schritt erfolgten, fälschlicherweise Spanien zugeordnet worden, während 39 % weiterhin nicht-klassifiziert blieben und die restlichen 4 % einem anderen falschen Land zugeordnet wurden. Dies spiegelt die bereits mehrfach beschriebene Problematik des geringen Probenumfangs wider und verdeutlicht die Notwendigkeit einer größeren Datengrundlage, anhand derer dieses Modell möglicherweise besser funktionieren würde. Aus diesem Grund wurde zusätzlich noch einmal eine reduzierte MCIM Berechnung ohne Beteiligung der spanischen Proben durchgeführt, die erwartungsgemäß besser ausfiel, sowohl im Fall des Training-Datensatzes (mit Spanien: 73,3 % korrekt bei 24,2 % falsch; ohne Spanien: 77,4 % korrekt bei 22,1 % falsch) als auch des Vorhersage-Datensatzes (mit Spanien: 58,0 % korrekt bei 38,0 % falsch; ohne Spanien: 67,3 % korrekt bei 30,6 % falsch).

Für das SIMCA Modell wurden auch hier für nur 25 % der Proben diskrete Einfacklassifizierungen erhalten, während die restlichen Proben durch Mehrfachklassifizierungen nicht eindeutig zugeordnet werden konnten. Da das Ergebnis jedoch bei keiner der authentischen Proben falsch negativ ausfiel, wurde erneut die Kombination aus PCA-LDA, SVM bzw. MCIM mit der SIMCA als Kontrollmodell in Betracht gezogen. Dadurch konnte der Anteil an falsch klassifizierten Proben im Training-Datensatz für das MCIM sowie SVM um 53 % bzw. 10 % verringert werden, während der Anteil bei der PCA-LDA unverändert blieb. Der Korrektureffekt für die Vorhersage-Datensätze fiel hingegen deutlich schlechter aus, da bereits 52 % der nicht-authentischen Proben mit dem SIMCA-Modell falsch negativ vorhergesagt wurden. Zwar wurde in einigen Fällen der Anteil an falsch vorhergesagten Proben zugunsten nicht klassifizierter verändert, jedoch sank auch die Anzahl der korrekt vorhergesagten Proben drastisch, sodass bei der SVM als bestes Ergebnis nur noch 34 % der nicht-authentischen Proben richtig klassifiziert werden konnten. Aus diesem Grund ist die Kombination eines Vorhersagemodells mit der SIMCA, zumindest für die vorliegende Fragestellung, nicht zielführend. Es müsste jedoch überprüft werden, ob es sich bei den erhaltenen Resultaten im Rahmen der non-targeted Messungen um zufällige Ergebnisse handelt oder sich diese reproduzieren lassen und das Problem am Datensatz der targeted Messungen lokalisiert ist. Da das  $\gamma$ -Tocopherol als wichtige Diskriminanz-Variable für die targeted Auswertung nicht zur Verfügung stand, könnte deren Fehlen auch zu der Verschlechterung beigetragen haben.

Als Fazit lässt sich somit schlussfolgern, dass das SVM ohne SIMCA am besten zur Diskriminierung und Vorhersage von Haselnussproben im Zuge der targeted Analysen geeignet ist und sowohl für die Modell-, Validierungs- als auch Vorhersagegenauigkeit hervorragende Resultate erzielen kann. Obwohl es sich bei der targeted Messung im Vergleich zur non-targeted Applikation um eine verhältnismäßig einfache und kurze Analysenmethode handelt, ist sie in der Lage bessere Ergebnisse hinsichtlich der Vorhersage von unbekannten Haselnussproben zu erzielen. Die Anwendung des hochaufgelösten non-targeted UPLC-QTOF-MS stellt daher kein Vorteil gegenüber der einfacheren HPLC-QqQ-MS/MS Methode dar, weshalb letztere für zukünftige Messungen und die Implementierung in Routineanwendungen zu bevorzugen ist.

Bei genauerer Betrachtung der Ergebnisse fällt auf, dass vier der sechs georgischen Proben falsch klassifiziert wurden, was auf eine verhältnismäßig schlechte Eignung der Methode für dieses Land schließen lässt. Nichtsdestotrotz lagen 93 % der quantifizierten, georgischen Metabolit-Gehalte innerhalb der entsprechenden Bereiche der authentischen Proben. Dies demonstriert, dass keine großen Abweichungen in den Metabolit-Profilen an sich existieren, sondern die Probleme vielmehr durch verschwommene Grenzen verursacht werden. Wegen der teilweise starken Überschneidungen der Metabolit-Gehalte der verschiedenen Länder werden die Grenzen in manchen Fällen unscharf und eine korrekte Vorhersage schwierig. In Bezug auf die französischen Proben konnte die eine nicht-authentische Probe jedoch richtig vorhergesagt werden, obwohl sich das Metabolit-Profil relativ stark von dem der authentischen Haselnüsse unterschied, was bereits in Abbildung 5-23 visuell auffiel. Da das metabolische Profil der französischen Haselnüsse allerdings verhältnismäßig stark von denen der anderen Länder abwich, kann eine höhere Variation für dennoch korrekte Klassifizierungen toleriert werden als bspw. bei georgischen Proben.

Nichtsdestotrotz besteht die Möglichkeit, dass der Unterschied zwischen nicht-authentischen und authentischen Proben entweder durch natürliche, biologische Varianzen oder wirtschaftliche Faktoren (z.B. Lagerung) verursacht wurden, wodurch eine schlechte Vorhersagbarkeit resultieren kann. Dies verdeutlicht erneut, dass eine konsequente Weiterentwicklung des statistischen Modells notwendig ist. Durch Ergänzung weiterer Proben zur Modellerstellung kann die statistische Sicherheit dahingehend verbessert werden, dass durch größere Probenmengen natürliche Extrema besser erfasst werden können. Diese Hypothese konnte bestätigt werden, wenn nicht-authentische und authentische Proben gemeinsam zur Modellerstellung verwendet wurden, da dann sowohl die Modell- als auch die Validierungsgenauigkeit nahezu unverändert bei 100 % bzw. 97,4 % lagen. Es muss jedoch bei allen Schlussfolgerungen beachtet werden, dass die tatsächliche Herkunft der nicht-authentischen Proben per Definition nicht eindeutig verifiziert werden konnte, sodass stets die Möglichkeit einer Falschdeklaration besteht, wodurch die Ergebnisse der Vorhersagegenauigkeit entsprechend verfälscht werden könnten. Andererseits könnten bestimmte Proben auch in Zukunft mit dieser Analysenmethode aufgrund zu stark überlappender Bereiche nichtunterscheidbar bleiben. Insbesondere Haselnussproben aus Grenzregionen werden eine Herausforderung darstellen, da diese sehr ähnliche metabolische Profile aufweisen werden. Dies würde dann problematisch werden, wenn die Anbauregionen innerhalb eines Landes räumlich stärker getrennt sind als die Nähe zu denen eines anderen Landes. In diesem Fall wäre eine alternative Klassifizierung in ggfs. multinationale Regionen analytisch sinnvoller, jedoch spiegelt diese Vorgehensweise nicht die wirtschaftliche Realität der Deklarierung von Handelsware wider, welche üblicherweise nur nach Ländern erfolgt. Glücklicherweise liegen die Haselnussanbaugebiete am derzeitigen Haselnussmarkt nicht in unmittelbarer Nähre zu denen anderer Länder.

Zusätzlich wurden die fünf ebenfalls gemessenen, gerösteten und enthäuteten Haselnussproben ebenfalls mittels der targeted LC-MS Methode analysiert, um die Auswirkung des Röstprozesses auf die Vorhersagbarkeit des SVM-Modells abzuschätzen. Überraschenderweise konnten auch hier 80 % korrekt vorhergesagt werden bei nur einer Falsch-Vorhersage. Die Metabolit-Profile der richtig vorhergesagten Proben stimmten größtenteils mit den authentischen überein, da sie in 93 % der Fälle in den entsprechenden Bereichen der Metabolit-Gehalte der authentischen Proben lagen. Weitere Analysen müssen diese vorläufigen Rückschlüsse allerdings noch verifizieren. Die Ergebnisse demonstrieren jedoch, dass die prinzipielle Anwendbarkeit des SVM-Modells auch auf geröstete Haselnüsse gegeben ist und dass die 19 Markersubstanzen nicht durch die Temperatur während des Röstprozesses sowie des Blanchierens (Enthäuten) beeinflusst werden. Obwohl die hohen Temperaturen im Zuge der Röstung theoretisch zu Veränderungen des Haselnuss-Metaboloms führen können, stehen diese Resultate in Einklang mit einer Studie von Locatelli et al. (2015), welche aus einer chemometrischen Auswertung schlussfolgerten, dass die Fettsäurezusammensetzung von Haselnüssen nicht durch das Rösten verändert wird und daher die Leistung von geographischen Herkunftsdiskriminierungen unbeeinflusst bleibt. Ähnliche Ergebnisse publizierten Alasalvar et al. (2010), die nur geringfügige Veränderungen des Fettsäureprofils bei verschiedenen türkischen Varietäten nach der Röstung feststellten. Dies bestätigt die Schlussfolgerung, dass die Lipid Fraktion von Haselnüssen nicht sehr sensitiv auf Temperatureinflüsse während des Röstprozesse reagieren und das Vorhersagemodell möglicherweise unverändert direkt und ohne Anpassungen auch auf geröstete Haselnussproben angewendet werden kann

# 5.5 Messung von gemischten Haselnussproben

In der Realität des internationalen Warenhandels ist neben der illegalen Falschdeklarierung kompletter Chargen oder Packungen ebenfalls mit Beimischungen von günstigen Haselnüssen zu höherpreisigen Haselnüssen zur betrügerischen Gewinnmaximierung zu rechnen. Solche Beimischungen geringer Mengen stellen ein erhebliches Problem für die Authentizitätsprüfung dar, da sich die Detektion derartiger Verfälschungen viel schwieriger gestaltet. Aus diesem Grund stellt diese Vorgehensweise für kriminell motivierte Unternehmer einen hohen Anreiz dar, da die Wahrscheinlichkeit einer Aufdeckung sehr niedrig ist.

Um zu überprüfen, inwieweit eine Quantifizierung einer Beimischung von Haselnüssen mit den entwickelten analytischen Methoden realisierbar ist, wurden Regressionsmodelle mittels multivariater Statistik erstellt, mit denen die prozentualen Anteile der verschiedenen Länder in den zu untersuchenden Haselnussproben quantifiziert werden können. Dazu wurden virtuelle Mischungen anhand der nach Abschnitt 5.3.2 analysierten authentischen Haselnussproben berechnet (siehe Abschnitt 6.4.6). Im Gegensatz zur Messung von realen Mischungen (Einwaage der Haselnussproben in unterschiedlichen Anteilen mit anschließender Extraktion und Messung mittels targeted LC-MS) bietet dies den Vorteil, dass in sehr kurzer Zeit ein großer Datensatz mit realitätsnahen Werten erhalten werden kann. Diese Werte entsprechen in etwa den Werten, die eine Messung von realen Mischungen ergeben würde. Für die Modellerstellung wurden auf diese Weise virtuelle Mischungen in den Verhältnissen 100/0, 80/20, 60/40, 50/50, 40/60, 20/80, 0/100 für jede Probe (Mittelwert der Dreifachbestimmung) mit dem Mittelwert der verschiedenen Länder (Deutschland, Frankreich, Georgien, Italien, Spanien und Türkei) jeweils für die Erntejahre 2014 und 2015 berechnet. Anhand dieses Datensatzes wurden multivariate Regressionsmodelle zur Bestimmung des Anteils der verschiedenen Länder in unbekannten Proben berechnet. Anhand realer Mischungen von jeweils einer zufällig gewählten, repräsentativen, authentischen Haselnussprobe eines jeden Landes wurden diese Modelle anschließend auf ihre Leistungsfähigkeit überprüft. Dazu wurde für jedes Land je eine authentische Haselnussprobe verwendet und ebenfalls in den Verhältnissen 100/0, 80/20, 60/40, 50/50, 40/60, 20/80, 0/100 in jeder möglichen Länder-Kombination in Zwei-Komponenten-Mischungen eingewogen, extrahiert und mit der targeted LC-MS Methode analysiert. Zusätzlich wurden Drei-Komponenten-Mischungen im Verhältnis 33/33/33 erstellt und ebenfalls analysiert. Nach der Vorhersage der prozentualen Anteile wurden allerdings nur diejenigen Werte akzeptiert, deren Messunsicherheit kleiner als der berechnete Anteil war. Anschließend wurden die berechneten Werte so normiert, damit die Summe der Anteile 100 % ergab. Als Regressionsmodelle wurden PLS-R, MLR und SVM-R getestet. Die Ergebnisse der Modellerstellung, -validierung und -anwendung auf reale Mischproben sind in Tabelle 5-18 zusammengefasst.

Tabelle 5-18: Vergleich der Regressionsmodelle mit RMSE und R<sup>2</sup> der Kalibrierung (Kal) und Validierung (Val), mittlere prozentuale Abweichung vom wahren Wert ( $A_{proz}$ ), mittlere absolute Abweichung vom wahren Wert ( $A_{abs}$ ), Anteil der wahren Werte innerhalb des Bereiches der Messunsicherheit ( $W_{Mu}$ ) und Anteil der korrekt vorhergesagten Proben ( $W_{korr}$ ; Werte aller Länder innerhalb des Bereiches der Messunsicherheit).

	RM	RMSE		R <sup>2</sup>		Aabs	WMu	Wkorr	
	Kal	Val	Kal	Val					
MLR	12,28 %	12,36 %	0,7927	0,7899	11,23 %	± 67,41 %	65,18 %	16,07 %	
PLS-R	12,60 %	12,67 %	0,7817	0,7794	10,79 %	$\pm$ 64,77 %	68,72 %	18,46 %	
	DE: 3,95 %	DE: 6,09 %	DE: 0,9786	DE: 0,9490					
	ES: 4,45 %	ES: 5,64 %	ES: 0,9622	ES: 0,9394					
SVM D	FR: 6,37 %	FR: 8,71 %	FR: 0,9560	FR: 0,9174	12 00 0/		61 50 0/	17 86 0/	
SVIVI-K	GE: 4,37 %	GE: 5,54 %	GE: 0,9654	GE: 0,9445	15,00 %	$3,00\% \pm 77,98\%$	± //,98 % 04,38 %	17,00 %	
	IT: 7,69 %	IT: 9,60 %	IT: 0,9079	IT: 0,8560					
	TR: 3,68 %	TR: 4,81 %	TR: 0,9769	TR: 0,9606					

Für die PLS-R wurde der NIPALS Algorithmus verwendet mit einer optimalen Anzahl an Faktoren von 12. Die Validierung des Regressionsmodells erfolgte via Kreuzvalidierung mit 10 Segmenten. Es wurde ein mäßiger Regressionskoeffizient von 0,7817 (0,7794 Validierung) erhalten bei einem Fehler (engl.: *root-mean-square error;* RMSE) von 12,60 % (12,67 % Validierung).

Im Fall der MLR wurde ein Signifikanz-Level von  $\alpha = 0,05$  zugrunde gelegt. Die Validierung des Regressionsmodells erfolgte via *Leverage Correction*. Es wurde ein ähnlich mäßiger Regressionskoeffizient von 0,7927 (0,7899 Validierung) bei einem RMSE von 12,28 % (12,36 % Validierung) erhalten.

Beim SVM-R wurden verschiedene SVM Typen (Nu-SVR und Epsilon-SVR), Kernel Typen (Linear, Polynomial 2. bis 4. Grades, Radial Basis Function und Sigmoid) mit jeweils entsprechenden Nu-, Gamma-, C- und Offset-Werten für Nu-SVR bzw. Epsilon-, Gamma-, C- und Offset-Werten für Epsilon-SVR getestet. Die optimalen Werte für Nu (0-1), Epsilon (0-1), C (0,01-100), Gamma (0,01-100) und Offset (0-4) wurden semi-automatisch über *die Grid Search Funktion* bestimmt. Daraus ergab sich, dass die besten Ergebnisse durch das Epsilon-SVR mit Radial Basis Function Kernel Typ bei einem C-Wert von 10, Gamma-Wert von 1 und Epsilon-Wert von 0,1 erhalten werden konnten. Bei der SVM-R musste für jedes Herkunftsland ein separates Modell erstellt werden. Die Validierung der Regressionsmodelle erfolgte jeweils via Kreuzvalidierung mit einer Segmentgröße von 10. Die Regressionsmodelle le erzielten im Vergleich zur MLR und PLS-R deutlich bessere Regressionskoeffizienten von 0,9079 - 0,9786 (0,8560 – 0,9606 Validierung) und geringere RMSE-Werte von 3,68 % - 7,69 % (4,81 % - 9,60 % Validierung).

Den Bewertungskriterien der multivariaten Regression ( $R^2$  und RMSE) zufolge sollte die SVM-R somit das beste Modell darstellen, welches einen akzeptablen linearen Zusammenhang zwischen Metabolit-Gehalten und Länderanteilen beschreiben kann. Nichtsdestotrotz zeigte die Anwendung der zuvor erstellten Modelle zur Vorhersage realer Mischproben, dass die SVM-R am wenigsten geeignet war. Die besten Resultate konnten mittels PLS-R erzielt werden. Die mittels PLS-R berechneten Werte für die realen Mischproben sind in Abschnitt 8.2 Tabelle 8-13 dargestellt. Während die mittlere Abweichung der vorhergesagten Anteile von den tatsächlichen Anteilen 10,79 % beträgt und eine gute Vorhersagerate suggeriert, liegt die summierte, absolute Abweichung der Werte je Probe bei  $\pm$  64,77 %. Unter Einbeziehung der Messunsicherheit, welche einen gewissen Vertrauensbereich um den

Mittelwert des vorhersagten Anteils angibt, deckten immerhin 68,72 % der Bereiche den wahren Wert ab. Allerdings konnten nur 12 Mischproben (18,46 %) vollständig korrekt vorhergesagt werden. Für diese Proben decken die Bereiche der Messunsicherheit für alle sechs Länder den wahren Wert ab. In den meisten anderen Fällen waren zumindest häufig die richtigen Länder beteiligt, auch wenn die Anteile meist nicht korrekt waren. Häufig waren allerdings die Tendenzen richtig. Hervorzuheben sind dabei die Resultate bei der Mischung von Haselnüssen aus der Türkei und Italien sowie Frankreich und Italien, bei welchen jeweils 4 der 5 Mischungen korrekt vorhergesagt werden konnten.

Unter Berücksichtigung der Tatsache, dass die Messunsicherheiten im Durchschnitt ± 18 % betrugen und die Richtigkeit der Vorhersagen auch nur als mäßig einzustufen ist, lässt sich schlussfolgern, dass mit den derzeitigen Modellen keine präzise Vorhersage der Länder-Anteile in Mischungen möglich ist. Die Schwierigkeiten der richtigen Vorhersage der Länderanteile liegen wahrscheinlich in den teilweise stark überlappenden Metabolit-Gehalten zwischen unterschiedlichen Ländern begründet, welche bereits bei der Klassifizierung (siehe Abschnitt 5.4) zu Problemen führten. Aus diesem Grund gestaltet sich die Berechnung von diskreten prozentualen Anteilen extrem schwierig und erscheint - zumindest in Hinsicht auf die vorliegende Applikation – als nicht möglich. Zwar werden größtenteils Werte vorhergesagt, die in der Nähe des wahren Wertes liegen, jedoch sind die Ungenauigkeiten und die Fehler zu ausgeprägt, als dass zuverlässige Prognosen erstellt werden könnten, auf dessen Basis Beanstandungen, Reklamationen o.ä. formuliert werden könnten. Hinzu kommt, dass üblicherweise damit zu rechnen ist, dass gezielte, anteilige Verfälschungen eher in kleinen Größenordnungen erfolgen (ca. 1 - 10 %), um die Nachweisbarkeit für Kontrollstellen zu erschweren. Eine derartige Auflösung ist mit der vorliegenden Applikation jedoch nicht zu erreichen und erscheint auch in Hinsicht auf alternative Applikationen als äußert ambitioniert, da die natürliche biologische Varianz derart groß ist, dass eine solch genaue und hochaufgelöste Differenzierung extrem schwer zu erreichen sein wird.

Ein möglicher Ansatz zur Lösung dieses Problems ist eine alternative Probenvorbereitung. Durch eine stichprobenartige Entnahme mehrerer Haselnüsse einer Charge, welche anschließend jedoch einzeln und nicht im Gesamten homogenisiert und anschließend extrahiert werden, sodass aus einer repräsentativen Stichprobe mit x Nüssen auch jeweils x Extrakte erhalten werden, könnte mithilfe des gut funktionierenden SVM-Klassifizierungsmodells (siehe Abschnitt 5.4.2) jede Nuss einzeln einem Land zugeordnet werden. Anschließend werden darauf basierend die Anteile der Länder berechnet. Der Nachteil dieser Praxis ist allerdings - neben dem stark erhöhten Zeitaufwand - die Möglichkeit von Fehlklassifizierungen einzelner Nüssen aufgrund individuell auftretender Veränderungen z.B. durch Missbildung, Schimmelpilz- oder Bakterienbefall oder Druckbelastung während des Transports, welche bei der Homogenisierung eines entsprechend großen Stichprobenumfangs nicht ins Gewicht fallen würden.

# 5.6 Überprüfung der Lagerstabilität

Um die Stabilität der Markersubstanzen zu evaluieren wurden Lagerungsversuche durchgeführt. Dazu wurde die Haselnüsse vor der massenspektrometrischen Analyse wie gewohnt aufgearbeitet (siehe Abschnitt 6.1.2 und 6.1.3.2) und die unterschiedlichen Aufarbeitungsbzw. Zwischenprodukte bei unterschiedlichen Temperaturen für definierte Zeiträume gelagert und anschließend mittels der Methode NTNP-pos analysiert. Es handelte sich dabei um ungeschälte Haselnüsse, geschälte Haselnusskerne, Haselnussgranulat (Haselnusshomogenisat), Haselnusslyophilisat und den messfertigen Extrakt. Diese wurden jeweils für 1, 2, 4 und 8 Wochen bei -20 °C, -40 °C und -80 °C gelagert. Darüber hinaus wurde der Extrakt statt bei -80 °C bei 4 °C gelagert, um eine Aufbewahrung im Kühlschrank oder gekühlten Autosampler zu simulieren. Unmittelbar nach Ablauf der Lagerzeit wurden die jeweiligen Zwischenprodukte entsprechend der noch ausstehenden Aufarbeitungs- und Extraktionsschritte extrahiert. Aus Gründen der Praktikabilität und besseren Vergleichbarkeit der Ergebnisse wurden die erhaltenen Extrakte bis zur Messung bei -80 °C zwischengelagert, da davon ausgegangen wurde, dass in diesem Zustand Abbaureaktionen nahezu vollständig zum Erliegen gelangen. Die Proben wurden jeweils in Doppelbestimmung angesetzt. Zur Bewertung der Lagerstabilität der ganzen Haselnüsse wurde die Probe FR116 verwendet, während Homogenisat, Lyophilisat und Extrakt anhand der Probe IT001 untersucht wurden. Aus den resultierenden Datensätzen wurden die Peakflächen der 20 Markersubstanzen ausgelesen. Allerdings konnten drei Metabolite (y-Tocopherol, DG(16:0/16:1) und TG(15:0/16:0/18:1)) wegen zu niedriger Signalintensitäten nicht quantifiziert werden. Die restlichen 17 Markersubstanzen wurden für die Bewertung herangezogen, wobei ihre jeweiligen Ergebnisse sehr ähnlich waren. Es ist somit davon auszugehen, dass die drei nicht quantifizierbaren Metabolite keine wesentlichen Zusatzinformationen generieren würden, weshalb auf eine erneute Durchführung der Analyse verzichtet wurde. Die Ergebnisse der Auswertung sind in Tabelle 5-19 bzw. Tabelle 5-20 als jeweilige Mittelwerte aller Markersubstanzen inklusive Konfidenzintervall ( $\alpha = 0.05$ ) zusammenfassend dargestellt.

		1 Woche	2 Wochen	4 Wochen	8 Wochen	Durchschnitt
	-20 °C	$5,80 \pm 4,36$	$-12,33 \pm 7,60$	$28,31 \pm 6,85$	$25,\!18\pm6,\!19$	$11,74 \pm 5,15$
Homogenisat	-40 °C	$15,\!40\pm 6,\!77$	$-6,42 \pm 6,09$	$30,04 \pm 8,14$	$16,46 \pm 7,01$	$13,\!87\pm5,\!65$
	-80 °C	$11,\!84 \pm 5,\!50$	$-9,17 \pm 7,41$	47,61 ± 11,81	$26,\!43 \pm 6,\!87$	$19,18 \pm 5,63$
	-20 °C	33,64 ± 10,38	$-8,61 \pm 5,79$	$38,33 \pm 10,31$	$14,04 \pm 8,86$	$19,35 \pm 7,07$
Lyophilisat	-40 °C	$14,\!64 \pm 6,\!81$	$2,\!50\pm6,\!35$	$32,\!68 \pm 8,\!45$	$26,75\pm9,32$	$19,14 \pm 6,28$
	-80 °C	$12,72 \pm 5,84$	$-17,43 \pm 6,87$	$28,76\pm9,55$	$30,52 \pm 6,99$	$13,\!64 \pm 5,\!27$
	4 °C	$48,\!55 \pm 19,\!00$	82,89 ± 25,61	$28,21 \pm 13,20$	$-12,30 \pm 6,73$	$36,83 \pm 16,14$
Extrakt	-20 °C	$-15,27 \pm 6,28$	$-11,08 \pm 9,32$	$17,10 \pm 9,26$	$96,53 \pm 28,44$	21,82 ± 11,21
	-40 °C	$56,31 \pm 16,09$	$43,\!89 \pm 15,\!88$	$18,\!97\pm9,\!36$	$78,17 \pm 23,34$	49,33 ± 15,79

Tabelle 5-19: Änderung der Peakflächen der Markersubstanzen im Verlauf der Lagerung für die Probenvorbereitungszwischen- und -endprodukte (Homogenisat, Lyophilisat und Extrakt) bei unterschiedlichen Temperaturen in Bezug auf die frische Probe.

Tabelle 5-20: Änderung der Peakflächen der Markersubstanzen im Verlauf der Lagerung für ganze
Haselnüsse mit bzw. ohne Schale bei unterschiedlichen Temperaturen in Bezug auf die frische Probe.

		1 Woche	2 Wochen	4 Wochen	8 Wochen	Durchschnitt
TT 1	-20 °C	$-17,64 \pm 4,08$	$-7,75 \pm 6,16$	$-5,58 \pm 6,60$	$-22,84 \pm 9,77$	$-13,45 \pm 5,02$
Haselnuss mit Schale	-40 °C	$-9,41 \pm 4,19$	$-15,46 \pm 4,57$	$-9,30 \pm 8,30$	$-10,98 \pm 5,58$	$-11,29 \pm 4,43$
Int Senate	-80 °C	$-14,65 \pm 5,81$	$-9,46 \pm 3,99$	$-17,75 \pm 4,27$	$-18,14 \pm 7,00$	$-15,00 \pm 4,35$
Haselnuss	-20 °C	$-21,78 \pm 6,18$	$-15,34 \pm 6,68$	$-15,35 \pm 5,78$	$-19,10 \pm 6,69$	-17,89 ± 5,44
ohne	-40 °C	$-17,82 \pm 6,37$	$-12,34 \pm 5,58$	$-16,42 \pm 6,22$	$-11,81 \pm 6,69$	$-14,60 \pm 4,55$
Schale	-80 °C	$-18,59 \pm 3,82$	$-19,87 \pm 9,12$	$-15,83 \pm 3,61$	$-16,12 \pm 4,96$	$-17,60 \pm 4,26$

Bei allen 17 Metaboliten war keine Abnahme des Gehalts beim Homogenisat, Lyophilisat und Extrakt festzustellen, während bei der geschälten oder ungeschälten ganzen Nuss 15 Metabolite eine Signalverringerung unabhängig von der Lagerungstemperatur erfuhren. Abbildung 5-26 und Abbildung 5-27 zeigen exemplarisch die gemessenen Peakflächen für TG(2:0/18:1/18:2) als repräsentativen Verlauf für nahezu alle Markersubstanzen. Lediglich TG(16:0/16:1/18:1) blieben TG(14:0/16:0/18:1) und bei allen Lagerungsformen und -temperaturen über den gesamten Zeitraum unverändert. Erwähnenswert ist jedoch, dass bei den verarbeiteten Haselnüssen (Homogenisat, Lyophilisat, Extrakt) für alle Metabolite durchgehend ein teilweise nicht unerheblicher Anstieg der Metabolit-Gehalte detektiert wurde. Dies ist mit hoher Wahrscheinlichkeit auf die allgemein eher schlechte Reproduzierbarkeit von QTOF-Geräten und eine verhältnismäßig große Varianz bei den Messungen zurückzuführen (vgl. Abschnitt 3.3.2.3). Da die Lagerstabilität der Haselnussproben allerdings im Vorfeld der Entwicklung einer targeted Methode evaluiert werden sollte, wurde diese Versuchsreiche anhand der Methode NTNP-pos durchgeführt. Insbesondere bei den



Abbildung 5-26: Exemplarische, repräsentative Darstellung der Peakflächen der Extrakte, Lyophilisate und Homogenisate bei unterschiedlichen Lagertemperaturen anhand TG(2:0/18:1/18:2), stellvertretend auch für die restlichen Markersubstanzen.



Abbildung 5-27: Exemplarische, repräsentative Darstellung der Peakflächen der ungeschälten und geschälten Haselnussproben bei unterschiedlichen Lagertemperaturen anhand TG(2:0/18:1/18:2), stellvertretend auch für die restlichen Markersubstanzen.

Messungen der frischen Haselnussproben führt ein sehr großer zufälliger Fehler zu einer starken Veränderung des Ausgangswertes, auf welchem die Berechnung der Signalveränderung der gelagerten Proben beruht. Würde man nur die frische Probe mit den jeweils größeren Signalen zur Berechnung verwenden, so ergäbe sich statt eines 24,62 %igen Anstiegs des Gesamt-Durchschnitts der Signaländerung der gelagerten Proben nun eine Verringerung von -2,96 %. Daher kann keine sichere Aussage über die absolute Veränderung des Metabolit-Gehalts in Relation zur frischen Probe getätigt werden. Unter Berücksichtigung der Tatsache, dass die verwendeten Haselnussproben sowieso nicht tatsächlich "frisch" im Sinne einer zeitnahen Untersuchung nach der Ernte vom Baum waren, sondern als Handelsproben bereits eine unbekannte Zeit vom Erzeuger, Zwischenhändler und/oder Einzelhändler gelagert wurden, erscheint eine solche Herangehensweise allerdings auch nicht zweckmäßig. Vielmehr sollte die Betrachtung des Verlaufes über die Lagerdauer als Rückschluss über die Auswirkung der Lagerungsform und -temperatur herangezogen werden. Der Verlauf der Signalintensitäten der verschiedenen Metabolite zeigte dabei im Wesentlichen einheitlich sowohl für ganze als auch gemahlene oder extrahierte Nüsse weder einen Abwärtstrend noch einen Aufwärtstrend, welche eine gerichtete Metabolit-Veränderung rechtfertigen würde. Die Werte fluktuierten in unterschiedlichem Maße mit unterschiedlichem Verlauf über die verschiedenen Zeitpunkte, sodass davon ausgegangen wird, dass es sich dabei um zufällige Schwankungen um einen Mittelwert handeln muss. Generell wiesen die Extrakt-Lagerungsproben dabei ein variableres Verhalten mit stärkeren Maxima/Minima auf. Da jedoch auch die Standardabweichungen der jeweiligen Extrakt-Proben im Gegensatz zu allen anderen Lagerungsformen größer ausfielen, kann davon ausgegangen werden, dass die starken Schwankungen auf eine große Messunsicherheit zurückzuführen sind. Wodurch diese letztendlich verursacht wurde, lässt sich aus den Daten allerdings nicht mit Sicherheit schlussfolgern. Da dieses Phänomen jedoch nur bei den Extrakten derart ausgeprägt ist, liegt die Ursache höchstwahrscheinlich nicht im analytischen Prozess.

Die Analysenergebnisse führen somit zu dem Rückschluss, dass die Haselnussproben unabhängig von Lagerungsform und -temperatur (-20 °C, -40 °C oder -80 °C) lagerstabil sind und ohne weiteres auch nach langer Zeit für Metabolomics-basierte Analysen der geographischen Herkunft mit den entwickelten Methoden genutzt werden können. Ebrahem (1992) zeigte in seinen Untersuchungen zur Lagerstabilität von diversen Haselnusssorten der Erntejahre 1986-1988 bereits, dass die Gehalte an Fettsäuren sowie  $\alpha$ -Tocopherol und Gesamt-Tocopherol bei der Lagerung über mehrere Monate nur geringfügig abnehmen und schlussfolgerte daher, dass Haselnüsse bereits ab einer Temperatur von < 5 °C für bis zu zwei Jahre lagerstabil sind. Unter anderen sind vermutlich die schützenden polyphenolischen Verbindungen der Cuticula für diese Langzeitstabilität verantwortlich. Darüber hinaus schützt der hohe natürliche Gehalt an Vitamin E (ca. 11,8 - 44,0 mg/100 g) die Lipide vor dem Ranzigwerden (Lipid-Peroxidation). Die Stabilität könnte allerdings im Einzelfall durch degradative Prozesse (z.B. einsetzender mikrobiologischen Verderb, mechanische Schäden oder Schädlingsbefall) irgendwann überschritten werden.

Auch Haselnussproben mit unterschiedlicher Lagerdauer können somit ohne weiteres für eine gemeinsame Analyse verwendet werden, da die als Markersubstanzen verwendeten Lipide sehr stabil sind. Die erzielten Ergebnisse der Forschungsarbeit in Bezug auf die entwickelten Metabolomics-Applikationen können daher als repräsentativ angesehen werden, obwohl die verwendeten Haselnussproben von den Projektpartnern zu unterschiedlichen Zeitpunkten geerntet, gelagert und versendet wurden. Die Schlussfolgerungen über die Stabilität der Markersubstanzen stehen darüber hinaus im Einklang mit dem Resultat der Analyse der gerösteten Haselnüsse, bei denen bereits eine Temperaturresistenz festgestellt wurde, wodurch die geographische Herkunft selbst bei gerösteten Haselnüssen zuverlässig vorhergesagt werden konnte (vgl. Abschnitt 5.4.2).

### 5.7 Schlussfolgerungen und zukünftige Herausforderungen

Die Ergebnisse der vorliegenden Forschungsarbeit demonstrieren die hervorragende Anwendbarkeit von LC-MS-basierten Metabolomics-Applikationen zur Bestimmung der geographischen Herkunft von Haselnüssen. Sie zeigen, dass mit den entwickelten Methoden unbekannte Proben zuverlässig vorhergesagt werden können. Die SVM-C war dabei am besten geeignet und erzielte eine 100 %ige Genauigkeit für die authentischen Proben und konnte 90 % der nicht-authentischen Proben korrekt klassifizieren. Eine exakte Vorhersage der prozentualen Länderanteile bei Mischproben erscheint anhand der derzeitigen Ergebnisse jedoch nicht realisierbar. Durch die Regressionsmodelle können lediglich Abschätzungen getroffen werden, wobei diese noch sehr fehleranfällig sind. Darüber hinaus reicht die Leistungsfähigkeit der Modelle nach derzeitigem Stand nicht aus, die regionale geographische Herkunft der Haselnussproben hochaufgelöst vorherzusagen. Eine Unterscheidung innerhalb eines Herkunftslandes nach Anbauregion ist somit nicht möglich. Durch eine Kombination dieser Applikation mit anderen apparativen Methoden könnten diese Punkte möglicherweise dennoch realisiert bzw. zumindest weiter verbessert werden. Als vielversprechender Ansatz kann in diesem Zusammenhang die Kombination aus Massenspektrometrie und NMR-Spektroskopie gesehen werden. Durch das sogenannte Data-Fusion werden Informationsinhalte kombiniert, welche mit einer einzigen Applikation alleine nicht abgebildet werden können. Der Nachteil dieser Vorgehensweise ist jedoch die sehr arbeits-, rechen- und zeitaufwendige Durchführung, welche für Routineanalytik Anwendungen eher unpraktisch ist.

Sowohl die Klassifizierungs- als auch Regressionsmodelle haben gemein, dass die statistischen Modelle durch weitere Proben weiter verbessert werden müssen. Darüber hinaus sollten zusätzliche Herkunftsländer implementiert werden, die die kommerzielle Marktsituation widerspiegeln. Hierzu zählen derzeit insbesondere Aserbaidschan und die USA. Zwar



Abbildung 5-28: Metabolit-Gehalte (Centered und Scaled) der authentischen Proben aus Aserbaidschan, USA und Kanada der targeted LC-MS Messung; siehe dazu im Vergleich die anderen Länder in Abbildung 5-22, Abschnitt 5.3.2.

relativ einfach in entsprechende Modelle integriert werden könnten, ohne die Vorhersageeffizienz deutlich zu verringern. Die Metabolit-Gehalte der jeweiligen Markersubstanzen bzw. die daraus resultierenden "Mini-Fingerprints" weisen verhältnismäßig große Unterschiede zu den bestehenden Ländern auf (siehe Abbildung 5-28).

Die statistischen Modelle in ihrer jetzigen Form stellen nur eine Momentaufnahme des metabolischen Fingerabdrucks der Haselnussproben dar und müssen durch kontinuierliche Messung von Haselnussproben weiterer Erntejahre stetig überarbeitet werden. Insbesondere natürliche klimatische Variationen, aber auch Änderungen in Anbau- und Erntepraktiken sowie der kommerzielle Ausbau der Haselnussproduktion in größeren oder neuen Anbaugebieten und die Verwendung anderer Sorten stellen die Vorhersagemodelle vor eine dauerhafte Herausforderung, sodass regelmäßige Aktualisierungen notwendig sind. Als Beispiel seien an dieser Stelle drei Extremfälle näher beschrieben:

- seit einigen Jahren betreibt der weltweit größte Haselnussverarbeiter Ferrero eine aggressive Expansionspolitik und treibt die Erschließung weiterer kommerzieller Anbaugebiete gezielt voran. Unter anderem gründeten sie dafür die Firmen AgriChile (seit 1991), AgriSudafrica (seit 2009) und AgriAustralia (seit 2011), welche lokale Farmer durch Beratungstätigkeiten unterstützen und Forschungsprojekte fördern.<sup>32</sup> Die Erntemenge ist seitdem stetig steigend, sodass mittel- oder langfristig mit einer Verschiebung des globalen Haselnussmarktes zu rechnen ist.<sup>16, 246</sup> Diese Länder könnten dadurch stärker in den Vordergrund treten und bei entsprechend guter Qualität und günstigen Preisen andere traditionelle Produktionsländer aus dem Markt verdrängen. Daher sollten auch diese Länder mittel- oder langfristig in die entsprechenden Modelle mit einbezogen werden. Aufgrund der im Verhältnis zu den traditionellen Anbauländern großen geographischen Entfernung ist allerdings auch mit entsprechend ausgeprägten Unterschieden zu rechnen, sodass eine gute Implementierbarkeit als wahrscheinlich erscheint.
- 2014 traten in der Türkei erhebliche Ernteeinbrüche (ca. 35 %) auf, da ein ungewöhnlich strenger Frost zusammen mit starken Hagelschauern Ende März viele der empfindlichen Blüten zerstörte und zusätzlich eine anormale Trockenheit große Teile des Landes heimsuchte.<sup>247</sup> Neben den Auswirkungen auf die Haselnusspreise und Exportmengen führten diese Phänomene mit großer Wahrscheinlichkeit auch zu einer Veränderung des Metaboloms der geernteten Haselnüsse im Vergleich zu anderen Erntejahren.

 seit einigen Jahren findet in den USA ein starker Wandel im Haselnussanbau aufgrund der aggressiven Pilzkrankheit "Eastern Filbert Blight" (verursacht durch *Anisogramma anomala*) statt, welche regelmäßig ganze Ernten vernichtet und nicht effektiv bekämpft werden kann.<sup>17</sup> Daher existieren dort starke Forschungsbestrebungen zur Selektion von resistenten Haselnusssorten oder -hybriden durch gezielte Züchtungsversuche. Die bisherigen dort typischerweise kommerziell angebauten Haselnussorten (v.a. Barcelona, Lewis, Ennis und Butler) werden daher mittelfristig durch resistente Sorten ersetzt werden, um eine regelmäßige Ernte garantieren zu können.<sup>248, 249</sup>

Das Ziel der geographischen Herkunftsbestimmung von Haselnüssen durch Metabolomics-Applikationen sollte daher im Sinne eines guten Qualitätsmanagements ein kontinuierlicher Verbesserungsprozess sein, welcher durch fortwährende Analysen möglichst vieler Haselnussproben und der anschließenden Zusammenführung generierter Datensätze aus diversen Laboratorien zu einer gemeinsamen, globalen Datenbank (z.B. *MetaboLights*<sup>250</sup>) führen sollte. Das Prinzip des *Data-Sharings* stellt einen Grundsatz des Metabolomics-Gedanken dar und ist zunehmend als Voraussetzung für die Nutzung eines vernetzten Wissenspools als objektive Grundlage für leistungsfähige Analysenergebnisse zu verstehen.<sup>251, 252</sup>

Aufgrund der vielen Störfaktoren und der sich teilweise jährlich neu ergebenden Herausforderungen ist eine Vorhersage der geographsichen Herkunft mit 100 %iger Sicherheit nicht realisierbar, sodass die juristische Standhaftigkeit der Analysenergebnisse nach derzeitigem Stand fraglich bleibt. Vielmehr sollte bei auffälligen chemischen Analysen zur Untersuchung der geographischen Herkunft die bislang durchgeführten und sehr zeitaufwändigen Dokumentenprüfungen angeschlossen werden, um einen etwaigen Ausgangsverdacht zu bestätigen. Insofern sollten die hier dargestellten Methoden eher als Screening-Methoden und weniger als gerichtsfeste Einzelmethoden verstanden werden. Die Kombination aus Analysenergebnis mit anschließender Dokumentenprüfung stellt dementsprechend nicht nur eine sinnvolle Ergänzung der bisherigen Beweislage dar, sondern zusätzlich eine Verringerung des zeitlichen Aufwands, da die Dokumentenprüfung nur bei auffälligen Proben durchgeführt würde.

Eine weitere Schlussfolgerung dieser Forschungsarbeit ist in Bezug auf die Bewertungskriterien der Leistungsfähigkeit von multivariaten statistischen Vorhersagemodellen zu treffen.

Während derartige Modelle üblicherweise ausschließlich per Kreuzvalidierung beurteilt werden und bei einer entsprechend guten Validierungsgenauigkeit davon ausgegangen wird, dass es sich um ein leistungsfähiges und robustes Vorhersagemodell handeln muss, bleibt die tatsächliche Anwendbarkeit fraglich. Die Ergebnisse der non-targeted und targeted Messungen demonstrieren diesbezüglich, dass diese Vorgehensweise allein nicht ausreichend ist (insbesondere bei den Regressionsmodellen, bei denen das SVM-R als vermeintlich bestes Modell gemäß Kalibrierungs- und Validierungsergebnis die schlechteste Vorhersageleistung für reale Mischproben aufwies). Zum einen hängt das Validierungsergebnis maßgeblich von der Anzahl der verwendeten Segmente ab und nähert sich mit steigender Segmentzahl einem modellspezifischem Maximum an (siehe Abbildung 5-25, Abschnitt 5.4.2), wobei große Segmente (= kleine Segmentanzahl) aussagekräftiger für die Robustheit eines Modells sind (vgl. Abschnitt 3.4.2). Da die Anzahl allerdings prinzipiell in Abhängigkeit der Menge der zugrundeliegenden Proben individuell gewählt werden kann bzw. muss, ist es möglich die Validierungsgenauigkeit durch Erhöhung der Segmentanzahl gezielt zu verbessern. Ein sinnvoller Vergleich unterschiedlicher Modelle und/oder Studien auf Basis der Validierungsgenauigkeit ist daher nur begrenzt möglich. Zum anderen zeigen die Resultate der vorliegenden Forschungsarbeit, dass die Vorhersageleistung für unbekannte, nicht zur Modellerstellung verwendete (nicht-authentische) Proben von der Validierungsgenauigkeit teilweise erheblich abwich, sowohl bei den non-targeted als auch targeted Messungen. Daher ist es fraglich, ob die Begutachtung der Validierungsgenauigkeit allein ausreichend ist, um die Effizienz eines Vorhersagemodells zu beurteilen.

Eine alternative Vorgehensweise ist die Teilung der authentischen Proben vor der statistischen Analyse in einen Training-Datensatz und einen Test-Datensatz, wobei die Training-Daten ausschließlich zur Modellerstellung genutzt werden (inkl. Validierung). Die Test-Daten dienen im Anschluss als "unbekannte" Proben zur Überprüfung der Vorhersageleistung. Dadurch, dass ausschließlich authentischen Proben verwendet werden, resultiert aber gleichzeitig auch der wesentliche Nachteil dieser Vorgehensweise, welcher wahrscheinlich auch für die Unterschiede zwischen Validierungsgenauigkeit und Vorhersagegenauigkeit bei der herkömmlichen Vorgehensweise verantwortlich ist. Authentische Proben spiegeln nicht zwangsläufig marktreale Bedingungen wider, wie sie ggfs. bei handelsüblichen (nichtauthentischen) Realproben anzutreffen sind. Dies kann bspw. Ernte- und Nacherntepraktiken sowie Lagerungs- und Transportbedingungen betreffen. Aus diesem Grund besteht die Möglichkeit, dass der Test-Datensatz eine zu hohe Vorhersagegenauigkeit vortäuscht, welche bei tatsächlich unbekannten (nicht-authentischen) Proben nicht erreicht werden kann. Dabei ist selbstverständlich stets zu berücksichtigen, dass bei nicht-authentischen Proben ein Restrisiko der Falschdeklaration besteht, wodurch falsche Klassifizierungen erklärt werden könnten. Khakimov et al. (2015) berechneten die Falschklassifizierungs-Fehler anhand eines PLS-DA Klassifizierungsmodells bei unterschiedlichen Validierungspraktiken und demonstrierten dadurch ebenfalls, dass die Verwendung eines unabhängigen Test-Datensatzes (≙ nicht-authentischen Proben) die Effizienz des Modells am realistischen beschreiben kann, während bspw. die Kreuzvalidierung überoptimistische Ergebnisse vortäuscht.

Daher ist eine Kombination aus Kreuzvalidierung und Vorhersage nicht-authentischer Proben für die Bewertung eines Vorhersagemodells anzuraten, da dies mehrere Faktoren gleichzeitig berücksichtigt und dadurch bessere Rückschlüsse erlaubt. Die alleinige Anwendung der Kreuzvalidierung hingegen ist als unzureichend anzusehen.<sup>84</sup>

Darüber hinaus zeigen die erhaltenen Resultate eindeutig, dass die Wahl der chemometrischen Methode das Ergebnis der statistischen Auswertung maßgeblich beeinflusst und dass die Nutzung einer einzigen, favorisierten Methode nicht zielführend ist. Vielmehr sollten in der Praxis unterschiedliche Ansätze getestet und die letztendlich zu verwendende Methode sorgfältig ausgewählt werden um zu garantieren, dass das entsprechende Modell ideal zu der vorliegenden Fragestellung passt und die beste Vorhersageleistung ermöglicht.<sup>189</sup> Im vorliegenden Fall war bspw. die Kombination aus SVM-C und SIMCA für die Klassifikation in der non-targeted Applikation am besten geeignet, während die SVM-C ohne SIMCA bei der Klassifikation sowie die PLS-R bei der Regression der targeted Applikation zu den besten Ergebnissen führten. Gleichzeitig müssen auch die jeweiligen optimalen Modellparameter individuell gewählt werden, um die zugrundeliegenden Datenstruktur bestmöglich widergeben zu können. (z.B. linearer Kernel Typ bei SVM-C für non-targeted und polynomialer Kernel Typ bei SVM-C für targeted) Außerdem sollten die Modellparameter nach dem Hinzufügen weiterer Daten (z.B. zusätzlicher Länder oder weiterer Erntejahre) stets erneut überprüft werden.

# 6 Material und Methoden

Eine tabellarische Auflistung aller im Rahmen der Arbeit verwendeten Chemikalien (Tabelle 8-1 und Tabelle 8-2), Lösungen und Puffer (Tabelle 8-3), Arbeitsgeräte und Software (Tabelle 8-4) sowie Verbrauchsmaterialien (Tabelle 8-5) befindet sich im Anhang Abschnitt 8.1.

### 6.1 Proben

### 6.1.1 Probenmaterial

Es wurden insgesamt 266 Proben von rohen Haselnüssen aus neun verschiedenen Ländern akquiriert, von denen 207 als authentisch einzustufen waren und 59 (davon fünf geröstete) als nicht-authentisch (siehe Tabelle 6-1). Die Haselnüsse stammten aus den jeweils wirtschaftlich relevanten Hauptanbaugebieten (vgl. Abbildung 3-2, Abschnitt 3.1). Eine detaillierte Auflistung aller Proben inklusive Probenbezeichnung, Metadaten und Lieferanten ist in Abschnitt 8.2 Tabelle 8-15 für die authentischen bzw. Tabelle 8-16 für die nicht-authentischen Haselnüsse dargestellt. Authentisch bedeutete in diesem Zusammenhang, dass die geographische Herkunft der Probe zweifelsfrei durch Projektpartner garantiert werden konnte oder die Haselnüsse direkt von Bauern bezogen wurden, während es sich bei den nicht-authentischen Proben überwiegend um Handelsware für die Süßwarenindustrie oder aus dem Einzelhandel handelte, bei denen die deklarierte Herkunft nicht eindeutig nachverfolgt werden konnte. Die Proben umfassten entweder ca. 1000 g (mind. 200 g) rohe Haselnusskerne mit Haut (Testa) oder 1500 g (mind. 600 g) Haselnüsse mit Schale. Die Ernte- sowie

Nacherntebehandlung (Trocknen, Sortieren, ggfs. Knacken, Verpacken) wurden von den Projektpartnern unter marktrealen Bedingungen durchgeführt. Die authentischen Proben wurden danach unverzüglich unter handelsüblichen Lagerungsbedingungen versendet, während für die nicht-authentischen Proben keine Informationen über die Zwischenlagerungszeit vorlagen. Die Proben wurden unverzüglich nach Erhalt eingefroren und bei -40 °C bis zur Weiterverarbeitung gelagert.

Tabelle	6-1:	Anzahl	der	authentischen	und
nicht-au	thent	ischen	Has	elnussproben	pro
Herkun	ftslan	d.			

Land	authentisch		nic authe	ht- ntisch
	2014	2015	2014	2015
Aserbaidschan	2	1	1	1
Deutschland	24	11	-	-
Frankreich	60	55	1	-
Georgien	8	2	3	3
Italien	8	14	2	8
Kanada	1	-	-	-
Spanien	3	3	1	-
Türkei	4	10	12	25
USA	-	1	1	-
unbekannt*	-	-	1	-

\*nur zur Methodenentwicklung verwendet

### 6.1.2 Probenaufarbeitung

In-Shell Haselnüsse wurden zunächst mit Hilfe einer mechanischen Handknackmaschine von der Schale befreit. Nach Schockgefrieren der Haselnusskerne mittels Flüssigstickstoff (-196 °C) wurden max. 500 g zusammen mit Trockeneisgranulat (-78 °C) im Verhältnis 1:1 mit einer Messermühle für 20 s bei 1000 rpm und anschließend 4x 30 s bei 4000 rpm mit jeweils 5 s Pause zur Homogenisierung gemahlen. Bei größeren Probenmengen wurden die einzelnen Homogenisate vereinigt und gründlich durchmischt. 100 g des Mahlguts wurden unter Lichtausschluss bei 0,1 mbar für 24 h  $\pm$  1 h gefriergetrocknet. Die Lagerung der Lyophilisate erfolgte anschließend bei -20 °C.

#### 6.1.3 Extraktion

#### 6.1.3.1 Initialmethode zur Methodenoptimierung

50 mg  $\pm$  0,5 mg Probenlyophilisat wurden in ein eisgekühltes 2-mL-Reaktionsgefäß eingewogen und mit zwei Stahlkugeln, 1000 µL vorgekühlter Extraktionslösung (4 °C) aus Chloroform/Isopropanol (1/2, v/v) versetzt und 3x für 5 min bei 30 Hz mit 5 min Inkubationszeit zwischen den Mahlvorgängen in einer Kugelmühle (TissueLyser) homogenisiert. Die Suspension wurde anschließend bei 4 °C für 30 min bei 16.000 g zentrifugiert und der Überstand in ein eisgekühltes Braunglasvial überführt, fest verschlossen und bis zur Analyse bei -20 °C gelagert. Im weiteren Verlauf der Methodenentwicklung wurden einige der Extrakte nach der Zentrifugation mittels Spritzenvorsatzfilter (Rotilabo PTFE; Porengröße 0,45 µm) membranfiltriert.

#### 6.1.3.2 Extraktion der unpolaren Metabolite für die non-targeted Analyse

50 mg  $\pm$  0,5 mg Probenlyophilisat wurden in ein eisgekühltes 2-mL-Reaktionsgefäß eingewogen und mit zwei Stahlkugeln, 1000 µL vorgekühlter Extraktionslösung (4 °C) aus Chloroform/Isopropanol (1/2, v/v) + 0,01 % BHT versetzt und für 3 min bei 3 m/s in einer Kugelmühle (BeadRuptor 24) homogenisiert. Die Suspension wurde anschließend bei 4 °C für 10 min bei 16.000 g zentrifugiert und der Überstand nach Membranfiltration mittels Spritzenvorsatzfilter (Rotilabo PTFE; Porengröße 0,45 µm) in ein eisgekühltes Braunglasvial überführt, fest verschlossen und bis zur Analyse bei -20 °C gelagert.

### 6.1.3.3 Extraktion der unpolaren Metabolite für die targeted Analyse

Die Extraktion erfolgte analog zur Extraktion der unpolaren Metabolite für die non-targeted Analyse (vgl. Abschnitt 6.1.3.2) mit der Änderung, dass 990  $\mu$ L des Extraktionsmittels sowie 10  $\mu$ L des internen Standard Mix (IS) (siehe Abschnitt 6.2.3.2) aus fünf stabilisotopenmarkierten Substanzen [PC(15:0/18:1(d7)), PE(15:0/18:1(d7)), PC(18:1(d7)/0:0), DG(15:0/18:1(d7)) und TG(15:0/18:1(d7)/15:0)] zur Extraktion verwendet wurden.

### 6.1.3.4 Extraktion der polaren Metabolite für die non-targeted Analyse

50 mg ± 0,5 mg Probenlyophilisat wurden in ein eisgekühltes 2-mL-Reaktionsgefäß eingewogen und mit zwei Stahlkugeln und zunächst mit 667  $\mu$ L vorgekühltem Methanol (4 °C) zur Proteinpräzipitation versetzt und für 1 min bei 3 m/s in einer Kugelmühle (BeadRuptor 24) homogenisiert. Anschließend wurden 333  $\mu$ L vorgekühltes Wasser (4 °C) hinzugegeben und für weitere 2 min bei 3 m/s homogenisiert. Die Suspension wurde anschließend bei 4 °C für 10 min bei 16.000 g zentrifugiert und der klare Überstand in ein eisgekühltes Braunglasvial überführt, fest verschlossen und bis zur Analyse bei -20 °C gelagert.

### 6.1.3.5 Qualitätskontrolle

Zur Überprüfung der Stabilität des massenspektrometrischen Systems im Laufe der Messungen wurde aus den nach Abschnitt 6.1.3.2, 6.1.3.3 und 6.1.3.4 extrahierten Proben jeweils eine QC-Probe erstellt (QC-targeted, QC-NTNP bzw. QC-NTP). Dafür wurden gleiche Teile der Extrakte zufällig ausgewählter Proben DE005, DE016, FR009, FR046, GE002, GE005, IT001, IT006, TR001 und TR002 als repräsentative Mischung vereint (QC-NTNP, QC-NTP bzw. QC-targeted).

# 6.2 Stammlösungen und Multistandards

# 6.2.1 Markersubstanzen

Tabelle 6-2 zeigt die 20 Markersubstanzen, deren Identität durch eine geeignete Strukturaufklärung (vgl. Abschnitt 6.3.7) ermittelt wurde. 10 der Markersubstanzen konnten als kommerziell erhältliche Referenzstandards erworben werden (siehe Tabelle 8-2, Abschnitt 8.1).

Stoffklasse	Markersubstanz	Summenformel	Monoisotopische Masse [Da]
	DG(16:0/16:1)	$C_{35}H_{66}O_5$	566,4910
Diagulghyaarala	DG(16:0/18:1)*	C37H70O5	594,5223
Diacyigiyceiole	DG(18:2/18:2)*	C39H68O5	616,5067
	DG(18:1/18:1)*	$C_{39}H_{72}O_5$	620,5380
Lysophosphatidylcholine	PC(18:2/0:0)*	C <sub>26</sub> H <sub>50</sub> NO <sub>7</sub> P	519,3325
	PC(16:0/18:3)	$C_{42}H_{78}NO_8P$	755,5465
Dhaanhatidadahalina	PC(16:0/18:2)*	$C_{42}H_{80}NO_8P$	757,5622
Phosphaticytcholine	PC(18:2/18:2)*	$C_{44}H_{80}NO_8P$	781,5622
	PC(18:1/18:2)	$C_{44}H_{82}NO_8P$	783,5778
	PE(16:0/18:2)*	C <sub>39</sub> H <sub>74</sub> NO <sub>8</sub> P	715,5152
Phosphatidylethanolamine	PE(18:2/18:2)*	$C_{41}H_{74}NO_8P$	739,5152
	PE(18:1/18:1)*	$C_{41}H_{78}NO_8P$	743,5465
Tocopherole	γ-Tocopherol*	$C_{28}H_{48}O_2$	416,3654
	TG(2:0/18:2/18:2)	$C_{41}H_{70}O_6$	658,5172
	TG(2:0/18:1/18:2)	$C_{41}H_{72}O_6$	660,5329
	TG(14:0/16:0/18:1)	$C_{51}H_{96}O_{6}$	804,7207
Triacylglycerole	TG(15:0/16:0/18:1)	$C_{52}H_{98}O_{6}$	818,7363
	TG(16:0/16:1/18:1)	$C_{53}H_{98}O_{6}$	830,7363
	TG(17:1/18:1/18:2)	$C_{56}H_{100}O_{6}$	868,7520
	TG(18:2/18:2/18:3)	$C_{57}H_{96}O_{6}$	876,7207

Tabelle 6-2: Markersubstanzen	fiir die	Bestimmung der	r geogranhischen	Herkunft.
Tabelle 0-2. Marker substanzen	iui uic	Destimining de	geographischen	IICI KUIIII.

\*Referenzstandard vorhanden

### 6.2.2 Stammlösungen

Von den 10 Referenzstandards sowie ergänzend von zwei kommerziell erhältlichen Triacylglycerolen (TG(16:0/18:1/18:1) und TG(16:0/18:1/18:2)) wurden Stammlösungen zu je ca. 10 mM durch Lösen oder Verdünnen mit Chloroform und/oder Methanol erstellt. Zusätzlich wurden sechs stabilisotopenmarkierte, deuterierte Standardsubstanzen als interne Standards verwendet ( $\alpha$ -Tocopherol-D6, PC(18:1d7/0:0), DG(15:0/18:1d7), PE(15:0/18:1d7), PC(15:0/18:1d7), TG(15:0/18:1d7/15:0)), deren voreingestellte Konzentration je 1 mg/mL (0,5 mg/mL  $\alpha$ -Tocopherol-D6), gelöst in Chloroform betrug. Jeder dieser internen Standards repräsentiert dabei eine der sechs Stoffklassen der Markersubstanzen.

### 6.2.3 Multistandards

#### 6.2.3.1 Methodenentwicklung und -validierung

Für die Grundkalibrierung der Methodenvalidierung wurden 25 Multistandards durch Mischen unterschiedlicher Volumina der Stammlösungen (vgl. Abschnitt 6.2.2) mit Methanol verschiedene Konzentrationen im Konzentrationsbereich von 0,02 - 1000  $\mu$ M erstellt. Die

daraus resultierenden Konzentrationen der Standardlösungen sind in Tabelle 6-3 dargestellt. Für die Matrixkalibrierung wurde ebenfalls 25 Multistandards derselben Konzentrationen erstellt, in diesem Fall durch Mischen unterschiedlicher Volumina der Stammlösungen (vgl. Abschnitt 6.2.2) mit Methanol und einem definierten Volumen Haselnussmatrix (QC-targeted, siehe Abschnitt 6.1.3.5), sodass der resultierende Multistandard 50 % Matrix beinhaltete. Der Multistandard der Grundkalibrierung Val20 (je ca. 100  $\mu$ M) wurde zusätzlich im Rahmen der Methodenentwicklung verwendet.

Tabelle 6-3: Konzentration der Metabolite in den Multistandards der Methodenvalidierung in  $\mu$ M. (aus Klockmann 2017)

Standard	DG	DG	DG	РС	РС	РС	PE	PE	PE	γ-
Stanuaru	(16:0/18:1)	(18:1/18:1)	(18:2/18:2)	(18:2/0:0)	(16:0/18:2)	(18:2/18:2)	(16:0/18:2)	(18:1/18:1)	)(18:2/18:2)	Tocopherol
Val1	0,0168	0,0230	0,0194	0,0192	0,0200	0,0200	0,0215	0,0199	0,0200	0,0150
Val2	0,0336	0,0460	0,0389	0,0385	0,0400	0,0400	0,0430	0,0398	0,0400	0,0300
Val3	0,0504	0,0691	0,0583	0,0577	0,0600	0,0599	0,0645	0,0597	0,0600	0,0450
Val4	0,0672	0,0921	0,0778	0,0770	0,0799	0,0799	0,0859	0,0796	0,0800	0,0600
Val5	0,0840	0,115	0,0972	0,0962	0,100	0,100	0,107	0,0995	0,100	0,0750
Val6	0,168	0,230	0,194	0,192	0,200	0,200	0,215	0,199	0,200	0,150
Val7	0,336	0,460	0,389	0,385	0,400	0,400	0,430	0,398	0,400	0,300
Val8	0,504	0,691	0,583	0,577	0,600	0,599	0,645	0,597	0,600	0,450
Val9	0,672	0,921	0,778	0,770	0,799	0,799	0,859	0,796	0,800	0,600
Val10	0,840	1,15	0,972	0,962	1,00	1,00	1,07	0,995	1,00	0,750
Val11	1,68	2,30	1,95	1,92	2,00	2,00	2,15	1,99	2,00	1,50
Val12	3,36	4,60	3,89	3,85	4,00	4,00	4,30	3,98	4,00	3,00
Val13	5,04	6,91	5,84	5,77	6,00	5,99	6,45	5,97	6,00	4,50
Val14	6,72	9,21	7,78	7,70	7,99	7,99	8,60	7,96	8,00	6,00
Val15	8,40	11,5	9,72	9,62	10,0	9,99	10,7	9,95	10,0	7,50
Val16	16,8	23,0	19,4	19,2	20,0	20,0	21,5	19,9	20,0	15,0
Val17	33,6	46,0	38,9	38,5	40,0	40,0	43,0	39,8	40,0	30,0
Val18	50,4	69,1	58,3	57,7	60,0	59,9	64,5	59,7	60,0	45,0
Val19	67,2	92,1	77,8	77,0	79,9	79,9	85,9	79,6	80,0	60,0
Val20	84,0	115	97,2	96,2	99,9	99,9	107	99,5	100	75,0
Val21	-	230	211	-	-	-	-	-	-	-
Val22	-	461	406	-	-	-	-	-	-	-
Val23	-	691	648	-	-	-	-	-	-	-
Val24	-	921	843	-	-	-	-	-	-	-
Val25	-	1150	1090	-	-	-	-	-	-	-

Tabelle 6-3 (Fortsetzung): Konzentration der Metabolite in den Multistandards der Methodenvalidierung in μM. (aus Klockmann 2017)

Standard	PC	РС	PE	DG	TG	TG	TG
Standard	(18:1d7/0:0)	(15:0/18:1d7)	(15:0/18:1d7)	(15:0/18:1d7)	(15:0/18:1d7/15:0)	(16:0/18:1/18:1)	(16:0/18:1/18:2)
Val1	0,0237	0,0166	0,0180	0,0213	0,0157	0,0201	0,0212
Val2	0,0474	0,0332	0,0360	0,0426	0,0315	0,0401	0,0424
Val3	0,0710	0,0499	0,0540	0,0639	0,0472	0,0602	0,0636
Val4	0,0947	0,0665	0,0720	0,0852	0,0630	0,0802	0,0848
Val5	0,118	0,0831	0,0900	0,106	0,0787	0,100	0,106
Val6	0,237	0,166	0,180	0,213	0,157	0,201	0,212
Val7	0,474	0,332	0,360	0,426	0,315	0,401	0,424
Val8	0,710	0,499	0,540	0,639	0,472	0,602	0,636

Material und Methoden

Standard	РС	РС	PE	DG	TG	TG	TG
~	(18:1d7/0:0)	(15:0/18:1d7)	(15:0/18:1d7)	(15:0/18:1d7)	(15:0/18:1d7/15:0)	(16:0/18:1/18:1)	(16:0/18:1/18:2)
Val9	0,947	0,665	0,720	0,852	0,630	0,802	0,848
Val10	1,18	0,831	0,900	1,07	0,787	1,00	1,06
Val11	2,37	1,66	1,80	2,13	1,58	2,01	2,12
Val12	4,74	3,33	3,60	4,26	3,15	4,01	4,24
Val13	7,10	4,99	5,40	6,39	4,73	6,02	6,36
Val14	9,47	6,65	7,20	8,52	6,30	8,03	8,48
Val15	11,8	8,31	9,00	10,6	7,88	10,0	10,6
Val16	23,7	16,6	18,0	21,3	15,8	20,1	21,2
Val17	47,4	33,2	36,0	42,6	31,5	40,1	42,4
Val18	71,0	49,9	54,0	63,9	47,3	60,2	63,6
Val19	94,7	66,5	72,0	85,2	63,0	80,2	84,8
Val20	118	83,1	90,0	106	78,7	100	106
Val21	-	-	-	-	-	201	212
Val22	-	-	-	-	-	401	425
Val23	-	-	-	-	-	602	636
Val24	-	-	-	-	-	802	848
Val25	-	-	-	-	-	1000	1060

#### 6.2.3.2 Interner Standard Mix

Der interne Standard Mix (IS) wurde durch Mischen unterschiedlicher Volumina der sechs Stammlösungen der stabilisotopenmarkierten Substanzen (vgl. Abschnitt 6.2.2) mit Methanol erstellt. Die daraus resultierenden Konzentrationen im IS sind in Tabelle 6-4 dargestellt. Tabelle 6-4: Konzentration der deuteriertenStandards im internen Standard Mix.

	Konzentration [mM]
DG(15:0/18:1d7)	1,70
PC(18:1d7/0:0)	1,89
PC(15:0/18:1d7)	1,33
PE(15:0/18:1d7)	1,44
α-Tocopherol-D6	1,38
TG(15:0/18:1d7/15:0)	1,26

#### 6.2.3.3 Extrakt-Verdünnungsreihe

Zur Abschätzung des linearen Arbeitsbereiches der Markersubstanzen, für die keine Standardsubstanzen erhältlich waren, und zur Bestimmung der benötigten Verdünnung des Extraktes zur Messung im linearen Arbeitsbereich der validierten Markersubstanzen wurde eine Verdünnungsreihe eines repräsentativen Extraktes (QC-targeted; siehe Abschnitt 6.1.3.5) erstellt. Dieser wurde mit Methanol mit den folgenden Verdünnungsfaktoren verdünnt: 1,18; 1,43; 1,82; 2,50; 4,00; 10,00; 11,76; 14,29; 18,18; 25,00; 40,00; 100,00; 117,65; 142,86; 181,82; 250,00; 400,00; 1000,00.
#### 6.2.4 Blindwert

Als Blindwertproben wurde jeweils das frisch angesetzte Extraktionsmittel für die entsprechende LC-MS-Methode verwendet. Für NTNP-pos/NTNP-neg sowie für targeted: Chloroform/Isopropanol (1/2, v/v) + 0,01 % BHT und für NTP-pos/NTP-neg: Wasser/Methanol (1/2, v/v).

# 6.3 Non-Targeted LC-MS (UPLC-ESI-QTOF-MS)

# 6.3.1 LC-MS Methoden für die Messung der unpolaren Metabolite (NTNP-pos und NTNP-neg)

Die apparative Analytik der nach Abschnitt 6.1.3.2 erstellten unpolaren Extrakte erfolgte auf einer Dionex UltiMate 3000 UPLC, welche mit einem maXis 3G ESI-QTOF Massenspektrometer gekoppelt war. Die Analyse erfolgte sowohl im positiven Ionenmodus (NTNP-pos) als auch im negativen Ionenmodus (NTNP-neg). Nachfolgend sind die Methodenparameter (siehe Tabelle 6-5) der beiden Methoden zusammenfassend dargestellt.

Autosampler						
Temperatur		4 °C				
Injektionsvolumen		1 µL				
	Säulenofen					
Säule	Thermo Fischer,	Accucore RP-MS, 150	) x 2,1 mm, 2,6 µm			
Säulenofentemperatur	30 °C					
	Pumpe					
Dauer	25 min					
Flussrate	350 μL/min					
Eluent A	H <sub>2</sub> O + 10 mM AF (pH 3,5)					
Eluent B	CHCl <sub>3</sub> /IPA (1/2, v/v) + 10 mM AF (pH 3,5)					
	Zeit	% Eluent A	% Eluent B			
	0	35	65			
	2	35	65			
Cradient	5	25	75			
Grautent	15	0	100			
	21	0	100			
	22	35	65			
	25	35	65			
	Massenspektromet	ter				
Ionisationsmethode		ESI				
Ionisationsmodus	positiv oder negativ					
Massenbereich		60 – 1000 Da				
Spectra Rate		1 Hz				

Tabelle 6-5: Methodenparameter für die non-targeted LC-MS Methoden für unpolare Metabolite. (NTNP-pos und NTNP-neg)

Massenspektrometer (Fortsetzung)					
Valibration	Lockmasse:	Hexakis(2,2- difluoroethoxy)phosphazen			
Kanoration	Kalibrierlösung:	Natriumformiat-Cluster			
	Calibration Mode	Enhanced Quadratic			
	Segment 1:	24 min, divert valve to source			
Segmente	Segment 2 (Kalibrierung):	0,5 min, divert valve to waste			
	Segment 3:	0,5 min, divert valve to source			
	End Plate Offset:	+/-500 V			
	Capillary:	4500 V			
Ionenquelle	Nebulizer Gas:	4 bar			
•	Drying Gas:	9 L/min			
	Drying Temperature:	200 °C			
	Funnel 1 RF:	300 Vpp			
Transfer	isCID Energy:	0 eV			
	Multipole RF:	250 Vpp			
Quadmunal	Ion Energy:	4 eV			
Quadrupoi	Low Mass:	60 m/z			
	Collision Energy:	10 eV			
Collision Cell	Collision RF:	300 Vpp			
	Transfer Time:	60 µs			
	Pre Puls Storage:	5 μs			
nur bei Aufnahme von	Scan Mode:	MS/MS			
MS/MS-Spektren	Collision Energy:	40 eV			

Material und Methoden

# 6.3.2 LC-MS Methoden für die Messung der polaren Metabolite (NTP-pos und NTP-neg)

Die apparative Analytik der nach Abschnitt 6.1.3.3 erstellten polaren Extrakte erfolgte auf einer Dionex UltiMate 3000 UPLC, welche mit einem maXis 3G ESI-QTOF Massenspektrometer gekoppelt war. Die Analyse erfolgte sowohl im positiven Ionenmodus (NTP-pos) als auch im negativen Ionenmodus (NTP-neg). Nachfolgend sind die optimierten Methodenparameter (siehe Tabelle 6-6) der beiden Methoden zusammenfassend dargestellt. Im Zuge der Methodenoptimierung wurden folgende Parameter variiert: Eluent B, Gradient, Fließgeschwindigkeit und Säulenofentemperatur (vgl. Tabelle 5-5, Abschnitt 5.1.1.2.1).

Tabelle 6-6: Methodenparameter für die non-targeted LC-MS Methoden für polare Metabolite. (NTP-pos und NTP-neg)

Autosampler					
Temperatur		4 °C			
Injektionsvolumen		4 μL			
	Säulenofen				
Säule	Cogent, Diamond Hyd	lride, 150 x 2,1 n	nm, 2,2 μm		
Säulenofentemperatur		30 °C			
	Pumpe				
Dauer	2	20 min			
Flussrate	400	) μL/min			
Eluent A	$H_2O + 5 m$	nM AF (pH 3,5)			
Eluent B	ACN/IPA (10/1, v/	(v) + 5  mM AF (	pH 3,5)		
	Zeit %	Eluent A	% Eluent B		
	0	0	100		
	2	0	100		
	5	15	85		
	7	15	85		
Gradient	9	50	50		
	10,5	50	50		
	11,5	100	0		
	15	100	0		
	16	0	100		
	20	0	100		
	Massenspektrometer				
Ionisationsmethode		ESI			
Ionisationsmodus	positiv	oder negativ			
Massenbereich	60 – 1000 Da				
Spectra Rate		1 Hz			
	Lockmasse:	Hexa	ıkis(2,2-		
Kalibration	Lockindsse.	difluoroethoxy)phosphazen			
Kanbration	Kalibrierlösung:	Natriumfo	nformiat-Cluster		
	Calibration Mode	I	-IPC		
	Segment 1:	19 min, divert	valve to source		
Segmente	Segment 2 (Kalibration):	0,5 min, divert	valve to waste		
	Segment 3:	0,5 min, divert	valve to source		
	End Plate Offset:	+/-	500 V		
	Capillary:	45	500 V		
Ionenquelle	Nebulizer Gas:	4	bar		
	Drying Gas:	91	_/min		
	Drying Temperature:	20	<u>)0 °C</u>		
The second se	Funnel I RF:	275 Vpp			
Transfer	isCID Energy:		) eV		
	Multipole RF:	25	0 Vpp		
Quadrupol	Ion Energy:	Ion Energy: 4 ev			
-	Low Mass: 60 m/Z				
	Collision Energy:				
Collision Cell		30	<u>o vpp</u>		
	i ransier 1 ime:	1	υμs		
	Dea Dal- 04	Pre Puls Storage: 8 us			
	Pre Puls Storage:	8	β μs		
nur bei Aufnahme von	Pre Puls Storage: Scan Mode:	<u> </u>	3 μs S/MS		

#### 6.3.3 Kalibration

Zusätzlich zur Kalibrierung vor jedem Batch mittels Natriumformiat-Cluster erfolgte in jeder einzelnen Messung die Injektion von 20 µL des Clusters im Segment 2 über eine ventilgesteuerte Injektionsschleife (Zufuhr über Spritzenpumpe mit einer Flussrate von 0,1 mL/h) zur Re-Kalibrierung der Datensätze nach der Akquisistion. Neben dieser internen Kalibrierung über den gesamten Massenbereich erfolgte die Kalibrierung des Datensatzes mittels Lockmassen-Kalibration über die gesamte chromatographische Dauer durch kontinuierliche Evaporation einer (Hexakis(2,2-difluoroethoxy)phosphazen)-Lösung, welches auf einem Trägermaterial in der Ionenquelle aufgebracht wurde. Die Kalibrierung der Datensätze wurde mit Hilfe der Software DataAnalysis 4.1 vorgenommen. Die Parameter der verschiedenen Kalibrationen sind in Tabelle 6-7 tabellarisch aufgeführt.

Parar	neter	positiver Ionenmodus	negativer Ionenmodus			
		Lockmassen-Kalibration				
Sig	nal	622,0290 m/z	566,0020 m/z			
Search Range		0,05 m/z	0,05 m/z			
Intensity Threshold		1000	1000			
Kalibration vor jedem Batch und interne Kalibration						
Calibra	tion List	Na Formate (pos)	Na Formate (neg)			
Calibration	NTNP	Enhanced Quadratic	Enhanced Quadratic			
Mode	NTP	HPC	HPC			
Search Range		0,05 m/z	0,05 m/z			
Intensity '	Threshold	1000	1000			

Tabelle 6-7	: Parameter	der	Kalibration	der	non-targeted	Messungen.
-------------	-------------	-----	-------------	-----	--------------	------------

#### 6.3.4 Auswertung

Die Auswertung der Datensätze erfolgte mittels DataAnalysis 4.1, ProfileAnalysis 2.1 und/oder QuantAnalysis 2.1. Die zur *Feature Detection* verwendeten Parameter sind in Tabelle 6-8 dargestellt.

Parameter	NTNP-pos	NTNP-neg	NTP-pos	NTP-neg
Algorithmus	Find Molecular Features	Find Molecular Features	Find Molecular Features	Find Molecular Features
S/N Threshold	5	4	2	4
Correlation Coefficient Threshold	0,7	0,6	0,7	0,6
Minimum Compound Length	8	8	6	8
Smoothing Width	3	3 3		3
Additional Smoothing	ja	ja	ja	ja
Positive Adducts	M+H, M+NH4, M+Na, M+K, M-H2O+H, M-CO2+H, M-NH3+H, M- CH3OH+H	-	M+H, M+NH4, M+Na, M+K, M-H <sub>2</sub> O+H, M- CO <sub>2</sub> +H, M- NH <sub>3</sub> +H, M- CH <sub>3</sub> OH+H	-
M-H, M-H <sub>2</sub> C M+Na-H <sub>2</sub> Negative Adducts - M+K-H <sub>2</sub> , M- M+HCOOH M+COOH		M-H, M-H <sub>2</sub> O-H, M+Na-H <sub>2</sub> , M+K-H <sub>2</sub> , M+Cl, M+HCOOH-H, M+CH <sub>3</sub> COOH-H	-	M-H, M-H <sub>2</sub> O-H, M+Na-H <sub>2</sub> , M+K-H <sub>2</sub> , M+Cl, M+HCOOH-H, M+CH <sub>3</sub> COOH-H
Proteomics, CHNO	ja	ja	ja	ja
Background Substrac- tion	Spectral	Spectral	Spectral	Spectral
ChromatogrammRange	1,1; 24,0; 0; 0	1,1; 21,0; 0; 0	1,3; 15,0; 0; 0	1,3; 15,0; 0; 0

Tabelle 6-8: Feature Detection Parameter der Methoden NTNP-pos/NTNP-neg und NTP-pos/NTP-neg.

Auf Basis der detektierten *Features* in den Datensätzen eines Batchs wurde mittels ProfileAnalysis 2.1 ein Bucket-Table erstellt, welcher alle *Features* mit identischer Ionenmasse und Retentionszeit, die in einem bestimmten Anteil des Batches enthalten sind, mit der jeweiligen Signalintensität zusammenfasst. Die dafür verwendeten Parameter sind in Tabelle 6-9 zusammengefasst.

Tabelle 6-9: Parameter der Bucket-Table Erstellung.

Parameter	Einstellung			
Use Time Alignment	ja			
<b>Retention Time Range</b>	Dauer des jeweiligen chromatographischen Laufs			
Mass Range	60 - 1000  m/z			
Evaluation Dangas	621,9-622,1 (positiv);			
Exclusion Ranges	555,9 – 556,1 (negativ); 665,9 – 666,1 (negativ)			
<b>ΔTime (Advanced bucketing)</b>	0,3 min			
<b>ΔMass (Advanced bucketing)</b>	5 ppm			
Get Parameters From Time Alignment	nein			
Split Buckets With Multiple Compounds	ja			
Normalization	Sum of bucket values in analysis			
Value Count Of Bucket	$\ge$ 90 %			
Value Count Of Group Attribute	> 50.%			
Within Bucket	$\geq 30\%$			

Parameter	Einstellung
Allow Empty Group Attributes	ja
Missing Value Substitution	Average bucket value of group attribute
<b>Bucket Value Transformation</b>	None

#### 6.3.5 Probenmessungen

Die nach Abschnitt 6.1.3.2 bzw. 6.1.3.4 hergestellten unverdünnten Extrakte der Haselnussproben der Erntejahre 2014 und 2015 wurden je nach Fragestellung mit einer der vier LC-MS Methoden (siehe Abschnitt 6.3.1 bzw. 6.3.2) jeweils in Einfachbestimmung analysiert. Die Akquisition erfolgte jeweils in randomisierter Reihenfolge unter wiederholter Messung eines entsprechenden QCs (siehe Abschnitt 6.1.3.5) und eines Blindwertes (bestehend aus dem jeweiligen purem Extraktionsmittel der verwendeten Methode) nach jeder 10. Injektion und initialer Messung von 5 QCs.

# 6.3.6 Software-basierte Identifizierung von Markersubstanzen zur geographischen Herkunftsbestimmung

Zur Bestimmung von potentiellen Markersubstanzen wurden sowohl ANOVA-Tests zum Vergleich von mehr als einem Mittelwert als auch T-Test zum Vergleich zweier Mittelwerte für jede mögliche Länderpaarkombination mit ProfileAnalysis 2.1 durchgeführt und die entsprechenden p-values berechnet. Darüber hinaus wurden die Variationskoeffizienten für jeden Metaboliten berechnet.

Vor der Durchführung der T-Tests wurde jeweils mit The Unscrambler X auf Normalverteilung nach DAVID ( $\alpha = 0,05$ ) getestet bzw. mittels F-Test auf Varianzenhomogenität ( $\alpha = 0,01$ ).

## 6.3.7 Strukturaufklärung relevanter Metabolite

Die Identifizierung einzelner Metabolite erfolgte durch den Vergleich der exakten Masse, des Isotopenverhältnisses und der MS/MS-Spektren mit frei verfügbaren Datenbanken (z.B. MetFrag, MassBank, LipidMaps, HMDB). Ein abschließender Beweis der Identität erfolgte über den Vergleich der chromatographischen und massenspektrometrischen Eigenschaften mit kommerziell erhältlichen Standardsubstanzen.

#### 6.3.8 Multivariate Datenanalyse

Die nach Abschnitt 6.3.4 erstellten Bucket-Tables wurden entweder direkt mit der Software ProfileAnalysis 2.1 analysiert (ANOVA und T-Test) oder in die Software The Unscrambler X exportiert. Nach einem *Centering (Median Centering)* und *Scaling (Interquartile Range Scaling)* des Datensatzes konnten folgende multivariate Analysenverfahren mit der Software The Unscrambler X durchgeführt werden:

- PCA (NIPALS Algorithmus, Full Cross Validation);
- PLS-DA (NIPALS Algorithmus, Full Cross Validation);
- SIMCA, wobei f
  ür jede Probengruppe (hier Herkunftsland) ein eigenst
  ändiges PCA Modell erstellt wird;
- PCA-LDA, ideale Parameter f
  ür die Methode (*Linear, Quadratic* oder *Mahalanobis*) sowie die Anzahl der verwendeten PCA *scores* individuell f
  ür jede Fragestellung optimiert. Anschließende Vorhersage unbekannter Proben;
- SVM-C (Kreuzvalidierung mit x Segmenten), ideale Parameter für den SVM Typ (C-SVC oder nu-SVC), Kernel Typ (Linear, Polynomial x-ten Grades, Radial basis function oder Sigmoid), Offset, C-value, Gamma und/oder Nu individuell für jede Fragestellung optimiert, über die Grid Search Funktion bestimmt. Anschließende Vorhersage unbekannter Proben;
- MCIM (berechnet mittels der Software Excel 2010; f
  ür detaillierte Beschreibung vgl. Abschnitt 5.4.1), Berechnung von Konfidenzintervallen der f
  ür jeden Metabolite und jedes Land aus gescaleten Peakfl
  ächen mit Signifikanzniveau α = 0,05.
  - Schritt 1: Berechnung der Diskriminanzwerte anhand der Diskriminanzvariablen für die jeweiligen Länderpaarungen (siehe Tabelle 5-16, Abschnitt 5.4.1) als Mittelwert aus unterem Konfidenzintervall Metabolit A (höherer Gehalt) und oberem Konfidenzintervall Metabolit B (niedrigerer Gehalt). Anschließend Klassifizierung unbekannter Proben entsprechender der Diskriminanzwerte für jede Länderpaarung; Zuordnung Land A pro Länderpaar, wenn mind. zwei der drei Diskriminanzvariablen Land A entsprechen; Zuordnung Land A gesamt, wenn Land A bei jeder Länderpaarung klassifiziert;
  - Schritt 2: Wenn Schritt 1 nicht eindeutig, dann Anzahl der Metabolite pro Land innerhalb der Konfidenzintervalle berechnen. Zuordnung Land A gesamt, wenn Anzahl f
    ür Land A mind. drei gr
    ößer ist als n
    ächstniedrigeres Land;

 Schritt 3: Wenn Schritt 1 und Schritt 2 nicht eindeutig, dann "nicht klassifiziert"

Falls mehrere Batches für eine gemeinsame statistische Auswertung miteinander kombiniert wurden, erfolgte das *Centering* und *Scaling* für jeden Batch separat.

# 6.4 Targeted LC-MS (HPLC-ESI-QqQ-MS/MS)

# 6.4.1 Targeted LC-MS Methode

Die apparative Analytik der nach Abschnitt 6.1.3.4 extrahierten unpolaren Extrakte erfolgte auf einer Agilent 1260 Infinity Series HPLC, welche mit einem API2000 Triple Quadrupol Massenspektrometer gekoppelt war. Im Zuge der Methodenentwicklung wurden folgende Parameter variiert: HPLC-Säule, Eluent B, Additive, Gradient, Fließgeschwindigkeit und Säulenofentemperatur (vgl. Tabelle 5-7 und Tabelle 5-8, Abschnitt 5.1.2.2). Die massenspektrometrische Analyse erfolgte im positiven Ionenmodus. Nachfolgend sind die Geräteparameter (Tabelle 6-10) der targeted LC-MS Methode zusammengefasst.

	Autosampler					
Temperatur	R	aumtemperatur				
Injektionsvolumen		3 µL				
	Säulenofen					
Säule	Agilent, Poroshell 1	20 EC-C18, 50 x 4	4,6 mm, 2,7 μm			
Säulenofentemperatur		30 °C				
	Pumpe					
Dauer		13,5 min				
Flussrate	500 µL/min					
Eluent A	$H_2O + 5 \text{ mM AF (pH 3,5)}$					
Eluent B	IPA +	5 mM AF (pH 3,5	5)			
	Zeit	% Eluent A	% Eluent B			
	0	15	85			
	2	15	85			
Gradient	5	0	100			
	9	0	100			
	9,5	15	85			
	13,5	15	85			
	Massenspektrometer					
Ionisationsmethode	ESI					
Ionisationsmodus		positiv				
Ionenquelle		Einmessen	LC-MS Messung			
Ionenquene	Ion Spray Voltage:	5500 V	5500 V			

Tabelle 6-10: Methodenparameter für die targeted LC-MS Methode.

Massenspektrometer (Fortsetzung)							
Ionenquelle (Fortsetzung)		Einmessen	LC-MS Messung				
	Temperature:	0 °C	450 °C				
	Ion Source Gas 1:	30 psi	30 psi				
	Ion Source Gas 2:	0 psi	70 psi				
	Curtain Gas:	10 psi	20 psi				
	Collision Gas:	5 psi	5 psi				
	Ion Spray Probe	2	2				
	Vertical Position:	2	5				
	Ion Spray Probe	0	5				
	Horizontal Position:	0	5				

Material und Methoden

Zur Erhöhung der Sensitivität und Selektivität wurde das MRM angewendet. Zur Ermittlung der idealen massenspektrometrischen Einstellungen für jede Substanz wurde aus den Stammlösungen (je ca. 10 mM; siehe Abschnitt 6.2.2) durch Verdünnung mit Methanol 1 mL einer 10 µM Lösung erstellt, welche per Direkteinlass mithilfe einer Spritzenpumpe mit einer Fließgeschwindigkeit von 8 µL/min in die ESI-Quelle des massenspektrometrischen Systems eingebracht wurde. Über die Funktion "Automatic Compound Optimization" der Software Analyst 1.6 wurden sowohl für die Precursor-Ionen, als auch für die zwei intensivsten Fragment-Ionen automatisch die optimalen Parameter der elektrischen Spannungen für die Ionenoptiken und Quadrupole bestimmt, um eine maximale Transmission zu gewährleisten. Das intensivste Fragment-Ion wird im Folgenden als "Quantifier" (Qnt) bezeichnet und für die quantitative Auswertung verwendet. Das zweitintensivste Fragment wird als "Qualifier" (Qal) bezeichnet und dient der Überprüfung der Identität des Precursor-Ions. Die ESI-Quellenparameter für die automatische Optimierung sind in Tabelle 6-10 unter "Einmessen" dargestellt. Die Einteilung in Perioden erfolgte auf Basis der Retentionszeiten der Substanzen im flüssigchromatographischen Lauf (Periode 1 = 5,5 min; Periode 2 = 8 min). Die Dwell Times der Quantifier wurden in Periode 1 auf 60 ms und in Periode 2 auf 100 ms festgelegt, während für die Qualifier eine Dwell Time von 20 ms (Periode 1 und 2) gewählt wurde. Die optimierten Parameter des massenspektrometrischen Systems, sowie die erhaltenen Fragmente der einzelnen Verbindungen und deren Einteilung in verschiedene Perioden sind in Tabelle 6-11 tabellarisch dargestellt.

S-h-ton-		Precursor	Fragment	Rt	DP	FP	EP	CEP	CE	СХР
	Substanz	[m/z]	[m/z]	[min]	[V]	[V]	[V]	[V]	[V]	[V]
	γ-Tocopherol Qnt	417,464	151,1	2,60	136	50	8,5	20	29	8
	γ-Tocopherol Qal	417,464	69,1	2,60	136	50	8,5	20	53	8
	PC(18:2/0:0) Qnt	520,353	184,2	1,19	36	370	10,5	26	35	6
	PC(18:2/0:0) Qal	520,353	104,2	1,19	36	370	10,5	26	39	4
	PC(18:1d7/0:0) Qnt	529,358	184,2	1,28	126	90	9	24	35	8
	PC(18:1d7/0:0) Qal	529,358	104,2	1,28	126	90	9	24	45	6
	DG(16:0/16:1) Qnt	584,525	311,4	3,49	21	370	11,5	26	35	18
	DG(16:0/16:1) Qal	584,525	313,4	3,49	21	370	11,5	26	35	18
le 1 in	DG(15:0/18:1d7) Qnt	605,431	570,6	3,81	41	160	6	38	25	16
ciod 5 m	DG(15:0/18:1d7) Qal	605,431	299,3	3,81	41	160	6	38	31	16
Pei 5,	DG(16:0/18:1) Qnt	612,607	339,4	4,26	21	370	11,5	26	31	10
	DG(16:0/18:1) Qal	612,607	313,5	4,26	21	370	11,5	26	31	8
	DG(18:2/18:2) Qnt	634,520	377,3	3,38	36	360	6,5	26	33	18
	DG(18:2/18:2) Qal	634,520	95,2	3,38	36	360	6,5	26	61	4
	DG (18:1/18:1) Qnt	638,640	399,5	4,43	21	370	7	26	37	10
	DG (18:1/18:1) Qal	638,640	83,2	4,43	21	370	7	26	69	4
	TG(2:0/18:2/18:2) Qnt	676.551	379,4	4,12	30	360	7	28	37	10
	TG(2:0/18:2/18:2) Qal	676,551	599,5	4,12	30	360	7	28	33	16
	TG(2:0/18:1/18:2) Qnt	678,570	381,3	4,73	30	360	7	28	37	10
	TG(2:0/18:1/18:2) Qal	678,570	379,3	4,73	30	360	7	28	37	10
	PE(15:0/18:1d7) Qnt	711,571	570,6	2,44	81	270	9	26	31	18
	PE(15:0/18:1d7) Qal	711,571	57,1	2,44	81	270	9	26	79	12
	PE(16:0/18:2) Qnt	716,581	575,6	2,41	36	350	9	30	31	16
	PE(16:0/18:2) Qal	716,581	81,2	2,41	36	350	9	30	83	2
	PE(18:2/18:2) Qnt	740,647	599,5	2,26	41	370	9	32	33	16
	PE(18:2/18:2) Qal	740,647	95,2	2,26	41	370	9	32	69	4
	PE(18:1/18:1) Qnt	744,470	603,6	2,82	126	370	9,5	30	29	16
	PE(18:1/18:1) Qal	744,470	69,0	2,82	126	370	9,5	30	87	8
	PC(15:0/18:1d7) Qnt	753,592	184,2	2,43	151	310	11	28	45	12
	PC(15:0/18:1d7) Qal	753,592	86,2	2,43	151	310	11	28	101	10
	PC(16:0/18:3) Qnt	756,550	184,2	2,13	56	360	10	30	43	8
	PC(16:0/18:3) Qal	756,550	86,1	2,13	56	360	10	30	97	12
	PC(16:0/18:2) Qnt	758,635	184,3	2,41	56	360	10	30	43	8
	PC(16:0/18:2) Qal	758,635	86,2	2,41	56	360	10	30	97	12
	PC(18:2/18:2) Qnt	782,596	184,2	2,25	76	350	7,5	32	45	8
	PC(18:2/18:2) Qal	782,596	86,1	2,25	76	350	7,5	32	97	10
	PC(18:1/18:2) Qnt	784,596	184,2	2,48	76	350	7,5	32	45	8
	PC(18:1/18:2) Qal	784,596	86,1	2,48	76	350	7,5	32	97	10
	TG(14:0/16:0/18:1) Qnt	822,755	549,5	8,19	66	370	9	36	33	16
	TG(14:0/16:0/18:1) Qal	822,755	577,5	8,19	66	370	9	36	33	16
0	TG(15:0/18:1d7/15:0) Qnt	829,782	570,6	8,15	131	160	8,5	28	33	16
de j	TG(15:0/18:1d7/15:0) Qal	829,782	523,5	8,15	131	160	8,5	28	35	14
s m	TG(15:0/16:0/18:1) Qnt	836,711	563,5	6,76	40	370	7	30	33	16
Pe	TG(15:0/16:0/18:1) Qal	836,711	537,5	6,76	40	370	7	30	33	16
	TG(16:0/16:1/18:1) Qnt	848,770	575,5	8,23	66	370	9	36	35	16
	TG(16:0/16:1/18:1) Qal	848,770	81,1	8,23	66	370	9	36	80	4
	TG(16:0/18:1/18:2) Qnt	874,829	81,0	8,45	56	360	9	36	87	10

#### Tabelle 6-11: MRM-Parameter der targeted LC-MS Methode. (modifiziert nach Klockmann 2017)

	Substanz	Precursor [m/z]	Fragment [m/z]	R <sub>t</sub> [min]	DP [V]	FP [V]	EP [V]	CEP [V]	CE [V]	CXP [V]
(g)	TG(16:0/18:1/18:2) Qal	874,829	95,1	8,45	56	360	9	36	85	4
etzun	TG(16:0/18:1/18:1) Qnt	876,806	95,1	8,26	46	270	10,5	36	83	4
orts.	TG(16:0/18:1/18:1) Qal	876,806	81,0	8,26	46	270	10,5	36	89	4
2 (F mii	TG(17:1/18:1/18:2) Qnt	886,787	587,5	8,20	60	360	9	38	33	16
de 8	TG(17:1/18:1/18:2) Qal	886,787	589,5	8,20	60	360	9	38	33	16
erio	TG(18:2/18:2/18:3) Qnt	894,756	597,5	7,76	66	370	9	38	33	16
Ρ	TG(18:2/18:2/18:3) Qal	894,756	599,5	7,76	66	370	9	38	33	16

Material und Methoden

R<sub>t</sub>: Retentionszeit; DP: *declustering potential*; FP: *focussing potential*; EP: *entrance potential*; CEP: *collision cell entrance potential*; CE: *collision energy*; CXP: *collision cell exit potential* 

# 6.4.2 Validierung

Die Validierung der targeted HPLC-ESI-QqQ-MS/MS Methode erfolgte gemäß anerkannter Richtlinien der FDA und der DIN 32645.<sup>212, 213</sup> Dazu wurde die Grundkalibrierreihe sowie die Matrixkalibrierreihe (vgl. Abschnitt 6.2.3.1) mit der targeted Methode (vgl. Abschnitt 6.4.1) jeweils in Fünffachbestimmung gemessen und anschließend folgende Parameter gemäß *Bioanalytical Method Validation* der FDA 2001 für jeden der enthaltenen Metabolite berechnet:

- <u>Empfindlichkeit (m)</u>: Entspricht der Steigung der Regressionsgeraden und ist ein Maß für die Sensitivität der Methode. Je größer die Empfindlichkeit, desto kleiner ist der noch messbare Konzentrationsunterschied.
- <u>Mittleres Begleitsignal (b)</u>: Entspricht dem Achsenabschnitt der Regressionsgeraden und ist ein Maß f
  ür den systematischen Fehler der Messung.
- <u>Bestimmtheitsmaß der Kalibriergeraden (R<sup>2</sup>)</u>: Ist ein Maß für die Stärke des linearen Zusammenhangs zwischen zwei Variablen und ergibt sich rechnerisch durch Quadrieren des PEARSON'schen Korrelationskoeffizienten. Je näher sich der Wert an 1 annähert, desto geringer ist die Streuung der Werte um die Regressionsgerade und als desto größer ist die Linearität anzusehen.
- <u>Verfahrensstandardabweichung (S<sub>V</sub>)</u>: Ist ein Gütemaß für die Leistungsfähigkeit einer Methode, die sich aus der Normierung der Präzision auf die Mitte des Arbeitsbereiches berechnet.

- <u>Verfahrensvariationskoeffizient (V<sub>KV</sub>)</u>: Ist ein Maß für die Robustheit einer Methode und erlaubt einen Vergleich verschiedener Verfahren, vorausgesetzt die Messwerte beziehen sich auf gleiche Konzentrationen.
- <u>Wiederholgrenze (r)</u>: Gibt die maximale Differenz zwischen zwei Messwerten an, die bei der Wiederholung einer Messung akzeptabel ist.
- <u>Linearer Arbeitsbereich (LA)</u>: Ist der Bereich, in dem ein Detektor ein Signal proportional zur Masse oder Konzentration eines Analyten liefert; die Empfindlichkeit also konstant ist. Das untere Ende wird auch als *Lower Limit of Quantitation* (LLOQ; dt. Bestimmungsgrenze) bezeichnet. Der lineare Arbeitsbereich wird mithilfe des Linearitätstest nach MANDEL bestimmt.<sup>253</sup> Die Überprüfung der Homoskedastizität erfolgte mittels BREUSCH-PAGAN-Test.<sup>254</sup>
- <u>Richtigkeit:</u> Beschreibt die Abweichung der gemittelten Messergebnisse zum tatsächlichen Konzentrationswert und entspricht der Richtigkeit der Kalibrierfunktion. Es wird empfohlen, die Richtigkeit für mindestens drei Konzentrationen im linearen Bereich anzugeben. Die Richtigkeit sollte nicht mehr als 15 % betragen; beim LLOQ nicht mehr als 20 %. Der lineare Arbeitsbereich wird soweit eingegrenzt, bis diese Kriterien erfüllt sind.
- <u>Präzision</u>: Beschreibt die zufällige Streuung von Messwerten und ist ein Maß für die Reproduzierbarkeit unabhängig berechneter Messwerte. Sie entspricht dem Variationskoeffizienten (CV). Es wird empfohlen, die Präzision für mindestens drei Konzentrationen im linearen Bereich anzugeben, wobei sie 15 % nicht überschreiten sollte; beim LLOQ 20 %. Der lineare Arbeitsbereich wird soweit eingegrenzt, bis diese Kriterien erfüllt sind.

Die Bestimmung der Nachweisgrenzen erfolgte gemäß DIN 32645 nach der Leerwertmethode. Hierfür wurden 10 unabhängig erstellte Blindwerte (vgl. Abschnitt 6.2.) gemessen.

- <u>Nachweisgrenze (NWG)</u>: Entspricht der niedrigsten Analytkonzentration, die bei einer einmaligen Messung qualitativ nachgewiesen werden kann.

Ausreißer bei Mehrfachbestimmungen wurden mittels Ausreißertest nach DEAN und DIXON identifiziert und eliminiert. Die Formeln zur Berechnung sämtlicher Parameter sind im Anhang unter Abschnitt 8.1.5 aufgeführt.

# 6.4.3 Bestimmung der Extrakt-Verdünnungsstufe zur Quantifizierung

Zur Bestimmung der benötigten Verdünnung der Haselnussextrakte für eine Quantifizierung im linearen Arbeitsbereich wurde eine Verdünnungsreihe eines repräsentativen Extraktes (siehe Abschnitt 6.2.3.3) in Fünffachbestimmung mit der targeted LC-MS Methode gemessen und anschließend entsprechend der Vorgaben aus Abschnitt 6.4.2 der lineare Arbeitsbereich bestimmt. Aus diesen Informationen wurde die für die Quantifizierung des jeweiligen Metabolits zu verwendende Verdünnung des Extraktes bestimmt (siehe Tabelle 6-12).

## 6.4.4 Berechnung der Korrekturfaktoren

Zur Berücksichtigung des unterschiedlichen Detektionsvermögens wurde für jeden der 20 Markersubstanzen ein Korrekturfaktor aus den Ergebnissen der Grundkalibrierung berechnet. Die Einzel-Korrekturfaktoren für die Konzentrationen jeweiligen wurden anhand der Gleichung 8-16 (Anhang, Abschnitt 8.1.5) berechnet. Der jeweils Berechnung der Metabolitzur Konzentration verwendete Korrekturfaktor wurde anschließend aus den die gemittelten Einzelwerten für Konzentrationen im übereinstimmenden linearen Arbeitsbereich berechnet. Für die Markersubstanzen ohne Referenzstandard wurde ein abgeschätzter Korrekturfaktor durch Extrapolation anhand der Werte derselben Stoffklasse

Tabelle 6-12: Korrekturfaktoren und zur Auswertungverwendete Verdünnung des Haselnussextraktes zurQuantifizierung mittels internem Standard.(modifiziert nach Klockmann 2017)

Metabolit	Verdünnung	KF
PC(18:2/0:0)	unverdünnt	0,0815
DG(16:0/16:1)	unverdünnt	0,32951
DG(16:0/18:1)	unverdünnt	0,4685
DG(18:2/18:2)	unverdünnt	0,2130
DG(18:1/18:1)	unverdünnt	0,3071
TG(2:0/18:2/18:2)	unverdünnt	0,3983 <sup>2</sup>
TG(2:0/18:1/18:2)	unverdünnt	0,3983 <sup>2</sup>
PE(16:0/18:2)	unverdünnt	0,1721
PE(18:2/18:2)	unverdünnt	0,1494
PE(18:1/18:1)	unverdünnt	0,6333
PC(16:0/18:3)	unverdünnt	0,05391
PC(16:0/18:2)	unverdünnt	0,0565
PC(18:2/18:2)	unverdünnt	0,0514
PC(18:1/18:2)	unverdünnt	0,05391
TG(14:0/16:0/18:1)	1:100	0,3983 <sup>2</sup>
TG(15:0/16:0/18:1)	unverdünnt	0,3983 <sup>2</sup>
TG(16:0/16:1/18:1)	1:100	0,3983 <sup>2</sup>
TG(17:1/18:1/18:2)	1:100	0,3983 <sup>2</sup>
TG(18:2/18:2/18:3)	1:100	0,3983 <sup>2</sup>

<sup>1</sup>extrapoliert anhand der Stoffklasse <sup>2</sup>extrapoliert anhand TG(16:0/18:1/18:1) und

TG(16:0/18:1/18:2)

erhalten. Die resultierenden Korrekturfaktoren sind in Tabelle 6-12 dargestellt.

## 6.4.5 Probenmessungen

Die nach Abschnitt 6.1.3.3 erhaltenen Extrakte der Haselnussproben aus den Erntejahren 2014 und 2015 wurden mit der targeted LC-MS Methode (siehe Abschnitt 6.4.1) jeweils in

Dreifachbestimmung sowohl unverdünnt als auch 1:100 verdünnt analysiert. Die Akquisition erfolgte jeweils in randomisierter Reihenfolge unter wiederholter Messung eines entsprechenden QCs (siehe Abschnitt 6.1.3.5) und eines Blindwertes (bestehend aus dem reinem Extraktionsmittel) nach jeder 10. Injektion.

Die Quantifizierung der absoluten Metabolit-Gehalte erfolgte über die internen Standards nach Gleichung 8-17 (Anhang, Abschnitt 8.1.5). Dabei wurde für jeden der Metabolite diejenige interne Standardsubstanz verwendet, die der entsprechenden Stoffklasse zugehörig war. Für die statistische Auswertung wurden für jeden Metaboliten die gemäß Tabelle 6-12 (Abschnitt 6.4.4) definierten Verdünnungen zur Quantifizierung verwendet und die entsprechend resultierenden Werte in einer Datenmatrix vereint.

#### 6.4.6 Mischungen

Die Proben DE035, FR095, GE010, IT020 und TR012 als zufällig gewählte, repräsentative Proben wurden in unterschiedlichen Mischungskombinationen eingewogen, sodass in Summe jeweils ca. 50 mg Haselnussgranulat zur Extraktion erhalten wurde. Jede mögliche Länder-Kombination an Zwei-Komponenten-Mischungen wurde in den Verhältnissen 80/20, 60/40, 50/50, 40/60 und 20/80 erstellt. Darüber hinaus wurden die Proben der fünf Länder jeweils ungemischt verwendet. Zusätzlich wurden Drei-Komponenten-Mischungen für jede mögliche Länder-Kombination im Verhältnis 33/33/33 erstellt. Die so erstellten Proben wurden nach Abschnitt 6.1.3.3 extrahiert und jeweils mit der targeted LC-MS Methode (siehe Abschnitt 6.4.1) sowohl unverdünnt als auch 1:100 verdünnt analysiert. Die Akquisition erfolgte jeweils in randomisierter Reihenfolge unter wiederholter Messung eines Blindwertes (bestehend aus dem reinem Extraktionsmittel) nach jeder 10. Injektion.

Die Berechnung der Metabolit-Gehalte und die Erstellung der Datenmatrix erfolgte analaog zu Abschnitt 6.4.5.

Zur Erstellung der Regressionsmodelle wurden virtuelle Mischungen der authentischen Proben berechnet. Dazu wurden für jedes Land im jeweiligen Erntejahr 2014 bzw. 2015 die mittleren Metabolit-Gehalte berechnet. Diese Mittelwertproben (insgesamt 12, da zwei Erntejahre mit je sechs Ländern) wurden anschließend jeweils mit jeder authentischen Probe (Mittelwert der Dreifachbestimmung pro Probe) verrechnet, sodass virtuelle Mischungen in den Verhältnissen 80/20, 60/40, 50/50, 40/60 und 20/80 berechnet wurden. Zusammen mit den ursprünglichen, zur Berechnung verwendeten 100 %-Proben (Mittelwert der Dreifachbes-

stimmung pro Probe) wurden diese virtuellen Mischungen als Training-Datensatz zur Modellerstellung in der multivariaten Datenanalyse verwendet.

## 6.4.7 Univariate und multivariate Datenanalyse

Zur Untersuchung der Signifikanz von Unterschieden zwischen Metabolit-Gehalten unterschiedlicher Länder wurden sowohl ANOVA-Tests zum Vergleich von mehr als einem Mittelwert als auch T-Test zum Vergleich zweier Mittelwerte für jede mögliche Länderpaarkombination mit der Software ProfileAnalysis 2.1 durchgeführt und die entsprechenden pvalues berechnet.

Vor der Durchführung der T-Tests wurde jeweils mit der Software The Unscrambler X auf Normalverteilung nach DAVID ( $\alpha = 0,05$ ) getestet bzw. mittels F-Test auf Varianzenhomogenität ( $\alpha = 0,01$ ).

Die berechneten Metabolit-Gehalte wurden entweder mit der Software ProfileAnalysis 2.1 analysiert (ANOVA und T-Test) oder in die Software The Unscrambler X exportiert. Nach einem *Centering (Median Centering)* und *Scaling (Interquartile Range Scaling)* des Datensatzes konnten folgende multivariate Analysenverfahren mit The Unscrambler X angewendet werden:

- PCA analog zu Abschnitt 6.3.8;
- PLS-R und PLS-DA (NIPALS Algorithmus, Validierung via Full Cross Validation);
- SIMCA, analog zu Abschnitt 6.3.8;
- PCA-LDA, analog zu Abschnitt 6.3.8;
- SVM-C, analog zu Abschnitt 6.3.8;
- SVM-R (Kreuzvalidierung mit x Segmenten), ideale Parameter für den SVM Typ (Epsilon-SVR oder nu-SVR), Kernel Typ (Linear, Polynomial x-ten Grades, Radial basis function oder Sigmoid), Offset, Epsilon, C, Gamma und/oder Nu individuell für jede Fragestellung optimiert, über die Grid Search Funktion bestimmt. Anschließende Vorhersage unbekannter Proben
- MLR (Signifikanz-Level  $\alpha = 0,05$ , Validierung via *Leverage Correction*)
- MCIM, analog zu Abschnitt 6.3.8, jedoch inkl. Spanien und mit teilweise anderen Diskriminanzmetaboliten (siehe Tabelle 8-14, Abschnitt 5.4.2).

Die berechneten *Scaling* Parameter für die authentischen Proben (siehe Tabelle 6-13) wurden anschließend zum *Scaling* der nicht-authentischen Proben verwendet. Die Berechnung der gescaleten Werte erfolgte nach Gleichung 8-19.

Metabolit	Median [mg/kg]	IQR [mg/kg]
PC(18:2/0:0)	3,59	1,86
DG(16:0/16:1)	1,36	0,49
DG(16:0/18:1)	718,13	234,68
DG(18:2/18:2)	1139,35	579,50
DG(18:1/18:1)	3160,60	1054,72
TG(2:0/18:2/18:2)	196,68	170,68
TG(2:0/18:1/18:2)	828,48	516,95
PE(16:0/18:2)	89,78	39,31
PE(18:2/18:2)	78,66	63,42
PE(18:1/18:1)	599,64	151,57
PC(16:0/18:3)	3,58	1,77
PC(16:0/18:2)	83,07	29,47
PC(18:2/18:2)	101,47	75,69
PC(18:1/18:2)	225,92	213,68
TG(14:0/16:0/18:1)	127,13	55,77
TG(15:0/16:0/18:1)	19,11	6,87
TG(16:0/16:1/18:1)	3357,39	1071,21
TG(17:1/18:1/18:2)	411,14	176,24
TG(18:2/18:2/18:3)	692,67	292,14

Tabelle 6-13: Scaling Parameter,	berechnet :	anhand d	ler 202 :	authentischen	Haselnussproben.	(modifiziert
nach Klockmann 2017)					_	

# 6.5 Lagerungsversuche

Es wurden ungeschälte Haselnüsse, geschälte Haselnusskerne, Haselnussgranulat (Haselnusshomogenisat) und Haselnusslyophilisat in Doppelbestimmung bei unterschiedlichen Temperaturen (-20 °C, -40 °C und -80 °C) für definierte Zeiträume (1 Woche, 2 Wochen, 4 Wochen und 8 Wochen) gelagert und anschließend mittels der Methode NTNP-pos (siehe Abschnitt 6.3.1) analysiert. Die Lagerung der Extrakte erfolgte abweichend bei 4 °C, -20 °C und -40 °C. Zur Auswertung wurden mittels QuantAnalysis 2.1 die Peakflächen der 20 Markersubstanzen berechnet. Die Haselnussproben wurden wie gewohnt aufgearbeitet (siehe Abschnitt 6.1.2 und 6.1.3.2). Unmittelbar nach Ablauf der Lagerzeit wurden die jeweiligen Zwischenprodukte entsprechend der noch ausstehenden Aufarbeitungs- und Extraktionsschritte (siehe Abschnitt 6.1.3.2) extrahiert und bis zur Analyse bei -80 °C zwischengelagert. Für die Lagerungsversuche der geschälten und ungeschälten ganzen Haselnüsse wurde die Probe FR116 verwendet, während Homogenisat, Lyophilisat und Extrakt anhand der Probe IT001 untersucht wurden.

# 7 Literaturliste

(1) Janick, J.; Paull, R. E.: *The Encyclopedia of Fruit & Nuts.* CAB International, Wallingford, UK, **2008**.

(2) Thompson, M. M.; Lagerstedt, H. B., Hazelnuts. In *Fruit Breeding: Nuts*, 3 Auflage.; Janick, J.; Moore, J. N. (Hrsg.), John Wiley & Sons, New York, **1996**; S. 125-184.

(3) Mehlenbacher, S. A., Hazelnuts (*Corylus*). In *Genetic resources of temperate fruit and nut crops*, Moore, J. N.; Ballington, J. R. (Hrsg.), Acta Horticulturae, Leuven, Belgien, **1991**; S. 791-838.

(4) Schütt, P.: *Enzyklopädie der Sträucher: die große Enzyklopädie mit über 400 Farbfotos unter Mitwirkung von 30 Experten.* Nikol, Hamburg, **2006**.

(5) Duke, J. A.: *Handbook of Nuts: Herbal Reference Library*. CRC Press, Boca Raton, FL, **2001**.

(6) Woodruff, J. G.: *Tree Nuts: Production, Processing, Products.* 2. Auflage, AVI Publishing, Westport, CT, **1979**.

(7) Wilkinson, J.: *Nut Grower's Guide: The Complete Handbook for Producers and Hobbyists.* Landlinks Press, Collingwood, V.I.C., **2005**.

(8) Leslie, J. F.; Bandyopadhyay, R.; Visconti, A.: *Mycotoxins: Detection Methods, Management, Public Health and Agricultural Trade.* CAB International, London, UK, **2008**.

(9) Tous, J.; Girona, J.; Tasias, J., *Cultural practices and costs in hazelnut production*, III International Congress on Hazelnut, Me, G.; Radicati, L., (Hrsg.), *Acta Hort.*, Alba, Italien, **1992**; S. 395-418.

(10) Pallottino, F.; Menesatti, P.; Costa, C.; Paglia, G.; De Salvador, F. R.; Lolletti, D. Image analysis techniques for automated hazelnut peeling determination. *Food Bioprocess Technol.* **2010**, *3*, 155-159.

(11) Cristofori, V.; Ferramondo, S.; Bertazza, G.; Bignami, C. Nut and kernel traits and chemical composition of hazelnut (*Corylus avellana* L.) cultivars. *J. Sci. Food Agric.* **2008**, *88*, 1091-1098.

(12) Bayerische Landesanstalt für Landwirtschaft (LFL), Endbericht: Unter welchen Bedingungen kann der Haselnussanbau eine wirtschaftliche und planzenbauliche Alternative zum Tabakanbau bieten - Fortsetzung (2009-2011): Ist der Anbau von Haselnüssen in Bayern wirtschaftlich möglich, Freising, **2012**.

(13) Fideghelli, C.; De Salvador, F. R., *World hazelnut situation and perspectives*, VII International Congress on Hazelnut, Varvaro, L.; Franco, S., (Hrsg.), *Acta Hort.*, Viterbo, Italien, **2008**; S. 39-52.

(14) Snare, L. Hazelnut production. *Primefacts. Profitable & Sustainable Primary Industry.* **2008**, 765, 1-8.

(15) Nitsch, C., Durchführung eines Sorten- und Anbauversuches zur Ermittlung der Erträge, der Qualität und der Pflanzengesundheit, AELF Fürth, Gartenbauzentrum Bayern Mitte, **2015**.

(16) Ellena, M.; Sandoval, P.; Gonzalez, A.; Jequier, J.; Contreras, M.; Grau Beretta, P. Chilean hazelnut situation and perspectives. *Acta Hortic.* **2012**, 329-342.

(17) Muehlbauer, M. F.; Honig, J. A.; Capik, J. M.; Vaiciunas, J. N.; Molnar, T. J. Characterization of eastern filbert blight-resistant hazelnut germplasm using microsatellite markers. *J. Am. Soc. Hortic. Sci.* **2014**, *139*, 399-432.

(18) Fontana, M.; Somenzi, M.; Tesio, A., *Cultivation, harvest and postharvest aspects that influence quality and organoleptic properties of hazelnut production and related final products*, VIII International Congress on Hazelnut, Grau Beretta, P.; Ellena, M., (Hrsg.), *Acta Hort.*, Temuco City, Chile, **2012**; S. 311-314.

(19) European Commission, *DOOR Database of Origin & Registration*. http://ec.europa.eu/agriculture/quality/door/list.html, Zugriff am **21.02.2017**.

(20) Alasalvar, C.; Shahidi, F.: *Tree nuts - Composition, Phytochemicals and Health Effects.* CRC Press Taylor & Francis Group, Boca Raton, FL, **2009**.

(21) Souci, S. W.; Fachmann, W.; Kraut, H.: *Die Zusammensetzung der Lebensmittel, Nährwert-Tabellen.* 7. Auflage, MedPharm Scientific Publishers, Stuttgart, **2008**.

(22) Parcerisa, J.; Richardson, D. G.; Rafecas, M.; Codony, R.; Boatella, J. Fatty acid, tocopherol and sterol content of some hazelnut varieties (*Corylus avellana* L.) harvested in Oregon (USA). *J. Chromatogr. A.* **1998**, *805*, 259-268.

(23) Ciemniewska-Żytkiewicz, H.; Pasini, F.; Verardo, V.; Bryś, J.; Koczoń, P.; Caboni, M. F. Changes of the lipid fraction during fruit development in hazelnuts (*Corylus avellana* L.) grown in Poland. *Eur. J. Lipid Sci. Technol.* **2014**, n/a-n/a.

(24) Ciemniewska-Zytkiewicz, H.; Verardo, V.; Pasini, F.; Brys, J.; Koczon, P.; Caboni, M. F. Determination of lipid and phenolic fraction in two hazelnut (*Corylus avellana* L.) cultivars grown in Poland. *Food Chem.* **2015**, *168*, 615-622.

(25) Bada, J. C.; León-Camacho, M.; Prieto, M.; Alonso, L. Characterization of oils of hazelnuts from Asturias, Spain. *Eur. J. Lipid Sci. Technol.* **2004**, *106*, 294-300.

(26) Amaral, J. S.; Casal, S.; Seabra, R. M.; Oliveira, B. P. P. Effects of roasting on hazelnut lipids. *J. Agric. Food Chem.* **2006**, *54*, 1315-1321.

(27) Crews, C.; Hough, P.; Godward, J.; Brereton, P.; Lees, M.; Guiet, S.; Winkelmann, W. Study of the main constituents of some authentic hazelnut oils. *J. Agric. Food Chem.* **2005**, *53*, 4843-4852.

(28) Rezaei, F.; Bakhshi, D.; Ghazvini, R. F.; Majd, D. J.; Pourghayoumi, M. Evaluation of fatty acid content and nutritional properties of selected native and imported hazelnut (*Corylus avellana* L.) varieties grown in Iran. *J. Appl. Bot. Food Qual.* **2014**, *87*.

(29) Solar, A.; Stampar, F. Characterisation of selected hazelnut cultivars: phenology, growing and yielding capacity, market quality and nutraceutical value. *J. Sci. Food Agric.* **2011**, *91*, 1205-1212.

(30) Coetsee, J, Suid-Afrika kan pas aangee met haselneute, Landbouweekblad, 2010.

(31) Baldwin, B.; Gilchrist, K.; Snare, L.: *Hazelnut Variety Assessment for South-Eastern Autralia*. RIRDC, Barton, A.C.T, **2007**.

(32) Ferrero International S. A., *Sustainable Agricultural Practices - Ferrero Agricultural Enterprises for the Cultivation of Hazelnuts*, http://www.rspo.org/file/acop2014/ferrero-trading-lux-sa/M-Policies-to-PNC-waterland.pdf, Zugriff am **22.11.2016**.

(33) United Nations, UNECE Standard DDP-04 Concerning the Marketing and Commercial Quality Control of Hazelnut Kernels, New York Genf, **2010**.

(34) Garrone, W.; Vacchetti, M. La qualità delle nocciole in rapporto alle esigenze dell'industria dolciaria utilizzatrice. *Acta Hortic.* **1994**, *351*, 641-648.

(35) Liland, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC, Trends Anal. Chem.* **2011**, *30*, 827-841.

(36) Amaral, J. S.; Casal, S.; Citová, I.; Santos, A.; Seabra, R. M.; Oliveira, B. P. P. Characterization of several hazelnut (*Corylus avellana* L.) cultivars based in chemical, fatty acid and sterol composition. *Eur. Food Res. Technol.* **2006**, *222*, 274-280.

(37) Europäische Komission, Verordnung (EG) Nr. 510/2006 'Nocciola del Piemonte/Nocciola Piemonte' g.g.A. *Amtsblatt der Europäischen Union.* **2012**, *C330*, 17-22.

(38) Food and Agriculture Organization of the United Nations, *FAOstat Food and agricultural data*, updated on Mar 10, 2017; http://www.fao.org/faostat/en/#data, Zugriff am **18.03.2017**.

(39) Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics.* **2001**, *2*, 155-168.

(40) Nicholson, J. K.; Lindon, J. C.; Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. **1999**, *29*, 1181-1189.

(41) Nicholson, J. K.; Lindon, J. C. Systems biology: Metabonomics. *Nature*. **2008**, *455*, 1054-1056.

(42) Faber, J. H.; Malmodin, D.; Toft, H.; Maher, A. D.; Crockford, D.; Holmes, E.; Nicholson, J. K.; Dumas, M. E.; Bausengaard, D. Metabonomics in diabetes research. *J. Diabetes Sci. Technol.* **2007**, *1*, 549-557.

(43) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* **2005**, *24*, 613-646.

(44) Oms-Oliu, G.; Odriozola-Serrano, I.; Martín-Belloso, O. Metabolomics for assessing safety and quality of plant-derived food. *Food Res. Int.* **2013**, *54*, 1172-1183.

(45) Harris, E. D. *Biochemical facts behind the definition and properties of metabolites*. Texas A&M University, College Station, TX, **2014**.

(46) Sangwan, N. S.; Tiwari, P.; Mishra, S. K.; Yadav, R. K.; Tripathi, S.; Kushwaha, A. K.; Sangwan, R. S., Plant Metabolomics: An Overview of Technology Platforms for Applications in Metabolism. In *PlantOmics: The Omics of Plant Science*, Barh, D.; Khan, M. S.; Davies, E. (Hrsg.), Springer India, Neu Delhi, **2015**; S. 257-298.

(47) Dixon, R. A.; Strack, D. Phytochemistry meets genome analysis, and beyond. *Phytochemistry*. **2003**, *62*, 815-816.

(48) Trethewey, R. N. Metabolite profiling as an aid to metabolic engineering in plants. *Curr. Opin. Plant Biol.* **2004**, *7*, 196-201.

(49) Cubero-Leon, E.; Peñalver, R.; Maquet, A. Review on metabolomics for food authentication. *Food Res. Int.* **2014**, *60*, 95-107.

(50) Danezis, G. P.; Tsagkaris, A. S.; Brusic, V.; Georgiou, C. A. Food authentication: state of the art and prospects. *Curr. Opin. Food Sci.* **2016**, *10*, 22-31.

(51) Esslinger, S.; Riedl, J.; Fauhl-Hassek, C. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Res. Int.* **2014**, *60*, 189-204.

(52) Hu, C.; Xu, G. Mass-spectrometry-based metabolomics analysis for foodomics. *TrAC*, *Trends Anal. Chem.* **2013**, *52*, 36-46.

(53) Garcia-Canas, V.; Simo, C.; Herrero, M.; Ibanez, E.; Cifuentes, A. Present and future challenges in food analysis: Foodomics. *Anal. Chem.* **2012**, *84*, 10150-10159.

(54) Parcerisa, J.; Rafecas, M.; Castellote, A. I.; Codony, R.; Farràn, A.; Garcia, J.; López, A.; Romero, A.; Boatella, J. Influence of variety and geographical origin on the lipid fraction of hazelnuts (*Coryllus avellana* L.) from Spain: (II). Triglyceride composition. *Food Chem.* **1994**, *50*, 245-249.

(55) Ulrich, D.; Nothnagel, T.; Schulz, H. Influence of cultivar and harvest year on the volatile profiles of leaves and roots of carrots (*Daucus carota* spp. *sativus* Hoffm.). *J. Agric. Food Chem.* **2015**, *63*, 3348-3356.

(56) Kim, H. K.; Verpoorte, R. Sample preparation for plant metabolomics. *Phytochem. Anal.* **2010**, *21*, 4-13.

(57) Parcerisa, J.; Codony, R.; Boatella, J.; Rafecas, M. Triacylglycerol and phospholipid composition of hazelnut (*Corylus avellana* L.) lipid fraction during fruit development. *J. Agric. Food Chem.* **1999**, *47*, 1410-1415.

(58) Fiehn, O.; Sumner, L. W.; Rhee, S. Y.; Ward, J.; Dickerson, J.; Lange, B. M.; Lane, G.; Roessner, U.; Last, R.; Nikolau, B. Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics.* **2007**, *3*, 195-201.

(59) Moritz, T.; Johansson, A. I., Plant Metabolomics. In *Metabolomics, Metabonomics and Metabolite Profiling*, Griffiths, W. J. (Hrsg.), The Royal Society of Chemistry, Cambridge, **2008**; S. 254-272.

(60) Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155-171.

(61) Ernst, M.; Silva, D. B.; Silva, R. R.; Vencio, R. Z.; Lopes, N. P. Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. *Nat. Prod. Rep.* **2014**, *31*, 784-806.

(62) Schmitke, J. L.; Wescott, C. R.; Klibanov, A. M. The mechanistic dissection of the plunge in enzymatic activity upon transition from water to anhydrous solvents. *J. Am. Chem. Soc.* **1996**, *118*, 3360-3365.

(63) Weckwerth, W.: *Metabolomics: Methods and Protocols*. Springer Science & Business Media, Berlin Heidelberg, **2007**.

(64) Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K. Metabolomics: back to basics. *Phytochem. Rev.* **2008**, *7*, 525-537.

(65) Wolfender, J.-L.; Rudaz, S.; Hae Choi, Y.; Kyong Kim, H. Plant metabolomics: from holistic data to relevant biomarkers. *Curr. Med. Chem.* **2013**, *20*, 1056-1090.

(66) Yuliana, N. D.; Jahangir, M.; Verpoorte, R.; Choi, Y. H. Metabolomics for the rapid dereplication of bioactive compounds from natural sources. *Phytochem. Rev.* **2013**, *12*, 293-304.

(67) Martin, A. C.; Pawlus, A. D.; Jewett, E. M.; Wyse, D. L.; Angerhofer, C. K.; Hegeman, A. Evaluating solvent extraction systems using metabolomics approaches. *RSC Adv.* **2014**.

(68) Tambellini, N. P.; Zaremberg, V.; Turner, R. J.; Weljie, A. M. Evaluation of extraction protocols for simultaneous polar and non-polar yeast metabolite analysis using multivariate projection methods. *Metabolites.* **2013**, *3*, 592-605.

(69) Theodoridis, G.; Gika, H. G.; Wilson, I. D. LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *TrAC, Trends Anal. Chem.* **2008**, *27*, 251-260.

(70) Luo, B.; Groenke, K.; Takors, R.; Wandrey, C.; Oldiges, M. Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. *J. Chromatogr.* **2007**, *1147*, 153-164.

(71) Dettmer, K.; Aronov, P. A.; Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.

(72) Qi, X.; Chen, X.; Wang, Y.: *Plant Metabolomics: Methods and Applications*. Springer, Berlin Heidelberg, **2014**.

(73) Kuang, H.; Li, Z.; Peng, C.; Liu, L.; Xu, L.; Zhu, Y.; Wang, L.; Xu, C. Metabonomics approaches and the potential application in foodsafety evaluation. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 761-774.

(74) Vernocchi, P.; Vannini, L.; Gottardi, D.; Del Chierico, F.; Serrazanetti, D. I.; Ndagijimana, M.; Guerzoni, M. E. Integration of datasets from different analytical techniques

to assess the impact of nutrition on human metabolome. *Front. Cell. Infect. Microbiol.* 2012, 2, 156.

(75) Ibáñez, C.; García-Cañas, V.; Valdés, A.; Simó, C. Novel MS-based approaches and applications in food metabolomics. *TrAC, Trends Anal. Chem.* **2013**, *52*, 100-111.

(76) Ellis, D. I.; Dunn, W. B.; Griffin, J. L.; Allwood, J. W.; Goodacre, R. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*. **2007**, *8*, 1243-1266.

(77) Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263-269.

(78) Kaddurah-Daouk, R.; Kristal, B. S.; Weinshilboum, R. M. Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 653-683.

(79) Luykx, D. M. A. M.; Van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **2008**, *107*, 897-911.

(80) Zhao, Y.; Zhang, B.; Chen, G.; Chen, A.; Yang, S.; Ye, Z. Recent developments in application of stable isotope analysis on agro-product authenticity and traceability. *Food Chem.* **2014**, *145*, 300-305.

(81) Rossmann, A. Determination of stable isotope ratios in food analysis. *Food Rev. Int.* **2001**, *17*, 347-381.

(82) Förstel, H.; Boner, M.; Zahnen, J., *Projektabschlussberich: Überprüfung der Herkunftsdeklaration von Holz mittels Isotopenverteilung*, WWF Deutschland (Hrsg.), Frankfurt am Main, **2008**.

(83) Locatelli, M.; Coisson, J. D.; Travaglia, F.; Bordiga, M.; Arlorio, M. Impact of roasting on identification of hazelnut (*Corylus avellana* L.) origin: A chemometric approach. *J. Agric. Food Chem.* **2015**.

(84) Moscetti, R.; Radicetti, E.; Monarca, D.; Cecchini, M.; Massantini, R. Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin. *J. Sci. Food Agric.* **2014**.

(85) Ciarmiello, L. F.; Mazzeo, M. F.; Minasi, P.; Peluso, A.; De Luca, A.; Piccirillo, P.; Siciliano, R. A.; Carbone, V. Analysis of different European hazelnut (*Corylus avellana* L.) cultivars: authentication, phenotypic features and phenolic profiles. *J. Agric. Food Chem.* **2014**.

(86) Caligiani, A.; Coisson, J. D.; Travaglia, F.; Acquotti, D.; Palla, G.; Palla, L.; Arlorio, M. Application of <sup>(1)</sup>H NMR for the characterisation and authentication of "Tonda Gentile Trilobata" hazelnuts from Piedmont (Italy). *Food Chem.* **2014**, *148*, 77-85.

(87) Sciubba, F.; Di Cocco, M. E.; Gianferri, R.; Impellizzeri, D.; Mannina, L.; De Salvador, F. R.; Venditti, A.; Delfini, M. Metabolic profile of different Italian cultivars of hazelnut (*Corylus avellana*) by nuclear magnetic resonance spectroscopy. *Nat. Prod. Res.* **2014**, 1-7.

(88) Locatelli, M.; Coïsson, J. D.; Travaglia, F.; Cereti, E.; Garino, C.; D'Andrea, M.; Martelli, A.; Arlorio, M. Chemotype and genotype chemometrical evaluation applied to authentication and traceability of "Tonda Gentile Trilobata" hazelnuts from Piedmont (Italy). *Food Chem.* **2011**, *129*, 1865-1873.

(89) Oddone, M.; Aceto, M.; Baldizzone, M.; Musso, D.; Osella, D. Authentication and traceability study of hazelnuts from piedmont, Italy. *J. Agric. Food Chem.* **2009**, *57*, 3404-3408.

(90) Hoffmann, J. F.; Carvalho, I. R.; Barbieri, R. L.; Rombaldi, C. V.; Chaves, F. C. Butia spp. (*Arecaceae*) LC-MS-based metabolomics for species and geographical origin discrimination. *J. Agric. Food Chem.* **2017**.

(91) Chae, Y. K.; Kim, S. H. Discrimination of rice products by geographical origins and cultivars by two-dimensional NMR spectroscopy. *Bull. Korean Chem. Soc.* **2016**, *37*, 1612-1617.

(92) D'Urso, G.; Maldini, M.; Pintore, G.; d'Aquino, L.; Montoro, P.; Pizza, C. Characterisation of *Fragaria vesca* fruit from Italy following a metabolomics approach through integrated mass spectrometry techniques. *LWT--Food Sci. Technol.* **2016**, *74*, 387-395.

(93) Marquetti, I.; Link, J. V.; Lemes, A. L. G.; Scholz, M. B. d. S.; Valderrama, P.; Bona, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Comput. Electron. Agric.* **2016**, *121*, 313-319.

(94) Gil-Solsona, R.; Raro, M.; Sales, C.; Lacalle, L.; Díaz, R.; Ibáñez, M.; Beltran, J.; Sancho, J. V.; Hernández, F. J. Metabolomic approach for extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography–quadrupole time-of-flight mass spectrometry. *Food Control.* **2016**.

(95) Arana, V. A.; Medina, J.; Esseiva, P.; Pazos, D.; Wist, J. Classification of coffee beans by GC-C-IRMS, GC-MS, and <sup>(1)</sup>H-NMR. *J. Anal. Methods Chem.* **2016**, *2016*, 8564584.

(96) Tomita, S.; Nemoto, T.; Matsuo, Y.; Shoji, T.; Tanaka, F.; Nakagawa, H.; Ono, H.; Kikuchi, J.; Ohnishi-Kameyama, M.; Sekiyama, Y. A NMR-based, non-targeted multistep metabolic profiling revealed L-rhamnitol as a metabolite that characterised apples from different geographic origins. *Food Chem.* **2015**, *174*, 163-172.

(97) Tingting, S.; Xiaobo, Z.; Jiyong, S.; Zhihua, L.; Xiaowei, H.; Yiwei, X.; Wu, C. Determination geographical origin and flavonoids content of goji berry using near-infrared spectroscopy and chemometrics. *Food Anal. Methods.* **2015**, *9*, 68-79.

(98) Li, L.; Lu, X.; Zhao, J.; Zhang, J.; Zhao, Y.; Zhao, C.; Xu, G. Lipidome and metabolome analysis of fresh tobacco leaves in different geographical regions using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2015**.

(99) Lee, J. E.; Lee, B. J.; Chung, J. O.; Kim, H. N.; Kim, E. H.; Jung, S.; Lee, H.; Lee, S. J.; Hong, Y. S. Metabolomic unveiling of a diverse range of green tea (*Camellia sinensis*) metabolites dependent on geography. *Food Chem.* **2015**, *174*, 452-459.

(100) Hori, K.; Kiriyama, T.; Tsumura, K. A liquid chromatography time-of-flight mass spectrometry-based metabolomics approach for the discrimination of cocoa beans from different growing regions. *Food Anal. Methods.* **2015**, *9*, 738-743.

(101) Sciubba, F.; Capuani, G.; Di Cocco, M. E.; Avanzato, D.; Delfini, M. Nuclear magnetic resonance analysis of water soluble metabolites allows the geographic discrimination of pistachios (*Pistacia vera*). *Food Res. Int.* **2014**, *62*, 66-73.

(102) Diaz, R.; Pozo, O. J.; Sancho, J. V.; Hernandez, F. Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Chem.* **2014**, *157*, 84-93.

(103) Bondia-Pons, I.; Savolainen, O.; Törrönen, R.; Martinez, J. A.; Poutanen, K.; Hanhineva, K. Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Res. Int.* **2014**, *63*, 132-138.

(104) Bassbasi, M.; De Luca, M.; Ioele, G.; Oussama, A.; Ragno, G. Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. *J. Food Compos. Anal.* **2014**, *33*, 210-215.

(105) De Vos, J.; Broeckhoven, K.; Eeltink, S. Advances in ultrahigh-pressure liquid chromatography technology and system design. *Anal. Chem.* **2016**, *88*, 262-278.

(106) Van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. v. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* **1956**, *5*, 271-289.

(107) Gritti, F.; Leonardis, I.; Abia, J.; Guiochon, G. Physical properties and structure of fine core–shell particles used as packing materials for chromatography: Relationships between particle characteristics and column performance. *J. Chromatogr. A.* **2010**, *1217*, 3819-3843.

(108) Dejaegher, B.; Mangelings, D.; Vander Heyden, Y. Method development for HILIC assays. J. Sep. Sci. 2008, 31, 1438-1448.

(109) Dunn, W. B.; Ellis, D. I. Metabolomics: Current analytical platforms and methodologies. *TrAC, Trends Anal. Chem.* **2005**, *24*, 285-294.

(110) Rojo, D.; Barbas, C.; Rupérez, F. J. LC-MS metabolomics of polar compounds. *Bioanalysis.* **2012**, *4*, 1235-1243.

(111) Buszewski, B.; Noga, S. Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique. *Anal. Bioanal. Chem.* **2012**, *402*, 231-247.

(112) Naidong, W. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. *J. Chromatogr. B.* **2003**, 796, 209-224.

(113) Spagou, K.; Tsoukali, H.; Raikos, N.; Gika, H.; Wilson, I. D.; Theodoridis, G. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *J. Sep. Sci.* **2010**, *33*, 716-727.

(114) Pesek, J. J.; Matyska, M. T.; Fischer, S. M.; Sana, T. R. Analysis of hydrophilic metabolites by high-performance liquid chromatography-mass spectrometry using a silica hydride-based stationary phase. *J. Chromatogr. A.* **2008**, *1204*, 48-55.

(115) Liu, Z.; Rochfort, S. Recent progress in polar metabolite quantification in plants using liquid chromatography-mass spectrometry. *J. Integr. Plant Biol.* **2014**, n/a-n/a.

(116) Sandoval, J. E.; Pesek, J. J. Synthesis and characterization of a hydride-modified porous silica material as an intermediate in the preparation of chemically bonded chromatographic stationary phases. *Anal. Chem.* **1989**, *61*, 2067-2075.

(117) Borges, E. M. Silica, hybrid silica, hydride silica and non-silica stationary phases for liquid chromatography. *J. Chromatogr. Sci.* **2015**, *53*, 580-597.

(118) Bawazeer, S.; Sutcliffe, O. B.; Euerby, M. R.; Watson, D. G. A comparison of the chromatographic properties of silica gel and silicon hydride modified silica gels. *J. Chromatogr. A.* **2012**, *1263*, 61-67.

(119) Kostiainen, R.; Kauppila, T. J. Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. *J. Chromatogr. A.* **2009**, *1216*, 685-699.

(120) de Hoffmann, E.; Stroobant, V.: *Mass spectrometry: Principles and applications*. 3. Auflage, John Wiley & Sons, New York, **2013**.

(121) Watson, J. T.; Sparkman, O. D.: *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation.* John Wiley & Sons, New York, **2013**.

(122) Hommerson, P.; Khan, A. M.; de Jong, G. J.; Somsen, G. W. Ionization techniques in capillary electrophoresis-mass spectrometry: Principles, design, and application. *Mass Spectrom. Rev.* **2011**, *30*, 1096-1120.

(123) Nguyen, S.; Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. *Proc. Natl. Acad. Sci.* **2007**, *104*, 1111-1117.

(124) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular beams of macroions. *J. Chem. Phys.* **1968**, *49*, 2240.

(125) Iribarne, J. V.; Thomson, B. A. On the evaporation of small ions from charged droplets. J. Chem. Phys. 1976, 64, 2287.

(126) Annesley, T. M. Ion suppression in mass spectrometry. *Clinical Chemistry*. **2003**, *49*, 1041-1044.

(127) Draper, J.; Lloyd, A. J.; Goodacre, R.; Beckmann, M. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics.* **2013**, *9*, S4-S29.

(128) King, R.; Bonfiglio, R.; Fernandez-Metzler, C.; Miller-Stein, C.; Olah, T. Mechanistic investigation of ionization suppression in electrospray ionization. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 942-950.

(129) Baldwin, M. A. Mass spectrometers for the analysis of biomolecules. *Methods Enzymol.* 2005, 402, 3-48.

(130) Perez, J. J.; Flanigan, P. M.; Karki, S.; Levis, R. J. Laser electrospray mass spectrometry minimizes ion suppression facilitating quantitative mass spectral response for multi-component mixtures of proteins. *Anal. Chem.* **2013**, *85*, 6667-6673.

(131) Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C.; O'Donnell, C. J.; Carr, S. A.; Mootha, V. K.; Florez, J. C.; Souza, A.; Melander, O.; Clish, C. B.; Gerszten, R. E. Metabolite profiles and the risk of developing diabetes. *J. Nat. Med.* **2011**, *17*, 448-453.

(132) Rodgers, M. T.; Ervin, K. M.; Armentrout, P. B. Statistical modeling of collisioninduced dissociation thresholds. *J. Chem. Phys.* **1997**, *106*, 4499.

(133) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, *4*.

(134) Guo, B.; Chen, B.; Liu, A.; Zhu, W.; Yao, S. Liquid chromatography-mass spectrometric multiple reaction monitoring-based strategies for expanding targeted profiling towards quantitative metabolomics. *Curr. Drug Metab.* **2012**, *13*, 1226-1243.

(135) El-Aneed, A.; Cohen, A.; Banoub, J. Mass spectrometry, review of the basics: Electrospray, MALDI, and commonly used mass analyzers. *Appl. Spectrosc. Rev.* **2009**, *44*, 210-230.

(136) Wang, X.; Wang, S.; Cai, Z. The latest developments and applications of mass spectrometry in food-safety and quality analysis. *TrAC, Trends Anal. Chem.* **2013**, *52*, 170-185.

(137) Gross, J. H.; Beifuss, K.: *Massenspektrometrie: Ein Lehrbuch*. Springer, Berlin Heidelberg New York, **2012**.

(138) Marchese, S.; Gentili, A.; Perret, D.; Ascenzo, G. D.; Pastori, F. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of non-steroidal antiinflammatory drugs in surface water by liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 879-886.

(139) Xiao, J. F.; Zhou, B.; Ressom, H. W. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC, Trends Anal. Chem.* **2012**, *32*, 1-14.

(140) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001**, *36*, 849-865.

(141) Rubert, J.; Zachariasova, M.; Hajslova, J. Advances in high-resolution mass spectrometry based on metabolomics studies for food--a review. *Food Addit. Contam., Part A.* **2015,** *32*, 1685-1708.

(142) Griffiths, W. J.; Wang, Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem. Soc. Rev.* **2009**, *38*, 1882-1896.

(143) Oldiges, M.; Lutz, S.; Pflug, S.; Schroer, K.; Stein, N.; Wiendahl, C. Metabolomics: Current state and evolving methodologies and tools. *Appl. Microbiol. Biotechnol.* **2007**, *76*, 495-511.

(144) Bowen, B. P.; Northen, T. R. Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1471-1476.

(145) Williams, M. L.; Wainer, I. W. Role of chiral chromatography in therapeutic drug monitoring and in clinical and forensic toxicology. *Ther. Drug Monit.* **2002**, *24*, 290-296.

(146) Liland, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC, Trends Anal. Chem.* **2011**, *30*, 827-841.

(147) Kessler, W.: *Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik.* Wiley VCH, Weinheim, **2007**.

(148) Katajamaa, M.; Oresic, M. Data processing for mass spectrometry-based metabolomics. J. Chromatogr. A. 2007, 1158, 318-328.

(149) Zhao, Y.; Zhao, J.; Zhao, C.; Zhou, H.; Li, Y.; Zhang, J.; Li, L.; Hu, C.; Li, W.; Peng, X.; Lu, X.; Lin, F.; Xu, G. A metabolomics study delineating geographical location-associated primary metabolic changes in the leaves of growing tobacco plants by GC-MS and CE-MS. *Sci. Rep.* **2015**, *5*, 16346.

(150) Ma, X. D.; Fan, Y. X.; Jin, C. C.; Wang, F.; Xin, G. Z.; Li, P.; Li, H. J. Specific targeted quantification combined with non-targeted metabolite profiling for quality evaluation of *Gastrodia elata* tubers from different geographical origins and cultivars. *J. Chromatogr. A.* **2016**, *1450*, 53-63.

(151) Mehl, F.; Marti, G.; Boccard, J.; Debrus, B.; Merle, P.; Delort, E.; Baroux, L.; Raymo, V.; Velazco, M. I.; Sommer, H.; Wolfender, J. L.; Rudaz, S. Differentiation of lemon essential oil based on volatile and non-volatile fractions with various analytical techniques: a metabolomic approach. *Food Chem.* **2014**, *143*, 325-335.

(152) Souto, U. T. C. P.; Pontes, M. J. C.; Silva, E. C.; Galvão, R. K. H.; Araújo, M. C. U.; Sanches, F. A. C.; Cunha, F. A. S.; Oliveira, M. S. R. UV–Vis spectrometric classification of coffees by SPA–LDA. *Food Chem.* **2010**, *119*, 368-371.

(153) Feudo, G. L.; Macchione, B.; Naccarato, A.; Sindona, G.; Tagarelli, A. The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin. *Food Res. Int.* **2011**, *44*, 781-788.

(154) Godelmann, R.; Fang, F.; Humpfer, E.; Schutz, B.; Bansbach, M.; Schafer, H.; Spraul, M. Targeted and nontargeted wine analysis by <sup>(1)</sup>H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: grape variety, geographical origin, year of vintage. *J. Agric. Food Chem.* **2013**, *61*, 5610-5619.

(155) Teye, E.; Huang, X.; Dai, H.; Chen, Q. Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochim. Acta, Part A.* **2013**, *114*, 183-189.

(156) Vaclavik, L.; Schreiber, A.; Lacina, O.; Cajka, T.; Hajslova, J. Liquid chromatography-mass spectrometry-based metabolomics for authenticity assessment of fruit juices. *Metabolomics*. **2012**, *8*, 793-803.

(157) Uarrota, V. G.; Moresco, R.; Coelho, B.; da Costa Nunes, E.; Peruch, L. A. M.; de Oliveira Neubert, E.; Rocha, M.; Maraschin, M. Metabolomics combined with chemometric

tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chem.* **2014**, *161*, 67-78.

(158) Ren, S.; Hinzman, A. A.; Kang, E. L.; Szczesniak, R. D.; Lu, L. J. Computational and statistical analysis of metabolomics data. *Metabolomics*. **2015**, *11*, 1492-1513.

(159) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Liu, S.; Zhang, Y.; Liang, Y. Recent advances in chemometric methods for plant metabolomics: A review. *Biotechnol. Adv.* **2014**.

(160) Khakimov, B.; Gürdeniz, G.; Engelsen, S. B. Trends in the application of chemometrics to foodomics studies. *Acta Aliment.* **2015**, *44*, 4-31.

(161) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics.* **2006**, *7*, 142.

(162) Gromski, P. S.; Xu, Y.; Hollywood, K. A.; Turner, M. L.; Goodacre, R. The influence of scaling metabolomics data on model classification accuracy. *Metabolomics*. **2014**.

(163) Bro, R.; Smilde, A. K. Centering and scaling in component analysis. J. Chemom. 2003, 17, 16-33.

(164) Hitchcock, C.; Sober, E. Prediction versus accommodation and the risk of overfitting. *Br. J. Philos. Sci.* **2004**, *55*, 1-34.

(165) Dietterich, T. Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*. **1995**, *27*, 326-327.

(166) Domingos, P. A few useful things to know about machine learning. *Commun. ACM.* **2012**, *55*, 78-87.

(167) Babyak, M. A. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* **2004**, *66*, 411-421.

(168) Cai, E., Machine Learning Lesson of the Day - Overfitting and Underfitting, The Chemical Statistician, https://chemicalstatistician.wordpress.com/?s=Machine+Learning+Lesson+of+the+Day+%E2 %80%93+Overfitting+and+Underfitting, veröffentlicht am 19.03.2014.

(169) Srivastava, T., *How to Avoid Over-Fitting Using Regularization*, AnalyticsVidhya, https://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/, veröffentlicht am **08.02.2015**.

(170) Ambrosius, W. T.: Topics in Biostatistics. Humana Press, Totowa, NJ, 2007.

(171) Slinker, B. K.; Glantz, S. A. Multiple linear regression. *Circulation*. **2008**, *117*, 1732-1737.

(172) Montgomery, D. C.; Peck, E. A.; Vining, G. G.: *Introduction to Linear Regression Analysis*. John Wiley & Sons, New York, **2015**.

(173) Jolliffe, I. T.: *Principal Component Analysis*. Springer Science & Business Media, Berlin Heidelberg, **2013**.

(174) Fahrmeir, L.; Brachinger, W.; Hamerle, A.; Tutz, G.: *Multivariate statistische Verfahren.* 2. Auflage, Walter de Gruyter, Berlin, **1996**.

(175) Shapoval, L. *Identifizierung des Primärtumors aus Hirnmetastasen mittels IR-spektroskopischer Methoden und multivariater Statistik*. Dissertation, Technische Universität Dresden, Dresden, 2005.

(176) Seifert, S. Klassifizierung von Pollenproben mit spektroskopischen und spektrometrischen Methoden und multivariater Statistik. Dissertation, Humboldt-Universität zu Berlin, Berlin, 2016.

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

(178) Want, E.; Masson, P., Processing and analysis of GC/LC-MS-based metabolomics data. In *Metabolic Profiling: Methods and Protocols*, Metz, T. O., Ed. Springer Science and Business Media: New York, 2011; Vol. 708, pp 277-298.

(179) Wold, S.; Sjöström, M., SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In *Chemometrics: Theory and Application*, Kowalski, B. R. (Hrsg.), ACS Publications, Washington D.C., **1977**; Vol. 52, S. 243–282.

(180) Oliveri, P.; Downey, G. Multivariate class modeling for the verification of foodauthenticity claims. *TrAC, Trends Anal. Chem.* **2012**, *35*, 74-86.

(181) Richter, T. *Einsatz der FT-IR-Mikrospektroskopie und multivariater Auswertealgorithmen zur Identifizierung und Klassifizierung von Tumorgeweben*. Dissertation, Technische Universität Dresden, Dresden, **2002**.

(182) Sun, D. W.: *Infrared Spectroscopy for Food Quality Analysis and Control*. Academic Press, Amsterdam, Boston, **2009**.

(183) Henrion, R.; Henrion, G.: *Multivariate Datenanalyse: Methodik und Anwendung in der Chemie und verwandten Gebieten.* Springer, Berlin Heidelberg New York, **2013**.

(184) Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Hum. Genet.* **1936**, *7*, 179-188.

(185) Webb, A. R.: Statistical Pattern Recognition. John Wiley & Sons, New York, 2003.

(186) Massart, D. L.; Vandeginste, B. G. M.: *Handbook of Chemometrics and Qualimetrics*. Elsevier, Amsterdam, **1998**.

(187) Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung.* 14. Auflage, Springer Gabler, Berlin Heidelberg, **2015**.

(188) Bishop, C. M.: *Pattern Recognition and Machine Learning*. Springer, Berlin Heidelberg, **2006**.

(189) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Anal. Chim. Acta.* **2016**, *914*, 17-34.

(190) Vapnik, V. N.: Statistical Learning Theory. Wiley, New York, 1998.

(191) Xu, Y.; Zomer, S.; Brereton, R. G. Support vector machines: A recent method for classification in chemometrics. *Crit. Rev. Anal. Chem.* **2006**, *36*, 177-188.

(192) Hsu, C.-W.; Lin, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415-425.

(193) Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121-167.

(194) Klockmann, S.; Reiner, E.; Bachmann, R.; Hackl, T.; Fischer, M. Food fingerprinting: Metabolomic approaches for geographical origin discrimination of hazelnuts (*Corylus avellana*) by UPLC-QTOF-MS. *J. Agric. Food Chem.* **2016**, *64*, 9253-9262.

(195) Klockmann, S.; Reiner, E.; Cain, N.; Fischer, M. Food targeting: Geographical origin determination of hazelnuts (*Corylus avellana*) by LC-QqQ-MS/MS-based targeted metabolomics application. *J. Agric. Food Chem.* **2017**, *65*, 1456-1465.

(196) Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2012**, *403*, 1523-1548.

(197) Duportet, X.; Aggio, R. B. M.; Carneiro, S.; Villas-Bôas, S. G. The biological interpretation of metabolomic data can be misled by the extraction method used. *Metabolomics*. **2011**, *8*, 410-421.

(198) Wu, X.; Li, N.; Li, H.; Tang, H. An optimized method for NMR-based plant seed metabolomic analysis with maximized polar metabolite extraction efficiency, signal-to-noise ratio, and chemical shift consistency. *Analyst.* **2014**, *139*, 1769-1778.

(199) Mushtaq, M. Y.; Choi, Y. H.; Verpoorte, R.; Wilson, E. G. Extraction for metabolomics: access to the metabolome. *Phytochem. Anal.* **2014**, *25*, 291-306.

(200) dos Santos, R. R.; Moreira, D. M.; Kunigami, C. N.; Aranda, D. A. G.; Teixeira, C. M. L. L. Comparison between several methods of total lipid extraction from *Chlorella vulgaris* biomass. *Ultrason. Sonochem.* **2015**, *22*, 95-99.

(201) Bligh, E. G.; Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911-917.

(202) Folch, J.; Lees, M.; Sloane-Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissue. *J. Biol. Chem.* **1957**, *226*, 497-509.

(203) Shiva, S.; Vu, H. S.; Roth, M. R.; Zhou, Z.; Marepally, S. R.; Nune, D. S.; Lushington, G. H.; Visvanathan, M.; Welti, R., Lipidomic Analysis of Plant Membrane Lipids by Direct Infusion Tandem Mass Spectrometry. In *Plant Lipid Signaling Protocols*, Munnik, T.; Heilmann, I. (Hrsg.), Humana Press, Totowa, NJ, **2013**; S. 79-91.

(204) Patterson, A. D.; Maurhofer, O.; Beyoglu, D.; Lanz, C.; Krausz, K. W.; Pabst, T.; Gonzalez, F. J.; Dufour, J. F.; Idle, J. R. Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer Res.* **2011**, *71*, 6590-6600.

(205) Wang, Y.-S.; Yao, H.-Y.; Xue, H.-W. Lipidomic profiling analysis reveals the dynamics of phospholipid molecules in *Arabidopsis thaliana* seedling growth. *J. Integr. Plant Biol.* **2016**.

(206) Li, W.; Zhang, J.; Tse, F. L. S.: *Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations.* John Wiley & Sons, New York, **2013**.

(207) Callahan, D. L.; De Souza, D.; Bacic, A.; Roessner, U. Profiling of polar metabolites in biological extracts using diamond hydride-based aqueous normal phase chromatography. *J. Sep. Sci.* **2009**, *32*, 2273-2280.

(208) Choi, M.-Y.; Choi, W.; Park, J. H.; Lim, J.; Kwon, S. W. Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. *Food Chem.* **2010**, *121*, 1260-1268.

(209) Song, H. H.; Kim, D. Y.; Woo, S.; Lee, H. K.; Oh, S. R. An approach for simultaneous determination for geographical origins of Korean *Panax ginseng* by UPLC-QTOF/MS coupled with OPLS-DA models. *J. Ginseng Res.* **2013**, *37*, 341-348.

(210) Gika, H.; Theodoridis, G. Sample preparation prior to the LC-MS-based metabolomics/metabonomics of blood-derived samples. *Bioanalysis*. **2011**, *3*, 1647-1661.

(211) Michopoulos, F.; Lai, L.; Gika, H.; Theodoridis, G.; Wilson, I. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *J. Proteome Res.* **2009**, *8*, 2114-2121.

(212) Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CV), *Guidance for Industry: Bioanalytical Method Validation*, US Department of Health and Human Services, Washington D. C., **2001**.

(213) Deutsches Institut für Normung (DIN), DIN 32645:2008-11 Chemical Analysis - Decision Limit, Detection Limit and Determination Limit Under Repeatability Conditions - Terms, Methods, Evaluation, 2008.

(214) Savage, G. P.; McNeil, D. L.; Dutta, P. C. Lipid composition and oxidative stability of oils in hazelnuts (*Corylus avellana* L.) grown in New Zealand. *J. Am. Oil Chem. Soc.* **1997**, 74, 755-759.

(215) Benitez-Sánchez, P. L.; León-Camacho, M.; Aparicio, R. A comprehensive study of hazelnut oil composition with comparisons to other vegetable oils, particularly olive oil. *Eur. Food Res. Technol.* **2003**, *218*, 13-19.

(216) Alasalvar, C.; Amaral, J. S.; Satır, G.; Shahidi, F. Lipid characteristics and essential minerals of native Turkish hazelnut varieties (*Corylus avellana* L.). *Food Chem.* **2009**, *113*, 919-925.

(217) Ayyildiz, H. F.; Topkafa, M.; Kara, H.; Sherazi, S. T. H. Evaluation of fatty acid composition, tocols profile, and oxidative stability of some fully refined edible oils. *Int. J. Food Prop.* **2015**, *18*, 2064-2076.

(218) Vujević, P.; Petrović, M.; Vahčić, N.; Milinović, B.; Čmelik, Z. Lipids and minerals of the most represented hazelnut varieties cultivated in Croatia. *Ital. J. Food Sci.* **2014**, *26*, 24-30.

(219) Köksal, A. İ.; Artik, N.; Şimşek, A.; Güneş, N. Nutrient composition of hazelnut (*Corylus avellana* L.) varieties cultivated in Turkey. *Food Chem.* **2006**, *99*, 509-515.

(220) Amaral, J. S.; Casal, S.; Alves, M. R.; Seabra, R. M.; Oliveira, B. P. P. Tocopherol and tocotrienol content of hazelnut cultivars grown in Portugal. *J. Agric. Food Chem.* **2006**, *54*, 1329-1336.

(221) Esche, R.; Müller, L.; Engel, K.-H. Online LC-GC-based analysis of minor lipids in various tree nuts and peanuts. *J. Agric. Food Chem.* **2013**, *61*, 11636-11644.

(222) Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B.* **2008**, *871*, 299-305.

(223) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B.* **1995**, 289-300.

(224) Devlin, B.; Roeder, K.; Wasserman, L. False discovery or missed discovery? *Heredity*. **2003**, *91*, 537-538.

(225) Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*. **2012**, *4*, 2249-2264.

(226) Diedrich, M.; Henschel, K. P. The natural occurrence of unusual fatty acids. Part 1. Odd numbered fatty acids. *Mol. Nutr. Food Res.* **1990**, *34*, 935-943.

(227) Rezanka, T.; Sigler, K. Odd-numbered very-long-chain fatty acids from the microbial, animal and plant kingdoms. *Prog. Lipid Res.* **2009**, *48*, 206-238.

(228) Holčapek, M.; Jandera, P.; Zderadička, P.; Hrubá, L. Characterization of triacylglycerol and diacylglycerol composition of plant oils using high-performance liquid chromatography–atmospheric pressure chemical ionization mass spectrometry. *J. Chromatogr. A.* **2003**, *1010*, 195-215.

(229) Kalo, P. J.; Ollilainen, V.; Rocha, J. M.; Malcata, F. X. Identification of molecular species of simple lipids by normal phase liquid chromatography–positive electrospray tandem mass spectrometry, and application of developed methods in comprehensive analysis of low erucic acid rapeseed oil lipids. *Int. J. Mass Spectrom.* **2006**, *254*, 106-121.

(230) Brydwell, W. C.: Modern Methods for Lipid Analysis by Liquid Chromatography: Mass Spectrometry and Related Techniques. AOCS Press, Urbana, IL, **2005**.

(231) Lampi, A.-M.; Nurmi, T.; Ollilainen, V.; Piironen, V. Tocopherols and tocotrienols in wheat genotypes in the healthgrain diversity screen. *J. Agric. Food Chem.* **2008**, *56*, 9716-9721.

(232) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. **2007**, *3*, 211-221.

(233) Deutscher Wetterdienst, *Klimadaten Deutschland und Klimadaten weltweit - Monatswerte*, http://www.dwd.de/, Zugriff am **22.02.2017**.

(234) World Meteorological Organization - World Weather Information Service, *Climatological Information*, http://worldweather.wmo.int/en/home.html, Zugriff am **22.02.2017**.

(235) Holcapek, M.; Jirasko, R.; Lisa, M. Recent developments in liquid chromatographymass spectrometry and related techniques. *J. Chromatogr. A.* **2012**, *1259*, 3-15.

(236) Picó, Y.; Blasco, C.; Farr, M.; Barcel, D. Analytical utility of quadrupole time-of-flight mass spectrometry for the determination of pesticide residues in comparison with an optimized column high-performance liquid chromatography/tandem mass spectrometry method. *J. AOAC Int.* **2009**, *92*, 734-744.

(237) Di Stefano, V.; Avellone, G.; Bongiorno, D.; Cunsolo, V.; Muccilli, V.; Sforza, S.; Dossena, A.; Drahos, L.; Vekey, K. Applications of liquid chromatography-mass spectrometry for food analysis. *J. Chromatogr. A.* **2012**, *1259*, 74-85.

(238) Kitteringham, N. R.; Jenkins, R. E.; Lane, C. S.; Elliott, V. L.; Park, B. K. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B.* **2009**, *877*, 1229-1239.

(239) Wasserstein, R. L.; Lazar, N. A. The ASA's statement on p-values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129-133.

(240) Kıralan, S.; Yorulmaz, A.; Şimşek, A.; Tekin, A. Classification of Turkish hazelnut oils based on their triacylglycerol structures by chemometric analysis. *Eur. Food Res. Technol.* **2014**.

(241) Gómez-Ariza, J. L.; Arias-Borrego, A.; García-Barrera, T.; Beltran, R. Comparative study of electrospray and photospray ionization sources coupled to quadrupole time-of-flight mass spectrometer for olive oil authentication. *Talanta*. **2006**, *70*, 859-869.

(242) Lerma-García, M. J.; Lusardi, R.; Chiavaro, E.; Cerretani, L.; Bendini, A.; Ramis-Ramos, G.; Simó-Alfonso, E. F. Use of triacylglycerol profiles established by high performance liquid chromatography with ultraviolet–visible detection to predict the botanical origin of vegetable oils. *J. Chromatogr. A.* **2011**, *1218*, 7521-7527.

(243) Chiavaro, E.; Vittadini, E.; Rodriguez-Estrada, M. T.; Cerretani, L.; Bendini, A. Differential scanning calorimeter application to the detection of refined hazelnut oil in extra virgin olive oil. *Food Chem.* **2008**, *110*, 248-256.

(244) Alasalvar, C.; Pelvan, E.; Topal, B. Effects of roasting on oil and fatty acid composition of Turkish hazelnut varieties (*Corylus avellana* L.). *Int. J. Food Sci. Nutr.* **2010**, *61*, 630-642.

(245) Ebrahem, K. S. *Factors Influencing Storage Quality of Hazelnut Varieties*. Dissertation, Oregon State University, Corvallis, OR, **1992**.

(246) Baldwin, B.; Guisard, Y., *The status and future challenges for the Australian hazelnut industry*, VIII International Congress on Hazelnut, Grau Beretta, P.; Ellena, M., (Hrsg.), *Acta Hort.*, Temuco City, Chile, **2012**; S. 321-327.

(247) Paulson, J. K., *Global Agricultural Information Network (GAIN) Report - 2014 Turkey Tree Nuts Annual*, USDA Foreign Agricultural Service, **2014**.

(248) Julian, J.; Seavert, C.; Olsen, J. L., *An economic evaluation of the impact of eastern filbert blight resistant hazelnut cultivars in Oregon, USA*, VII International Congress on Hazelnut, Varvaro, L.; Franco, S., (Hrsg.), *Acta Hort.*, Viterbo, Italien, **2008**; S. 725-732.

(249) Sathuvalli, V. R.; Mehlenbacher, S. A.; Peterschmidt, B. C.; Smith, D. C., *Current status of eastern filbert blight resistance sources and mapping*, VIII International Congress on Hazelnut, Grau Beretta, P.; Ellena, M., (Hrsg.), *Acta Hort.*, Temuco City, Chile, **2012**; S. 27-33.

(250) Steinbeck, C.; Conesa, P.; Haug, K.; Mahendraker, T.; Williams, M.; Maguire, E.; Rocca-Serra, P.; Sansone, S.-A.; Salek, R. M.; Griffin, J. L. MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics*. **2012**, *8*, 757-760.

(251) Field, D.; Sansone, S.-A.; Collis, A.; Booth, T.; Dukes, P.; Gregurick, S. K.; Kennedy, K.; Kolar, P.; Kolker, E.; Maxon, M. 'Omics data sharing. *Science*. **2009**, *326*, 234-236.

(252) Chervitz, S. A.; Deutsch, E. W.; Field, D.; Parkinson, H.; Quackenbush, J.; Rocca-Serra, P.; Sansone, S.-A.; Stoeckert, C. J.; Taylor, C. F.; Taylor, R. Data standards for omics data: the basis of data sharing and reuse. *Methods Mol. Biol.* **2011**, *719*, 31-69.

(253) Mandel, J.: The statistical analysis of experimental data. Interscience, New York, 1964.

(254) Breusch, T. S.; Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*. **1979**, *47*, 1287-1294.

(255) Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA), *GESTIS-Stoffdatenbank - Gefahrstoffinformationssystem der Deutschen Gesetzlichen Unfallversicherung*, www.dguv.de/ifa/stoffdatenbank, Zugriff am **09.01.2017**.

(256) Europäische Chemikalienagentur (ECHA), *Datenbank des C&L-Verzeichnisses*, https://echa.europa.eu/de/information-on-chemicals/cl-inventory-database, Zugriff am **09.01.2017**.

(257) Graf, U.; Henning, H. J.; Stange, K.; Wilrich, P. T.: Formeln und Tabellen der angewandten mathematischen Statistik. 3. Auflage, Springer, Berlin Heidelberg, **1997**.

(258) Roy, I.; Gupta, M. N. Freeze-drying of proteins: Some emerging concerns. *Biotechnol. Appl. Biochem.* **2004**, *39*, 165-177.

(259) Zhang, Y.; Li, F.; Huang, F.; Xie, G.; Wei, R.; Chen, T.; Liu, J.; Zhao, A.; Jia, W. Metabolomics analysis reveals variation in Schisandra chinensis cetabolites from different origins. *J. Sep. Sci.* **2014**, *37*, 731-737.

(260) Murata, N.; Los, D. A. Membrane fluidity and temperature perception. *Plant Physiol.* **1997**, *115*, 875-879.

(261) Creek, D. J.; Jankevics, A.; Breitling, R.; Watson, D. G.; Barrett, M. P.; Burgess, K. E. Toward global metabolomics analysis with hydrophilic interaction liquid chromatographymass spectrometry: improved metabolite identification by retention time prediction. *Anal. Chem.* **2011**, *83*, 8703-8710.

(262) Ellis, J. K.; Athersuch, T. J.; Thomas, L. D.; Teichert, F.; Pérez-Trujillo, M.; Svendsen, C.; Spurgeon, D. J.; Singh, R.; Järup, L.; Bundy, J. G.; Keun, H. C. Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.* **2012**, *10*, 61.

(263) Zang, Q.; Keire, D. A.; Wood, R. D.; Buhse, L. F.; Moore, C. M. V.; Nasr, M.; Al-Hakim, A.; Trehy, M. L.; Welsh, W. J. Determination of galactosamine impurities in heparin samples by multivariate regression analysis of their <sup>1</sup>H-NMR spectra. *Anal. Bioanal. Chem.* **2012**, *399*, 635-649.

# 8 Anhang

#### 8.1 Material

#### 8.1.1 Chemikalienverzeichnis

Im folgenden Verzeichnis sind die verwendeten Chemikalien tabellarisch mit entsprechender GHS-Gefahrenbeurteilung aufgeführt.

 Tabelle
 8-1:
 Verwendete
 Chemikalien
 mit
 CAS-Nummer
 und
 Gefahrstoffinformationen
 gemäß

 Produktspezifikationen der Hersteller bzw.
 Chemikalien-Datenbanken.<sup>255, 256</sup>
 255
 256

Substanz	CAS- Nummer	GHS- Symbol	Hazard Statement	Precaution Statement		
Acetonitril	75-05-8	02, 07	225, 302, 312, 319, 332	210, 280, 305+351+338		
Ameisensäure	64-18-6	02, 05	226, 314	280, 305+351+338, 310		
Ammoniak 25 %	7664-41-7	05, 09, 314, 400, 335 07 314, 400, 335 310, 301+330+33		280, 273, 305+351+338, 309, 310, 301+330+331		
Ammoniumformiat	540-69-2	07	315, 319, 335	261, 305+351+338		
Butylhydroxytoluol	128-37-0	07, 09	302, 319, 411	273, 305+351+338		
Chloroform	67-66-3	07, 08	302, 315, 319, 331, 351, 361d, 372	201, 202, 260, 264, 280, 281, 301+312, 302+352, 304+340, 305+351+338, 308+313, 311, 321,		
Ethanol	64-17-5	02, 07	225, 319	210, 233, 305+351+338		
1-Pentadecanoyl-2-oleoyl(d7)-sn- glycero-3-phosphocholine	-	unbekannt				
1-Pentadecanoyl-2-oleoyl(d7)-sn- glycero-3-phosphoethanolamine	-	unbekannt				
1-Pentadecanoyl-2-oleyol(d7)-sn- glycerol	-	unbekannt				
Essigsäure	64-19-7	02,05	226, 314	280, 305+351+338, 310		
n-Hexan	110-54-3	02, 07, 08, 09	225, 304, 315, 336, 361f, 373, 411	210, 240, 273, 301+310, 331, 302+352, 403+235		
Kohlenstoffdioxid, fest, tiefkalt	124-38-9	04 281 282, 336+315, 403		282, 336+315, 403		
1,2-Dilinoleoyl-sn-glycero-3- phosphocholine	998-06-1	kein gefährlicher Stoff nach GHS				
1,2-Dilinoleoyl-sn-glycero-3- phosphoethanolamine	20707-71-5	kein gefährlicher Stoff nach GHS				
1,3-Dilinoleoyl-rac-glycerol	15818-46-9	kein gefährlicher Stoff nach GHS				
1-Linoleoyl-2-hydoxy-sn-glycero- 3-phosphocholine	22252-07-9	kein gefährlicher Stoff nach GHS				
Methanol	67-56-1	02, 06, 08	225, 301, 311, 331, 370	210, 260, 280, 301+310, 307+311		
Natriumhydroxid	1310-73-2	05	314	260, 264, 280, 301+330+331, 303+361+353, 304+340, 305+351+338, 310, 321, 405, 501		
1,2-Dioleoyl-sn-glycerol	2442-61-7	kein gefährlicher Stoff nach GHS				
1-Oleoyl(d7)-2-hydoxy-sn-glycero- 3-phosphocholine	-	unbekannt				
rac-1-Oleoyl-2,3-dipalmitoyl-	1867-91-0	kein gefährlicher Stoff nach GHS				
Substanz	CAS- Nummer	GHS- Symbol	Hazard Statement	Precaution Statement		
--	----------------	----------------------------------	----------------------------------	---	--	--
glycerol						
1,2-Dioleoyl-3-palmitoyl-rac- glycerol	65390-75-2		kein gefährlic	her Stoff nach GHS		
1,2-Dioleoyl-sn- phosphoethanolamine	4004-05-1	07, 08	315, 351, 373	201, 202, 260, 264, 280, 281, 302+352, 308+313		
1-Palmitoyl-2-linoleoyl-sn-glycero- 3-phosphocholine	6931-84-6		kein gefährlicher Stoff nach GHS			
1-Palmitoyl-2-linoleoyl-rac- glycerol-3-phosphoethanolamine	26662-95-3		kein gefährlicher Stoff nach GHS			
1-Palmitoyl-2-oleoyl-sn-glycerol	29541-66-0		kein gefährlic	her Stoff nach GHS		
1-Palmitoyl-2-oleoyl-3-linoleoyl- rac-glycerol	2680-59-3	kein gefährlicher Stoff nach GHS				
1,3-Dipentadecanoyl-2-oleyol(d7)- glycerol	-	unbekannt				
Hexakis(2,2- difluoroethoxy)phosphazen	186817-57-2	kein gefährlicher Stoff nach GHS				
2-Propanol	67-63-0	02, 07	225, 319, 336	210, 233, 305+351+338		
Stickstoff, flüssig, tiefkalt	7727-37-9	04	281	282, 336+315, 403		
α-Tocopherol(d6)	107651-33-2	unbekannt				

Im folgenden Verzeichnis sind die verwendeten Chemikalien inklusive Hersteller und Reinheit tabellarisch aufgeführt.

Substanz	Hersteller	Reinheit
Acetonitril LC-MS grade	Carl Roth, Karlsruhe, Deutschland	≥99,95 %
Ameisensäure	Carl Roth, Karlsruhe, Deutschland	> 99 %
Ammoniak 25 %	Carl Roth, Karlsruhe, Deutschland	> 99,9 %
Ammoniumformiat	Carl Roth, Karlsruhe, Deutschland	> 95 %
Butylhydroxytoluol	Carl Roth, Karlsruhe, Deutschland	≥99,8 %
Chloroform HPLC grade	Carl Roth, Karlsruhe, Deutschland	≥99,9 %
Ethanol HPLC grade	Carl Roth, Karlsruhe, Deutschland	≥99,9 %
1-Pentadecanoyl-2-oleoyl(d7)-sn-glycero-3-	Avanti Polar Lipids,	$\geq$ 99 %
phosphocholine	Alabaster, AL, USA	
1-Pentadecanoyl-2-oleoyl(d7)-sn-glycero-3-	Avanti Polar Lipids,	≥99 %
phosphoethanolamine	Alabaster, AL, USA	
1 Pontadaganavi 2 alaval(d7) sn givaaral	Avanti Polar Lipids,	$\geq$ 99 %
1-1 entadecanoyi-2-oleyoi(u7)-sii-giycei oi	Alabaster, AL, USA	
Essigsäure	Carl Roth, Karlsruhe, Deutschland	≥98 %
n-Hexan	Carl Roth, Karlsruhe, Deutschland	≥99 %
Kohlenstoffdioxid, fest, tiefkalt	Linde Pullach Deutschland	keine
	Linde, Funden, Deutsemand	Angabe
1.2-Dilinoleovl-sn-glycero-3-phosphocholine	Avanti Polar Lipids,	$\geq$ 99 %
	Alabaster, AL, USA	22.24
1.2-Dilinoleovl-sn-glycero-3-phosphoethanolamine	Avanti Polar Lipids,	$\geq$ 99 %
-,	Alabaster, AL, USA	
1,3-Dilinoleovl-rac-glycerol	Cayman Chemical,	$\geq 95 \%$
	Ann Arbor, MI, USA	
1-Linoleoyl-2-hydoxy-sn-glycero-	Echelon Biosciences,	keine
3-phosphocholine	Salt Lake City, UI, USA	Angabe
Methanol LU-MS grade	Carl Roth, Karlsruhe, Deutschland	<u>299,95 %</u>
Natriumhydroxid	Carl Roth, Karlsruhe, Deutschland	<u>≥ 99 %</u>
1,2- Dieleevl en glygerel	Sigma-Aldrich, München, Deutschland	<u>≥</u> 9/‰
Dioleoyi-sii-giyceroi		

Substanz	Hersteller	Reinheit
Bubstanz	Avanti Polar Lipids	> 99 %
1-Oleoyl(d7)-2-hydoxy-sn-glycero-3-phosphocholine	Alabaster, AL, USA	_ // /0
	Toronto Research Chemicals,	98 %
rac-1-Oleoyi-2,3-dipalmitoyi-giycerol	Toronto, ON, Kanada	
1.2 Dialagy 3 nalmitayl rag glygaral	Toronto Research Chemicals,	97 %
1,2-Dioleoyi-3-painitoyi-rac-giyceroi	Toronto, ON, Kanada	
1,2-Dioleoyl-sn-phosphoethanolamine	Sigma-Aldrich, München, Deutschland	$\geq$ 97 %
1 Delmiterel 2 linelaard en ekseene 3 nhoenhoeheline	Avanti Polar Lipids,	$\geq$ 99 %
1-Pannitoy1-2-infoleoy1-sii-giycero-3-phosphocholine	Alabaster, AL, USA	
1-Palmitoyl-2-linoleoyl-rac-glycerol-3-	Cayman Chemical,	≥98 %
phosphoethanolamine	Ann Arbor, MI, USA	
	Avanti Polar Lipids,	≥99 %
1-Paimitoyi-2-oleoyi-sn-giycerol	Alabaster, AL, USA	
1 Delucitored 2 closed 2 lineland use showed	Cayman Chemical,	≥85 %
1-Paimitoyi-2-oleoyi-5-iinoleoyi-rac-giyceroi	Ann Arbor, MI, USA	
1.2 Dimente de com cul 2 claus (d7) chucanal	Avanti Polar Lipids,	≥99 %
1,5-Dipentadecanoyi-2-oleyoi(d7)-giyceroi	Alabaster, AL, USA	
Herely's (2.) diffuence the subheast here.	Santa Cruz Biotechnology,	keine
Hexakis(2,2-dilluoroetnoxy)phosphazen	Dallas, TX, USA	Angabe
2-Propanol LC-MS grade	Carl Roth, Karlsruhe, Deutschland	≥ <u>99,95 %</u>
Stielestoff flüssig tiefkelt	Linda Bullach Doutschland	keine
Suckston, nussig, ueikait	Linde, Pullacii, Deutschiand	Angabe
a-Tocopherol(d6)	Sigma-Aldrich, München, Deutschland	> 99,9 %

# 8.1.2 Verzeichnis über verwendete Lösungen und Puffer

Im folgenden Verzeichnis sind die verwendeten Lösungen und Puffer tabellarisch aufgeführt.

Tabelle 8-3: Verwendete Lösungen und Puffer.

Lösung/Puffer	Hersteller/Zusammensetzung			
Lookmassa	1 mg/mL Hexakis(2,2-difluoroethoxy)phosphazen in 2-			
Lockmasse	Propanol			
Natriumformiat Clustor	Wasser/Isopropanol/1 M NaOH/Ameisensäure (50/50/1/0,1,			
Natriumorimat-Cluster	v/v/v/v)			
400 mM Ammoniumformiat-Puffer zur	12,6 g Ammoniumformiat ad 500 mL mit Wasser auffüllen,			
Erstellung der Elutionsmittel	mit Ameisensäure auf pH 3,5 einstellen			
Extraktionsmittel unpolare Extraktion	Chloroform/Isopropanol (2/1, v/v) + 0,01 % BHT			
Extraktionsmittel polare Extraktion	Wasser/Methanol (1/2, v/v)			
	1,70 mM DG(15:0/18:1d7), 1,89 mM PC(18:1d7/0:0),			
Interner Standard Mir	1,33 mM PC(15:0/18:1d7), 1,44 mM PE(15:0/18:1d7),			
interner Standard Mix	1,26 mM TG(15:0/18:1d7/15:0) und 1,38 mM α-Tocopherol-			
	D6 in Methanol			

# 8.1.3 Verzeichnung über verwendete Geräte und Software

Im folgenden Verzeichnis sind die verwendeten Geräte und Software tabellarisch aufgeführt.

	Gerät	Hersteller	Spezifikationen	
	Analysenwaage	Kern & Sohn, Balingen, Deutschland	<u>Kern ABJ 220-4M</u> Maximalgewicht: 220 g Genauigkeit: 0,1 mg	
Gefriertrockner		Martin Christ, Osterode am Harz, Deutschland	Gamma 1-20	
	Handknackmaschine	Feucht Obsttechnik	Wal Man	
Triple Quadrupol		AB Sciex,	<u>API2000</u>	
-	Massenspektrometer	Foster City, CA, USA	mit ESI-Ionenquelle	
HPLC-ESI-MS/MS System	HPLC	Agilent Technologies, Santa Clara, CA, USA	Agilent 1260 Infinity Series Quaternary Pump (integrated Degasser): G1311B Autosampler: G1329B Thermostated Column Compartement: G1316A	
	Steuerungs- und Auswer- tungssoftware	AB Sciex, Foster City, CA, USA	Analyst Version 1.6	
Kolbenhubpipetten		Eppendorf, Hamburg, Deutschland	$\frac{\text{Research (variabel)}}{V = 0, 1-2, 5 \ \mu\text{L}}$ $V = 0, 5-10 \ \mu\text{L}$ $V = 10-100 \ \mu\text{L}$ $V = 100-1000 \ \mu\text{L}$	
Kugelmühle		Biolabproducts, Bebensee, Germany	<u>BeadRuptor 24</u> mit 2 mL Eppendorf Safelock Carriage kit	
		Quiagen, Venlo, Niederlande	TissueLyser	
	Kühl-/Gefrierschränke	verschiedene Hersteller	<u>4 °C, -20 °C, -40 °C, -80 °C</u>	
Messermühle		Retsch, Haan, Deutschland	Grindomix GM300 Mahlbehälter aus rostfreiem Stahl, Messer aus rostfreiem Stahl (Vollmetall)	
	NIR-Spektroskop	Bruker Optik, Ettlingen, Deutschland	<u>Tango</u> Software: OPUS Version 7.5	
pH-Meter		Mettler-Toledo, Gießen, Deutschland	<u>SevenEasy Meter S20K</u> , pH/mV/Temperatur Meter, Inlab 413 Electrode	
		CAMO Software, Oslo, Norwegen	The Unscrambler X, Version 10.3	
Software		Microsoft Corporation, Redmond WA USA	Microsoft Office Professional Plus 2010	
	OriginLab Corporation, Northampton, MA, USA	OriginPro 9.0.0G		
		Bruker Daltronik,	ProfileAnalysis 2.1	
		Bremen, Deutschland	QuantAnalysis 2.1	
	Ultraschallgerät	Bandelin Electronic, Berlin, Deutschland	Sonopuls HD 2200	
	Vortex	Bender und Hobein, Zürich, Schweiz	Vortex Genie 2	

# Tabelle 8-4: Spezifikationen der verwendeten Geräte und Software.

		- B		
	Gerät	Hersteller	Spezifikationen	
Wasseraufbereitungssystem		Merck Millipore, Darmstadt, Deutschland	Direct-Q 3 UV-R	
Zentrifugalmühle		Retsch, Haan, Deutschland	ZM1, 2 mm Rundlochung	
Zentrifuge		Sartorius Stedim Biotech, Göttingen, Deutschland	SIGMA Laborzentrifuge 1- 14	
Quadrupol-Time-of-Flight Massenspektrometer		Bruker Daltronik, Bremen, Deutschland	<u>maXis</u> mit ESI-Ionenquelle	
ıF-MS-System	UPLC	Thermo Fischer Scientific, Braunschweig, Deutschland	Dionex UltiMate 3000 Pump: HPG-3200RS Degasser: SRD-3200 Autosampler: WPS- 3000TRS Column Compartement: TCC-3000RS	
TQ-IS		Bruker Daltronik, Bremen, Deutschland Thermo Fischer Scientific	otof control Version 3.4	
UPLC-F	Steuerungs- und Auswer-	Braunschweig, Deutschland	Chromeleon Xpress Version 6.80	
	tungssonware	Bruker Daltronik, Bremen, Deutschland	HyStar Version 3.2	
		Bruker Daltronik, Bremen, Deutschland	DataAnalysis 4.1	

# 8.1.4 Verzeichnis über verwendete Verbrauchsmaterialien

Im folgenden Verzeichnis sind die verwendeten Verbrauchsmaterialien tabellarisch aufgeführt.

Verbrauchsmaterial	Hersteller	Spezifikationen		
Bördelkappen	Macherey und Nagel, Düren, Deutschland	N 11 Aluminium Bördelkappen, silber, Loch, Naturkautschuk/Butyl rot- orange/TEF farblos, Dicke 1 mm		
Einmalspatel	Carl Roth, Karlsruhe, Deutschland	Einmal-Kunststoff-Spatel Mikro, antistatisch, unsteril		
Einmalspritzen	Henke Sass Wolf, Tuttlingen, Deutschland	NORM-JECT, 2teilig, latex- und siliconölfrei, 1 mL		
Glasvials	Macherey und Nagel, Düren, Deutschland	Vials, V = 1,5 mL, 11,6 x 32 mm Braun, flacher Boden, enge Öffnung		
Mikroinlets für Vials	Macherey und Nagel, Düren, Deutschland	Mikroeinsätze für Flaschen mit enger Öffnung, klar, konisch, 15 mm Spitze $V = 100 \ \mu L$ , 5 x 31 mm		
Pipettenspitzen	Eppendorf, Hamburg, Deutschland	<u>epT.I.P.S.</u> 0,1–10 μL, 34 mm 2–200 μL, 53 mm 50–1000 μL, 71 mm		

Tabelle 8-5: Spezifikationen der verwendeten Verbrauchsmaterialien.

Verbrauchsmaterial	Hersteller	Spezifikationen		
Reaktionsgefäße	Eppendorf, Hamburg, Deutschland	Eppendorf Safe-Lock Gefäße, Standard $V = 2 \text{ mL}$		
Schraubdeckelröhren	Sarstedt, Nürnbrecht, Deutschland	Schraubröhre Spitzboden, unsteril, PP, V = 15  mL, 120 x 17 mm V = 50  mL, 114 x 28 mm		
	Carl Roth, Karlsruhe, Deutschland	Rotilabo, PTFE, unsteril, Porengröße 0,45 μm, Ø 0,15 mm		
Spritzenvorsatzfilter	Hahnemühle FineArt GmbH, Dasser, Deutschland	Regenerierte Cellulose (RC), unsteril, Porengröße 0,45 μm, Ø 0,13 mm; PTFE, unsteril, Porengröße 0,45 μm, Ø 0,13 mm		
	Macherey und Nagel, Düren, Deutschland	Chromafil Xtra, PTFE, Porengröße 0,45 μm, Ø 0,13 mm		
Stahlkugeln, entfettet	Fahrradgebrauchtmarkt, Hamburg, Deutschland	Radlagerkugeln, Edelstahl, Ø 3 mm, entfettet mittels SOXHLET-Extraktion mit n-Hexan		

# 8.1.5 Formeln zur statistischen Auswertung

Im Folgenden sind die verwendeten Formeln jeweils mit einer kurzen Erläuterung zur Bedeutung der Größen aufgeführt. Die Einheiten sind nur bei den Formeln mit angegeben, die ausschließlich für die Berechnung einer Art von Daten verwendet wurden.

# Test auf Normalverteilung nach DAVID

Der Test auf Normalverteilung nach DAVID gibt Auskunft darüber, ob bei einer vorgegebenen Wahrscheinlichkeit bei den erhaltenen Messwerten eine Normalverteilung vorliegt, welche für diverse statistische Verfahren Grundvoraussetzung ist. Ist der errechnete Prüfwert kleiner als der entsprechende Tabellenwert (Tabelle für den Test auf Normalverteilung nach DAVID)<sup>257</sup>, liegt bei den Messwerten mit einer Wahrscheinlichkeit von 95 % eine Normalverteilung vor.

$$PG = \frac{V}{s}$$
 Gleichung 8-1

PG: Prüfgröße ( $\alpha = 0,05$ )

V: Variationsbreite (Differenz zwischen dem höchsten und niedrigsten Wert)

s: Standardabweichung

#### Ausreißertest nach DEAN und DIXON

Ein Ausreißer stellt einen Wert dar, der sich stark von den übrigen Werten der Reihe unterscheidet, sodass es sich nicht mehr um eine statistische Abweichung handelt. Voraussetzung für die Durchführung des Ausreißertests ist die Normalverteilung der Messwerte, welche aus mehr als 3 Werten bestehen müssen. Ist der errechnete Prüfwert größer als der entsprechende Tabellenwert (Tabelle zum Ausreißertest nach DEAN und DIXON)<sup>257</sup>, handelt es sich bei dem Messwert mit einer Wahrscheinlichkeit von 95 % um einen Ausreißer. Zur Ermittlung des Tabellenwertes gilt der Freiheitsgrad f = n - 2.

$$PG = \frac{|x_1 - x_2|}{|x_1 - x_n|}$$
 Gleichung 8-2

PG: Prüfgröße ( $\alpha = 0,05$ )

- x<sub>1</sub>: kleinster bzw. größter Wert
- $x_2$ : benachbarter Wert von  $x_1$
- $x_n$ : weit entferntester Wert von  $x_1$

# Lineare Regression

Die lineare Regression stellt die Abhängigkeit zweier Variablen über das Prinzip der kleinsten Fehlerquadrate in einen linearen Zusammenhang. Die daraus resultierende Regressionsgerade kann durch Gleichung 8-3 beschrieben werden.

- y: Flächenintegral des Signals [counts]
- m: Steigung der Regressionsgeraden (Empfindlichkeit) [counts/µM]
- x: Konzentration der Probe [µM]
- b: Achsenabschnitt der Regressionsgeraden (mittleres Begleitsignal) [counts]

# Quadratische Regression

Die quadratische Regression stellt die Abhängigkeit zweier Variablen über das Prinzip der kleinsten Fehlerquadrate in einen quadratischen Zusammenhang. Die daraus resultierende Regressionskurve kann durch Gleichung 8-4 beschrieben werden.

$$y = a \cdot x^2 + b \cdot x + c$$
 Gleichung 8-4

- y: Flächenintegral des Signals [counts]
- a: Steigung der Regressionskurve [counts/ $\mu$ M<sup>2</sup>]
- x: Konzentration der Probe [µM]
- b: x-Achsenabschnitt der Regressionskurve [counts/ $\mu$ M<sup>1</sup>]
- c: y-Achsenabschnitt der Regressionskurve [counts]

# Reststandardabweichung der linearen Regression

$$S_{R} = \sqrt{\frac{\sum (y_{i} - \hat{y}_{L,i})^{2}}{n-2}}$$
Gleichung 8-5

- S<sub>R</sub>: Reststandardabweichung der linearen Regression [counts]
- yi: Gemessene Peakfläche an i-ter Stelle [counts]
- ŷ: Berechnete Peakfläche mittels Kalibrierfunktion der linearen Regression an i-ter Stelle [counts]
- n: Anzahl der Messwerte

# Reststandardabweichung der quadratischen Regression

$$S_{RQ} = \sqrt{\frac{\sum (y_i - \hat{y}_{Q,i})^2}{n-3}}$$
 Gleichung 8-6

- $S_{RQ}$ : Reststandardabweichung der quadratischen Regression [counts]
- yi: Gemessene Peakfläche an i-ter Stelle [counts]
- ŷ<sub>i</sub>: Berechnete Peakfläche mittels Kalibrierfunktion der quadratischen Regression an i-ter Stelle [counts]
- n: Anzahl der Messwerte

# Ausreißertest für lineare Funktionen

Ein Ausreißer in der linearen Regression stellt einen Wert dar, der sich stark von den übrigen Werten der Reihe unterscheidet, sodass es sich nicht mehr um eine statistische Abweichung handelt. Ist der errechnete Prüfgröße größer als der entsprechende Tabellenwert (Tabelle zur F-Verteilung)<sup>257</sup>, handelt es sich bei dem Messwert mit einer Wahrscheinlichkeit von 99 % um einen Ausreißer. Zur Ermittlung des Tabellenwertes gilt für die Freiheitsgrade  $f_1 = 1$  und  $f_2 = n - 3$ .

PG = 
$$\frac{((n-2) \cdot S_{R1}^2 - (n-3) \cdot S_{R2}^2)}{S_{R2}^2}$$

**Gleichung 8-7** 

PG: Prüfgröße ( $\alpha = 0,01$ )

n: Anzahl der Kalibrierpunkte

S<sub>R1</sub>: Reststandardabweichung der linearen Regression mit dem potentiellen Ausreißer [counts]

S<sub>R2</sub>: Reststandardabweichung der linearen Regression ohne den potentiellen Ausreißer [counts]

#### Linearitätstest nach MANDEL

Der Linearitätstest nach MANDEL prüft, ob die lineare Regression signifikant besser ist als die quadratische Regression und somit angewendet werden kann. Liegt die berechnete Prüfgröße (PG) unterhalb des entnommenen Tabellenwertes (Tabelle zur F-Verteilung)<sup>257</sup>, so ist die lineare Regression mit einer Wahrscheinlichkeit von 99 % der quadratischen Regression vorzuziehen. Zur Ermittlung des Tabellenwertes gilt für die Freiheitsgrade  $f_1 = 1$  und  $f_2 = n - 3$ .

$$PG = \frac{((N-2) \cdot S_R^2 - (N-3) \cdot S_{RQ}^2)}{S_{RQ}^2}$$

**Gleichung 8-8** 

PG: Prüfgröße

n: Anzahl der Kalibrierpunkte

S<sub>R</sub>: Reststandardabweichung der linearen Regression [counts]

S<sub>RQ</sub>: Reststandardabweichung der quadratischen Regression [counts]

#### Bestimmtheitsmaß

$$R^{2} = \left( \frac{\sum x_{i} \overline{x} \cdot \overline{x} \cdot (y_{i} \overline{y})}{\sqrt{\sum (x_{i} \overline{x})^{2} \cdot (y_{i} \overline{y})^{2}}} \right)^{2}$$

**Gleichung 8-9** 

- R<sup>2</sup>: Bestimmtheitsmaß
- x<sub>i</sub>: Konzentration an i-ter Stelle [µM]
- $\bar{x}$ : Mittelwert der Konzentrationen [ $\mu$ M]
- y<sub>i</sub>: Peakfläche an i-ter Stelle [counts]
- y:Mittelwert der Peakfläche [counts]

#### Verfahrensstandardabweichung

 $S_v = \frac{S_R}{m}$  Gleichung 8-10

Sv: Verfahrensstandardabweichung [µM]

S<sub>R</sub>: Reststandardabweichung [counts]

m: Empfindlichkeit (Steigung der Regressionsgeraden) [counts/µM]

#### Verfahrensvariationskoeffizient

- $VK_V = \frac{s_v}{\bar{x}} \cdot 100 \%$  Gleichung 8-11
- VK<sub>V</sub>: Verfahrensvariationskoeffizient [%]
- $S_V$ : Verfahrensstandardabweichung [ $\mu$ M]
- $\overline{x}$ : Mittelwert der Konzentrationen [ $\mu$ M]

# Wiederholgrenze

 $r = S_v \cdot 2$ 

**Gleichung 8-12** 

r: Wiederholgrenze [µM]

 $S_V$ : Verfahrensstandardabweichung [ $\mu$ M]

#### Nachweisgrenze (Leerwertmethode)

$$NWG = \frac{s \cdot t \cdot \sqrt{\frac{1}{p} + \frac{1}{n}}}{m}$$
 Gleichung 8-13

NWG: Nachweisgrenze [µM]

- s: Standardabweichung [counts]
- t: Tabellenwert aus einseitiger t-Tabelle für n-1 (Tabelle zur Student-t-Verteilung)<sup>257</sup>
- p: Anzahl an Parallelbestimmungen der Kalibrierreihe
- n: Anzahl unabhängig hergestellter Blindproben
- m: Empfindlichkeit (Steigung der Regressionsgeraden) [counts/µM]

#### 

 $CV = \frac{s}{\bar{x}} \cdot 100 \%$  Gleichung 8-14

CV: Variationskoeffizient bzw. Präzision [%]

s: Standardabweichung [µM]

 $\bar{x}$ : Mittelwert [ $\mu$ M]

#### **Richtigkeit**

$$Acc = \left| \frac{c_i - \hat{c}_i}{\hat{c}_i} \right| \cdot 100 \%$$

Acc: Richtigkeit [%]

ci: Berechnete Konzentration an i-ter Stelle [counts]

 $\hat{c}_i$ : Tatsächliche Konzentration an i-ter Stelle [counts]

### Berechnung des Korrekturfaktors zur Quantifizierung mittels internem Standard

$$KF_{M,i} = \frac{A_{M,i} \cdot c_{IS,i}}{A_{IS,i} \cdot c_{M,i}}$$
Gleichung 8-16

KF<sub>M,i</sub>: Korrekturfaktor des jeweiligen Metabolits bei i-ter Konzentration []

A<sub>M,i</sub>: Peakfläche des Metabolits im i-ten Standard der Grundkalibrierung [counts]

c<sub>IS,i</sub>: Konzentration der internen Standardsubstanz im i-ten Standard der Grundkalibrierung[µM]

A<sub>IS,i</sub>: Peakfläche der internen Standardsubstanz im i-ten Standard der Grundkalibrierung [counts]

 $c_{M,i}$ : Konzentration des jeweiligen Metabolits im i-ten Standard der Grundkalibrierung [ $\mu$ M]

#### Berechnung der Konzentrationen der einzelnen Schlüsselmetabolite in den Haselnussproben

$$G = \frac{A_{M} \cdot c_{IS} \cdot M \cdot V \cdot KF}{A_{IS} \cdot E}$$
 Gleichung 8-17

G: Gehalt des jeweiligen Metabolits in der Haselnussprobe [mg/kg]

A<sub>M</sub>: Peakfläche des Metabolits [counts]

- c<sub>IS</sub>: Konzentration der internen Standardsubstanz im Extrakt [µM]
- M: Molare Masse des jeweiligen Metabolits [g/mol]
- V: Probenvolumen [mL]
- KF: Korrekturfaktor []
- A<sub>IS</sub>: Peakfläche der internen Standardsubstanz [counts]
- E: Probeneinwaage (hier: 50 mg)

**Gleichung 8-15** 

# F-Test auf Varianzenhomogenität

Voraussetzung für den F-Test auf Varianzenhomogenität ist die Normalverteilung der Werte. Ist die errechnete Prüfgröße kleiner als der entsprechende Tabellenwert (Tabelle zur F-Verteilung)<sup>257</sup>, sind die Varianzen der beiden Messreihen mit einer Wahrscheinlichkeit von 99 % homogen. Zur Ermittlung des Tabellenwertes gilt für die Freiheitsgrade  $f_1 = n_a$  (Anzahl der Messwerte in Messreihe A) und  $f_2 = n_b$  (Anzahl der Messwerte in Messreihe B).

$$PG = \frac{S_{Ra}^2}{S_{Rb}^2}$$
 Gleichung 8-18

mit  $S_{Ra}^2 > S_{Rb}^2$ 

PG: Prüfgroße ( $\alpha = 0,01$ )

 $S_{Ra}$ : Standardabweichung der Grundgesamtheit A

S<sub>Rb</sub>: Standardabweichung der Grundgesamtheit B

Berechnung der gescaleten Werte (Interquartile Range Scaling)

$$\widetilde{\mathbf{x}}_{ij} = \frac{\mathbf{x}_{ij} - \mathbf{x}_{i_{median}}}{IQR_i}$$

**Gleichung 8-19** 

 $\boldsymbol{\tilde{x}}_{ij}$ : Wert nach dem Scaling von Metabolit i

x<sub>ii</sub>: Wert vor dem *Scaling* von Metabolit i

 $\dot{x_{i_{median}}}$ : Median Wert des authentischen Probensatzes von Metabolit i

IQR<sub>i</sub>: Interquartile des authentischen Probensatzes von Metabolit i



8.2 Ergänzungen zum Abschnitt 5 – Ergebnisse und Diskussion

Abbildung 8-1: Beispielchromatogramm eines repräsentativen Haselnussextraktes, gemessen mit NTNPneg.



Abbildung 8-2: Beispielchromatogramm eines in Chloroform gelösten Niederschlags einer mit Chloroform/Methanol (1/2, v/v) extrahierten Haselnussprobe (auch repräsentativ für Niederschläge der anderen Extraktionsmittel sowie für flüssige in Chloroform gelöste, ölige Phasen entsprechender Extrakte), gemessen in NTNP-pos.



Abbildung 8-3: Chromatogramme eines mit Chloroform/Isopropanol (1/2, v/v) extrahierten Blindwertes, (A) ohne und (B) mit Membranfiltration (Rotilabo, PTFE, 0,45 μm, 0,13 mm).



Abbildung 8-4: Beispielchromatogramm eines repräsentativen Haselnussextraktes, gemessen mit NTP-pos.



Abbildung 8-5: Beispielchromatogramm eines repräsentativen Haselnussextraktes, gemessen mit NTPneg.

Tabelle	8-6:	Getestete	Gradienten	in	Zuge	der	Methodenentwicklung	für	die	chromatog	raphische
Trennur	ng voi	n polaren N	Aetaboliten r	nit (	der noi	1-tar	geted Methode; Eluent A	<b>1</b> = I	I2O -	+ 5 mM AF;	Eluent B
= ACN/I	IPA 1	0/1 + 5  mM	I AF.								

	Gradient-	nt1		Gradient-	nt2	Gradient-nt3		
Zeit	% Eluent A	% Eluent B	Zeit	% Eluent A	% Eluent B	Zeit	% Eluent A	% Eluent B
0	0	100	0	0	100	0	0	100
6	0	100	2	0	100	6	0	100
8	40	60	8	20	80	8	15	85
11	40	60	13	20	80	12,5	15	85
13	100	0	14	100	0	13,5	100	0
17	100	0	17	100	0	17	100	0
19	0	100	19	0	100	19	0	100
23	0	100	23	0	100	23	0	100
	Gradient-nt4			Gradient-	nt5	Gradient-nt6		
Zeit	% Eluent A	% Eluent B	Zeit	% Eluent A	% Eluent B	Zeit	% Eluent A	% Eluent B
0	0	100	0	0	100	0	0	100
2	0	100	2	0	100	2	0	100
5	15	85	4	10	90	5	15	85
8	15	85	13	10	90	7	15	85
10	50	50	14	100	0	9	50	50
12	50	50	17	100	0	10,5	50	50
13	100	0	19	0	100	11,5	100	0
17	100	0	23	0	100	15	100	0
18	0	100				16	0	100
23	0	100				20	0	100

	Gradient	t-t1*		Gradient-t2 (is	okratisch) <sup>*</sup>
Zeit	% Eluent A	% Eluent B	Zeit	% Eluent A	% Eluent B
0	90 (85, 75,, 5)	10 (15, 25,, 95)	0	25 (20, 15,, 0)	75 (80, 85,, 100)
2	90 (85, 75,, 5)	10 (15, 25,, 95)	37	25 (20, 15,, 0)	75 (80, 85,, 100)
15	0	100			
30	0	100			
31	90 (85, 75,, 5)	10 (15, 25,, 95)			
37	90 (85, 75,, 5)	10 (15, 25,, 95)			

Tabelle 8-7: Getestete Gradienten in Zuge der Methodenentwicklung des chromatographischen Systemsder targeted Methode; Eluent  $A = H_2O + 5$  mM AF; Eluent B = IPA + 5 mM AF.

\* verschiedene Anfangs- und Endkonzentrationen

Metabolit	m [counts/uM]	b [counts]	R <sup>2</sup>		[ <b>u</b> M]
PC(18:2/0:0)	$2.61 \cdot 10^5$	$6.09 \cdot 10^4$	0.9976	0.577	- 9.62
DG(16:0/18:1)	8,24·10 <sup>4</sup>	$-1.78 \cdot 10^{5}$	0,9999	3.36 -	- 84,0
DG(18:2/18:2)	8,98·10 <sup>4</sup>	$-3,08 \cdot 10^4$	0,9971	1,94 -	- 58,4
DG(18:1/18:1)	6,56·10 <sup>4</sup>	$-4,21\cdot10^{4}$	0,9961	2,30 ·	- 92,1
PE(16:0/18:2)	9,99·10 <sup>4</sup>	$-4,64 \cdot 10^4$	0,9992	2,15 -	- 21,5
PE(18:2/18:2)	1,01.105	$3,21 \cdot 10^2$	0,9967	0,0400	) - 20,0
PE(18:1/18:1)	7,93·10 <sup>4</sup>	$-4,40\cdot10^{3}$	0,9996	0,398	- 59,7
PC(16:0/18:2)	$4,82 \cdot 10^5$	$5,18 \cdot 10^4$	0,9989	0,400	- 20,0
PC(18:2/18:2)	4,68·10 <sup>5</sup>	$4,35 \cdot 10^4$	0,9989	0,400	- 20,0
γ-Tocopherol	$1,37.10^{2}$	$3,13 \cdot 10^3$	0,9981	15,0 -	- 75,0
TG(16:0/18:1/18:1) <sup>1</sup>	$5,85 \cdot 10^4$	$1,92 \cdot 10^{5}$	0,9974	6,02 ·	- 20,1
TG(16:0/18:1/18:2) <sup>1</sup>	8,03·10 <sup>4</sup>	1,63·10 <sup>5</sup>	0,9963	4,24 ·	- 21,2
PC(18:1d7/0:0)	$4,13.10^{4}$	$4,50.10^{3}$	0,9982	0,118 - 23,7	
DG(15:0/18:1d7)	3,06.104	$-2,64 \cdot 10^{3}$	0,9979	0,106	- 106
PE(15:0/18:1d7)	$3,28.10^{4}$	6,48·10 <sup>3</sup>	0,9980	0,090	- 18,0
PC(15:0/18:1d7)	$7,25 \cdot 10^4$	$-2,45 \cdot 10^{2}$	0,9986	0,0499	9 - 33,3
TG(15:0/18:1d7/15:0)	1,69·10 <sup>5</sup>	$-2,15\cdot10^{3}$	0,9982	0,0315	5 - 6,30
	NWG [nM]	S <sub>R</sub> [counts]	S <sub>V</sub> [μM]	VKv [%]	r [µM]
PC(18:2/0:0)	1,36	$1,36.10^{5}$	0,523	12,0	1,05
DG(16:0/18:1)	3,57	$3,15\cdot10^{5}$	3,82	12,2	7,64
DG(18:2/18:2)	8,69	$1,89.10^{5}$	2,10	11,5	4,20
DG(18:1/18:1)	30,3	$2,32 \cdot 10^{5}$	3,53	11,8	7,07
PE(16:0/18:2)	4,25	$7,13 \cdot 10^4$	0,792	8,85	1,58
PE(18:2/18:2)	2,48	$5,82 \cdot 10^4$	0,571	16,0	1,15
PE(18:1/18:1)	3,69	$1,90.10^{5}$	2,39	18,9	4,79
PC(16:0/18:2)	1,70	$3,22 \cdot 10^5$	0,668	12,2	1,34
PC(18:2/18:2)	1,51	$2,97 \cdot 10^{5}$	0,634	12,0	1,27
γ-Tocopherol	81,5	$1,36 \cdot 10^3$	9,91	21,4	19,8
TG(16:0/18:1/18:1) <sup>1</sup>	978	$7,43 \cdot 10^4$	1,27	11,5	2,54
$TG(16:0/18:1/18:2)^{1}$	734	$7,73 \cdot 10^4$	0,963	9,46	1,93
PC(18:1d7/0:0)	0,200	$1,20.10^4$	0,292	5,57	0,584

Universität Hamburg, Hamburg School of Food Science, 2017

	NWG [nM]	S <sub>R</sub> [counts]	S <sub>V</sub> [μM]	VK <sub>V</sub> [%]	r [µM]
DG(15:0/18:1d7)	0,407	$4,72 \cdot 10^{4}$	1,55	6,97	3,09
PE(15:0/18:1d7)	0,374	$7,60.10^3$	0,232	5,82	0,463
PC(15:0/18:1d7)	0,192	$2,43 \cdot 10^4$	0,335	6,49	0,670
TG(15:0/18:1d7/15:0)	0,384	$1,42 \cdot 10^4$	0,0843	5,98	0,169

m: Empfindlichkeit; b: mittleres Begleitsignal; R<sup>2</sup>: PEARSON'scher Korrelationskoeffizient; LA: linearer Arbeitsbereich; NWG: Nachweisgrenze; S<sub>R</sub>: Reststandardabweichung der linearen Regression; S<sub>V</sub>: Verfahrensstandardabweichung; VK<sub>V</sub>: Verfahrensvariationskoeffizient; r: Wiederholgrenze <sup>1</sup>nur zur Methodenvalidierung verwendet

#### Tabelle 8-9: Validierungsparameter der Matrixkalibrierung. (modifiziert nach Klockmann 2017)

Metabolit	m [counts/µM]	b [counts]	R <sup>2</sup>	LA	μ <b>M</b> ]
PC(18:2/0:0)	2,01.105	1,74.105	0,9963	0,770	- 38,5
DG(16:0/18:1)	7,13.104	1,72.106	0,9944	16,8 -	· 67,2
DG(18:2/18:2)	7,46.104	3,03.106	0,9926	38,9	- 211
DG(18:1/18:1)	2,98.104	1,38.107	0,9784	92,1	- 691
PE(16:0/18:2)	7,76.104	3,27.105	0,9974	2,15 -	43,0
PE(18:2/18:2)	8,78.104	3,01.105	0,9981	8,00 -	40,0
PE(18:1/18:1)	7,46.104	5,83.105	0,9994	7,96 -	. 39,8
PC(16:0/18:2)	4,01.105	3,52.106	0,9951	2,00 -	40,0
PC(18:2/18:2)	3,68.105	3,92.106	0,9842	4,00 ·	40,0
γ-Tocopherol	1,92.102	6,18.102	0,9847	4,50 -	· 15,0
TG(16:0/18:1/18:1) <sup>1</sup>	-	-	-	-	
TG(16:0/18:1/18:2) <sup>1</sup>	-	-	-	-	•
PC(18:1d7/0:0)	2,48.104	1,86.105	0,9986	2,36 -	- 23,6
DG(15:0/18:1d7)	2,14.104	1,99.105	0,9995	4,25	- 106
PE(15:0/18:1d7)	1,43.104	1,07.105	0,9983	1,80 -	- 18,0
PC(15:0/18:1d7)	3,23.104	2,46.105	0,9993	3,32 -	- 49,8
TG(15.0/18.1d7/15.0)	6 66 103	434.104	0 9995	157.	. 78 6
10(15.0/10.10//15.0)	0,00 105	1,51101	0,7775	1,57	70,0
10(15.0/10.10//15.0)	<b>NWG [nM]</b>	S <sub>R</sub> [counts]	Sv [μM]	VKv [%]	r [µM]
PC(18:2/0:0)	<b>NWG [nM]</b> 1,77	$\frac{S_{R} [counts]}{4,24.105}$	Sv [μM]           2,11	<b>VKv [%]</b> 21,1	<b>r [μM]</b> 4,22
PC(18:2/0:0) DG(16:0/18:1)	NWG [nM]           1,77           5,32	SR [counts]           4,24·105           2,13·105	Sv [μM]           2,11           2,99	VKv [%]           21,1           6,81	<b>r [μM]</b> 4,22 5,98
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2)	NWG [nM]           1,77           5,32           12,1	$\begin{array}{c} \textbf{S}_{\textbf{R}} \ \hline \textbf{[counts]} \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline 2, 13 \cdot 105 \\ \hline 8, 61 \cdot 105 \end{array}$	<b>Sv [µM]</b> 2,11 2,99 11,5	VKv [%]           21,1           6,81           12,0	<b>r [μM]</b> 4,22 5,98 23,1
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1)	NWG [nM]           1,77           5,32           12,1           79,4	$\begin{array}{c} \textbf{S}_{\textbf{R}} \ \hline \textbf{[counts]} \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{8}, 61 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \end{array}$	S <sub>V</sub> [μM]           2,11           2,99           11,5           49,8	VKv [%]           21,1           6,81           12,0           15,7	<b>r [µM]</b> 4,22 5,98 23,1 99,6
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2)	NWG [nM]           1,77           5,32           12,1           79,4           4,71	$\begin{array}{c} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline 2, 13 \cdot 105 \\ \hline 8, 61 \cdot 105 \\ \hline 1, 48 \cdot 106 \\ \hline 1, 40 \cdot 105 \end{array}$	S <sub>V</sub> [μM]           2,11           2,99           11,5           49,8           1,81	VKv [%]           21,1           6,81           12,0           15,7           13,0	r [µM] 4,22 5,98 23,1 99,6 3,62
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15	$\begin{array}{r} \hline \textbf{S}_{\textbf{R}} \ \hline \textbf{[counts]} \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{8}, 61 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88	VKv [%]           21,1           6,81           12,0           15,7           13,0           14,8	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2) PE(18:1/18:1)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ 2, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{3}, 61 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57	VKv [%]           21,1           6,81           12,0           15,7           13,0           14,8           12,9	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2) PE(18:1/18:1) PC(16:0/18:2)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{8}, 61 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26	VKv [%]           21,1           6,81           12,0           15,7           13,0           14,8           12,9           17,0	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2) PE(18:1/18:1) PC(16:0/18:2) PC(18:2/18:2)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{3}, 61 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34	VKv [%]           21,1           6,81           12,0           15,7           13,0           14,8           12,9           17,0           15,4	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68
PC(18:2/0:0)           DG(16:0/18:1)           DG(18:2/18:2)           DG(18:1/18:1)           PE(16:0/18:2)           PE(18:2/18:2)           PE(18:1/18:1)           PC(16:0/18:2)           PC(16:0/18:2)           PC(18:2/18:2)           PC(18:2/18:2)           PC(18:2/18:2)           PC(18:2/18:2)           PC(18:2/18:2)           γ-Tocopherol	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10	VKv [%]           21,1           6,81           12,0           15,7           13,0           14,8           12,9           17,0           15,4           25,1	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21
PC(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2) PE(18:1/18:1) PC(16:0/18:2) PC(18:2/18:2) γ-Tocopherol TG(16:0/18:1/18:1) <sup>1</sup>	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \hline \textbf{-} \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10	VKv [%]         21,1         6,81         12,0         15,7         13,0         14,8         12,9         17,0         15,4         25,1	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 -
PC(18:2/0:0)           DG(16:0/18:1)           DG(18:2/18:2)           DG(18:1/18:1)           PE(16:0/18:2)           PE(18:2/18:2)           PE(18:1/18:1)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(18:2/18:2)           γ-Tocopherol           TG(16:0/18:1/18:1)^1           TG(16:0/18:1/18:2)^1	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \end{array}$	Sv [µM]         2,11         2,99         11,5         49,8         1,81         2,88         2,57         2,26         2,34         2,10	VKv [%]         21,1         6,81         12,0         15,7         13,0         14,8         12,9         17,0         15,4         25,1	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 -
РС(18:2/0:0) DG(16:0/18:1) DG(18:2/18:2) DG(18:1/18:1) PE(16:0/18:2) PE(18:2/18:2) PE(18:2/18:2) PE(18:1/18:1) PC(16:0/18:2) PC(18:2/18:2) γ-Tocopherol TG(16:0/18:1/18:1) <sup>1</sup> TG(16:0/18:1/18:2) <sup>1</sup> PC(18:1d7/0:0)	NWG [nM]           1,77           5,32           12,1           79,4           4,15           5,43           2,25           22,2           62,8           -           0,334	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \hline \textbf{-} \\ \hline \textbf{1}, 56 \cdot 104 \\ \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10           -           0,630	VKv [%] 21,1 6,81 12,0 15,7 13,0 14,8 12,9 17,0 15,4 25,1 - 6,40	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 - - 1,26
PC(18:2/0:0)           DG(16:0/18:1)           DG(18:2/18:2)           DG(18:1/18:1)           PE(16:0/18:2)           PE(18:2/18:2)           PE(18:1/18:1)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(18:2/18:2)           Y-Tocopherol           TG(16:0/18:1/18:1) <sup>1</sup> TG(16:0/18:1/18:2) <sup>1</sup> PC(18:1d7/0:0)           DG(15:0/18:1d7)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8           -           0,334           0,582	$\begin{array}{r} \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{S}_{\textbf{R}} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \hline \textbf{1}, 56 \cdot 104 \\ \hline \textbf{4}, 85 \cdot 104 \\ \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10           -           0,630           2,27	VKv [%]         21,1         6,81         12,0         15,7         13,0         14,8         12,9         17,0         15,4         25,1         -         6,40         5,85	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 - 1,26 4,54
PC(18:2/0:0)           DG(16:0/18:1)           DG(18:2/18:2)           DG(18:1/18:1)           PE(16:0/18:2)           PE(18:2/18:2)           PE(18:1/18:1)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(18:2/18:2)           γ-Tocopherol           TG(16:0/18:1/18:1) <sup>1</sup> TG(16:0/18:1/18:2) <sup>1</sup> PC(18:1d7/0:0)           DG(15:0/18:1d7)           PE(15:0/18:1d7)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8           -           0,334           0,582           0,857	$\begin{array}{r} \textbf{S_R [counts]} \\ \hline \textbf{S_R [counts]} \\ \hline \textbf{4,}24 \cdot 105 \\ \hline \textbf{2,}13 \cdot 105 \\ \hline \textbf{2,}13 \cdot 105 \\ \hline \textbf{2,}13 \cdot 105 \\ \hline \textbf{2,}105 \\ \hline \textbf{1,}48 \cdot 106 \\ \hline \textbf{1,}40 \cdot 105 \\ \hline \textbf{2,}53 \cdot 105 \\ \hline \textbf{1,}91 \cdot 105 \\ \hline \textbf{2,}53 \cdot 105 \\ \hline \textbf{3,}50 \cdot 105 \\ \hline \textbf{4,}03 \cdot 102 \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \hline \textbf{1,}56 \cdot 104 \\ \hline \textbf{4,}85 \cdot 104 \\ \hline \textbf{1,}01 \cdot 104 \\ \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10           -           0,630           2,27           0,704	VKv [%]         21,1         6,81         12,0         15,7         13,0         14,8         12,9         17,0         15,4         25,1         -         6,40         5,85         9,41	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 - 1,26 4,54 1,41
PC(18:2/0:0)           DG(16:0/18:1)           DG(18:2/18:2)           DG(18:1/18:1)           PE(16:0/18:2)           PE(18:2/18:2)           PE(18:1/18:1)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(16:0/18:2)           PC(18:2/18:2)           γ-Tocopherol           TG(16:0/18:1/18:1) <sup>1</sup> TG(16:0/18:1/18:2) <sup>1</sup> PC(18:1d7/0:0)           DG(15:0/18:1d7)           PE(15:0/18:1d7)           PC(15:0/18:1d7)	NWG [nM]           1,77           5,32           12,1           79,4           4,71           4,15           5,43           2,25           22,2           62,8           -           0,334           0,582           0,857           0,432	$\begin{array}{r} \textbf{S_R} [ \textbf{counts} ] \\ \hline \textbf{S_R} [ \textbf{counts} ] \\ \hline \textbf{4}, 24 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{2}, 13 \cdot 105 \\ \hline \textbf{1}, 48 \cdot 106 \\ \hline \textbf{1}, 40 \cdot 105 \\ \hline \textbf{2}, 53 \cdot 105 \\ \hline \textbf{1}, 91 \cdot 105 \\ \hline \textbf{9}, 05 \cdot 105 \\ \hline \textbf{8}, 50 \cdot 105 \\ \hline \textbf{4}, 03 \cdot 102 \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \hline \textbf{-} \\ \hline \textbf{1}, 56 \cdot 104 \\ \hline \textbf{4}, 85 \cdot 104 \\ \hline \textbf{1}, 01 \cdot 104 \\ \hline \textbf{3}, 73 \cdot 104 \\ \end{array}$	Sv [µM]           2,11           2,99           11,5           49,8           1,81           2,88           2,57           2,26           2,34           2,10           -           0,630           2,27           0,704           1,15	VKv [%]         21,1         6,81         12,0         15,7         13,0         14,8         12,9         17,0         15,4         25,1         -         6,40         5,85         9,41         6,57	r [µM] 4,22 5,98 23,1 99,6 3,62 5,76 5,13 4,51 4,68 4,21 - 1,26 4,54 1,41 2,31

m: Empfindlichkeit; b: mittleres Begleitsignal; R<sup>2</sup>: PEARSON'scher Korrelationskoeffizient; LA: linearer Arbeitsbereich; NWG: Nachweisgrenze; S<sub>R</sub>: Reststandardabweichung der linearen Regression; S<sub>V</sub>: Verfahrensstandardabweichung; VK<sub>V</sub>: Verfahrensvariationskoeffizient; r: Wiederholgrenze <sup>1</sup>nur zur Methodenvalidierung verwendet

	Grundkalibrierung			Matrixkalibrierung			
Metabolit	c	Präzision	Richtigkeit	с	Präzision	Richtigkeit	
	[µM]	[%]	[%]	[µM]	[%]	[%]	
	0,577	1,94	3,93	0,770	14,2	7,54	
PC(18:2/0:0)	3,85	11,4	10,9	7,70	3,35	18,2	
	9,62	6,05	4,96	38,5	7,89	6,91	
	3,36	9,36	10,7	16,8	4,96	10,0	
DG(16:0/18:1)	16,8	14,4	11,4	50,4	5,77	7,20	
	84,0	7,78	4,10	67,2	2,60	2,56	
	1,94	7,59	7,14	38,9	4,85	16,5	
DG(18:2/18:2)	7,78	16,1	12,8	77,8	3,15	4,17	
	58,4	5,67	5,27	211	9,09	6,67	
	2,30	6,08	5,61	92,1	2,69	12,6	
DG(18:1/18:1)	11,5	11,4	9,45	461	4,39	8,65	
	92,1	5,45	4,40	691	5,03	7,58	
	2,15	16,8	9,74	2,15	7,95	19,6	
PE(16:0/18:2)	6,45	10,4	6,56	10,7	10,5	14,4	
· · · ·	21,5	7,33	6,20	43,0	3,31	2,82	
	0,0400	5,09	6,88	8,00	5,29	10,0	
PE(18:2/18:2)	0,800	11,1	11,5	20,0	12,1	11,8	
· · · · ·	20,0	10,2	8,19	40,0	8,79	7,96	
	0,378	15,0	14,9	7,96	6,74	10,7	
PE(18:1/18:1)	7.96	18.6	14.9	19.9	8.50	14.2	
( / )	59.7	12.0	8.62	39.8	7.29	7.41	
	0.400	9.05	11.0	2.00	3.45	17.8	
PC(16:0/18:2)	2.00	10.5	14.0	9,99	8.49	16.4	
(	20.0	8.03	6.44	40.0	4.24	3.91	
	0.400	8.38	8.47	4.00	5.67	16.7	
PC(18.2/18.2)	4 00	5 22	13.2	20.0	14.2	16.4	
10(10.2/10.2)	20.0	7 44	5.83	40.0	4 27	4 26	
	15.0	10.3	17.4	4.50	11.2	18.3	
v-Tocopherol	45.0	16.5	18.1	6.00	8.13	15.4	
,	75.0	11.0	11.0	15.0	12.3	9.71	
	6.0	12.9	14.4	_		-	
TG(16·0/18·1/18·1)	10.0	10.7	8 91	-	-	-	
(	20.1	7.77	7.42	_	_	_	
	4 24	12.0	16.3	-	-	-	
TG(16·0/18·1/18·2)	8 48	8.17	8 38	-	-	-	
10(10:0,10:1,10:2)	21.2	5.08	4 89	-	-	-	
	0.118	6.62	11.6	2.36	3 28	14.4	
PC(18·1d7/0·0)	2 37	6 71	6 68	9.46	4 14	6 47	
10(10.14//0.0)	23.7	2.07	1.62	23.6	1 49	1.61	
	0.106	19.8	19.2	4 25	3.61	14 2	
DG(15.0/18.1d7)	6 39	2.32	4 35	21.3	3 99	5.17	
DO(10.0/10.10/)	106	2,32	2.84	106	3.62	2.82	
	0.0900	6 55	19.4	1.80	2 37	17.5	
PE(15:0/18:1d7)	1 80	7 43	7 21	7 19	3 74	7 18	
12(10.0/10.10/)	18.0	1 58	1 41	18.0	4 80	5 90	
	0.0499	3.96	143	3 3 2	5 09	14 7	
PC(15.0/18.1d7)	0.831	4 59	13.9	8 30	4 42	7 37	
10(10.0/10.107)	33.3	3.03	2.08	49.8	2 74	217	
	0.0315	2.60	13.7	1 57	2,77	14.3	
TG(15:0/18:1d7/15:0)	0.472	0.870	2.98	157	1 48	5 46	
1 O(15.0/10.10//15.0)	6 30	3 12	2,59	78.6	2.07	2.18	

Tabelle 8-10: Richtigkeits- und Präzisions-Werte der Grund- und Matrixkalibrierung. (modifiziert nach Klockmann 2017)

c = Konzentration

Metabolit	m [counts/µM]	b [counts]	<b>R</b> <sup>2</sup>	LA [Verdünnung]
DG(16:0/16:1)	$3,32 \cdot 10^3$	$1,72 \cdot 10^2$	0,9929	4,00 - unverdünnt
DG(16:0/18:1)	9,96·10 <sup>5</sup>	$-2,96\cdot10^{3}$	0,9992	250,00 - unverdünnt
DG(18:2/18:2)	1,37.107	$-7,75 \cdot 10^3$	0,9992	181,82 - unverdünnt
DG(18:1/18:1)	1,15.107	$2,91 \cdot 10^4$	0,9980	1000,00 - unverdünnt
PC(18:2/0:0)	3,12.105	9,08·10 <sup>3</sup>	0,9972	40,00 - unverdünnt
PC(16:0/18:3)	6,68·10 <sup>5</sup>	3,06·10 <sup>3</sup>	0,9983	400,00 - unverdünnt
PC(16:0/18:2	1,64.107	$1,42 \cdot 10^{5}$	0,9978	142,86 - unverdünnt
PC(18:2/18:2)	$2,58 \cdot 10^{7}$	4,91·10 <sup>5</sup>	0,9961	400,00 - unverdünnt
PC(18:1/18:2)	4,94·10 <sup>7</sup>	$-4,67 \cdot 10^3$	0,9988	1000,00 - unverdünnt
PE(16:0/18:2)	$2,09 \cdot 10^{6}$	$-8,04 \cdot 10^3$	0,9992	400,00 - unverdünnt
PE(18:2/18:2)	$2,04 \cdot 10^{6}$	$1,80.10^4$	0,9980	181,82 - unverdünnt
PE(18:1/18:1)	4,74·10 <sup>5</sup>	$-2,77\cdot10^{3}$	0,9982	100,00 - unverdünnt
γ-Tocopherol	-	-	-	-
TG(2:0/18:2/18:2)	3,01·10 <sup>5</sup>	$-1,56 \cdot 10^2$	0,9990	1000,00 - unverdünnt
TG(2:0/18:1/18:2)	$1,19.10^{6}$	$7,11.10^{2}$	0,9994	1000,00 - unverdünnt
TG(14:0/16:0/18:1)	1,36.106	3,16·10 <sup>3</sup>	0,9742	1000,00 - 100,00
TG(15:0/16:0/18:1)	$2,19 \cdot 10^4$	8,03·10 <sup>2</sup>	0,9955	10,00 - unverdünnt
TG(16:0/16:1/18:1)	4,07.107	$7,01.10^{4}$	0,9911	1000,00 - 100,00
TG(17:1/18:1/18:2)	9,53·10 <sup>6</sup>	$1,22 \cdot 10^4$	0,9938	1000,00 - 100,00
TG(18:2/18:2/18:3)	1,28.107	$4,24 \cdot 10^{3}$	0,9996	1000,00 - 11,76

 Tabelle 8-11: Parameter des linearen Arbeitsbereiches der Verdünnungsreihe eines repräsentativen

 Extrakts. (modifiziert nach Klockmann 2017)

m: Empfindlichkeit; b: mittleres Begleitsignal; R<sup>2</sup>: PEARSON'scher Korrelationskoeffizient; LA: linearer Arbeitsbereich

Tabelle 8	8-12:	<b>P-Values</b>	und	Varaitionskoeffizienter	der	Markersubstanzen,	gemessen	mit	NTNP-pos.
(modifizie	ert na	ch Klockı	mann	2016)					

Matabalit	p-value (	CV [0/]	
Wietabont	2014	2015	
DG(16:0/16:1)	8,10·10 <sup>-7</sup>	8,00·10 <sup>-13</sup>	5,29
DG(16:0/18:1)	3,14.10-9	1,28.10-7	7,43
DG(18:2/18:2)	2,31.10-10	2,31.10-17	7,33
DG(18:1/18:1)	8,97·10 <sup>-12</sup>	2,13.10-4	6,13
PC(18:2/0:0)	1,64.10-8	2,31.10-6	4,72
PC(16:0/18:3)	2,23.10-10	1,61.10-13	8,90
PC(16:0/18:2)	4,34·10 <sup>-7</sup>	6,19·10 <sup>-12</sup>	9,85
PC(18:2/18:2)	3,54·10 <sup>-13</sup>	4,87·10 <sup>-13</sup>	10,36
PC(18:1/18:2)	3,36.10-6	3,08.10-19	7,23
PE(16:0/18:2)	9,01·10 <sup>-9</sup>	4,76.10-14	10,53
PE(18:2/18:2)	8,30·10 <sup>-15</sup>	8,33·10 <sup>-6</sup>	8,17
PE(18:1/18:1)	2,03.10-6	1,54.10-10	8,91
γ-Tocopherol	3,09.10-6	7,13.10-5	6,00
TG(2:0/18:2/18:2)	5,00.10-8	5,59·10 <sup>-11</sup>	7,26
TG(2:0/18:1/18:2)	1,49.10-5	2,66.10-12	8,94

A	nł	nan	ıg
		in	5

Matabalit	p-value (A	CV [0/]	
Metabolit	2014	2015	— UV [70]
TG(14:0/16:0/18:1)	4,21.10-7	1,17.10-3	8,59
TG(15:0/16:0/18:1)	2,50.10-8	5,41.10-3	5,67
TG(16:0/16:1/18:1)	2,77.10-4	3,34.10-11	6,15
TG(17:1/18:1/18:2)	6,76·10 <sup>-8</sup>	2,13.10-10	7,42
TG(18:2/18:2/18:3)	5,18.10-7	7,58.10-5	8,18

CV: Variationskoeffizient

Tabelle 8-13: Ergebnisse der PLS-Regression zur Vorhersage der Anteile in Mischproben als Mittelwerte inkl. Messunsicherheiten.

	Türkei	Italien	Georgien	Frankreich	Spanien	Deutschland
TR-GE 80/20	$100 \pm 20$	-	-	-	-	-
TR-GE 60/40	$56\pm20$	-	$44 \pm 29$	-	-	-
TR-GE 50/50	$46 \pm 17$	-	$24 \pm 24$	-	$30 \pm 25$	-
TR-GE 40/60	$38 \pm 11$	-	$39 \pm 16$	-	$23 \pm 17$	-
TR-GE 20/80	$29 \pm 13$	-	$41 \pm 19$	-	$30\pm20$	-
TR-IT 80/20	$72 \pm 11$	$28 \pm 12$	-	-	-	-
TR-IT 60/40	$57\pm20$	$43 \pm 21$	-	-	-	-
TR-IT 50/50	$48 \pm 16$	$52 \pm 17$	-	-	-	-
TR-IT 40/60	$24 \pm 13$	$56 \pm 14$	-	-	$21 \pm 20$	-
TR-IT 20/80	$21 \pm 17$	$79 \pm 18$	-	-	-	-
IT-GE 80/20	$16 \pm 10$	$39 \pm 10$	$14 \pm 14$	$31 \pm 10$	-	-
IT-GE 60/40	$21 \pm 8$	$20\pm 8$	$20 \pm 11$	$23 \pm 8$	$15 \pm 11$	-
IT-GE 50/50	$18 \pm 18$	-	$47 \pm 26$	$35 \pm 18$	-	-
IT-GE 40/60	$30\pm 8$	$34\pm 8$	-	$17 \pm 8$	$19 \pm 12$	-
IT-GE 20/80	$34 \pm 15$	$17 \pm 15$	$24 \pm 21$	$25 \pm 15$	-	-
TR-FR 80/20	$52 \pm 7$	$10 \pm 7$	-	$27 \pm 7$	$11 \pm 11$	-
TR-FR 60/40	$36 \pm 11$	$18 \pm 11$	-	$28 \pm 11$	$18 \pm 16$	-
TR-FR 50/50	$29\pm9$	$10 \pm 9$	-	$43 \pm 9$	$17 \pm 13$	-
TR-FR 40/60	$25 \pm 14$	$17 \pm 14$	-	$58 \pm 14$	-	-
TR-FR 20/80	-	$51 \pm 19$	-	$49 \pm 19$	-	-
TR-DE 80/20	$34 \pm 15$	$43 \pm 16$	-	$23 \pm 15$	-	-
TR-DE 60/40	$36 \pm 14$	$16 \pm 14$	-	$20 \pm 14$	-	$28 \pm 14$
TR-DE 50/50	$39\pm14$	-	-	$29 \pm 14$	-	$32 \pm 13$
TR-DE 40/60	$24 \pm 22$	$36 \pm 23$	-	-	-	$40 \pm 21$
TR-DE 20/80	-	$34 \pm 28$	-	-	-	$66 \pm 27$
FR-DE 80/20	-	$24 \pm 17$	-	$54 \pm 16$	-	$22 \pm 16$
FR-DE 60/40	-	-	-	$67 \pm 30$	-	$33 \pm 29$
FR-DE 50/50	-	$49 \pm 28$	-	-	-	$51 \pm 26$
FR-DE 40/60	-	-	-	-	-	$100 \pm 50$
FR-DE 20/80	-	-	-	-	-	$100 \pm 48$
FR-IT 80/20	-	$20 \pm 10$	-	$56 \pm 10$	$23 \pm 14$	-
FR-IT 60/40	-	$32 \pm 16$	-	$68 \pm 16$	-	-
FR-IT 50/50	-	$47 \pm 16$	-	$53 \pm 16$	-	-
FR-IT 40/60	-	$48 \pm 21$	-	$52 \pm 21$	-	-
FR-IT 20/80	-	$59 \pm 22$	-	$41 \pm 21$	-	-
FR-GE 80/20	-	$18 \pm 7$	-	$46 \pm 7$	$19 \pm 11$	$17 \pm 7$
FR-GE 60/40	-	-	$19 \pm 9$	$54 \pm 6$	$13 \pm 9$	$14 \pm 6$
FR-GE 50/50	-	-	-	$100 \pm 30$	-	-
FR-GE 40/60	-	-	$25 \pm 12$	$60 \pm 8$	$15 \pm 12$	-

Anhang							
	Türkei	Italien	Georgien	Frankreich	Spanien	Deutschland	
FR-GE 20/80	-	-	$42 \pm 19$	$37 \pm 13$	$22 \pm 19$	-	
DE-IT 80/20	-	$51 \pm 30$	-	-	-	$49 \pm 28$	
DE-IT 60/40	-	$44\pm26$	-	$56 \pm 26$	-	-	
DE-IT 50/50	-	$47 \pm 23$	-	$31 \pm 22$	-	$22 \pm 22$	
DE-IT 40/60	-	$47\pm22$	-	$53 \pm 21$	-	-	
DE-IT 20/80	-	$70\pm29$	-	$30\pm28$	-	-	
DE-GE 80/20	-	-	-	-	-	$100 \pm 39$	
DE-GE 60/40	-	-	-	-	-	$100 \pm 47$	
DE-GE 50/50	$24 \pm 16$	-	-	$24 \pm 16$	-	$52 \pm 16$	
DE-GE 40/60	-	-	-	-	-	$100 \pm 48$	
DE-GE 20/80	-	-	$64 \pm 27$	$36 \pm 19$	-	-	
TR-FR-DE 33/33/33	-	-	-	$75 \pm 20$	-	$25 \pm 20$	
TR-FR-IT 33/33/33	-	$45\pm18$	-	$55 \pm 18$	-	-	
TR-FR-GE 33/33/33	-	-	$36 \pm 34$	$27 \pm 24$	$37 \pm 35$	-	
TR-DE-IT 33/33/33	-	$34\pm18$	-	$47 \pm 18$	-	$19 \pm 17$	
TR-DE-GE							
33/33/33	$21 \pm 16$	-	$23 \pm 22$	$28 \pm 16$	-	$28 \pm 15$	
TR-IT-GE 33/33/33	$21 \pm 17$	-	$36 \pm 24$	$43 \pm 17$	-	-	
FR-DE-IT 33/33/33	-	$47\pm20$	-	$31 \pm 19$	-	$22 \pm 19$	
FR-DE-GE 33/33/33	-	-	-	$36 \pm 30$	-	$64 \pm 29$	
FR-IT-GE 33/33/33	-	$32\pm23$	-	$68 \pm 22$	-	-	
DE-IT-GE 33/33/33	-	$33\pm19$	-	$39 \pm 19$	-	$28 \pm 18$	
TR 100	$83 \pm 15$	$17 \pm 16$	-	-	-	-	
GE 100	-	-	$70 \pm 26$	$30 \pm 19$	-	-	
IT 100	-	$58\pm21$	-	$42 \pm 21$	-	-	
DE 100	-	-	-	-	-	$100 \pm 48$	
FR 100	-	$36 \pm 16$	-	$64 \pm 16$	-	-	

	TR	GE	IT	FR	DE	ES
		TG(18:2/18:2/18:3)	PC(18:1/18:2)	PC(18:2/18:2)	PE(18:2/18:2)	TG/14:0/16:0/18:1)
TR	Х	TG(15:0/16:0/18:1)	PC(18:2/18:2)	PE(18:2/18:2)	PC(18:2/18:2)	DG(18:2/18:2)
		TG(14:0/16:0/18:1)	PC(16:0/18:2)	DG(18:2/18:2)	PC(16:0/18:2)	TG(16:0/16:1/18:1)
	TG(18:2/18:2/18:3)		PE(18:2/18:2)	DG(18:2/18:2)	PE(18:2/18:2)	DG(18:2/18:2)
GE	TG(15:0/16:0/18:1)	Х	PC(18:1/18:2)	PE(18:2/18:2)	PC(18:2/18:2)	TG(2:0/18:2/18:2)
	TG(14:0/16:0/18:1)		DG(18:2/18:2)	PC(18:2/18:2)	DG(18:2/18:2)	DG(18:1/18:1)
	PC(18:1/18:2)	PE(18:2/18:2)		TG(2:0/18:2/18:2)	PC(18:2/18:2)	TG(2:0/18:2/18:2)
IT	PC(18:2/18:2)	PC(18:1/18:2)	Х	PC(18:2/18:2)	PE(18:2/18:2)	TG(2:0/18:1/18:2)
	PC(16:0/18:2)	DG(18:2/18:2)		PE(18:2/18:2)	DG(18:1/18:1)	DG(18:1/18:1)
	PC(18:2/18:2)	DG(18:2/18:2)	TG(2:0/18:2/18:2)		PE(18:2/18:2)	PE(18:2/18:2)
FR	PE(18:2/18:2)	PE(18:2/18:2)	PC(18:2/18:2)	Х	DG(18:1/18:1)	PC(18:2/0:0)
	DG(18:2/18:2)	PC(18:2/18:2)	PE(18:2/18:2)		DG(16:0/18:1)	PC(18:2/18:2)
	PE(18:2/18:2)	PE(18:2/18:2)	PC(18:2/18:2)	PE(18:2/18:2)		PE(18:2/18:2)
DE	PC(18:2/18:2)	PC(18:2/18:2)	PE(18:2/18:2)	DG(18:1/18:1)	Х	PC(18:2/18:2)
	PC(16:0/18:2)	DG(18:2/18:2)	DG(18:1/18:1)	DG(16:0/18:1)		PC(16:0/18:3)
	TG/14:0/16:0/18:1)	DG(18:2/18:2)	TG(2:0/18:2/18:2)	PE(18:2/18:2)	PE(18:2/18:2)	
ES	DG(18:2/18:2)	TG(2:0/18:2/18:2)	TG(2:0/18:1/18:2)	PC(18:2/0:0)	PC(18:2/18:2)	Х
	TG(16:0/16:1/18:1)	DG(18:1/18:1)	DG(18:1/18:1)	PC(18:2/18:2)	PC(16:0/18:3)	

Tabelle 8-14: Verwendete Markersubstanzen für das MCIM-Modell der targeted Messungen. Mit dem non-targeted MCIM-Modell übereinstimmende Metabolite sind fett markiert.

Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
			Aserbaidschan		
AZ001	Schlüter & Maack GmbH	2014	unbekannt	unbekannt	unbekannt
AZ002	August Storck KG	2014	unbekannt	unbekannt	Ata Baba
AZ003	Schlüter & Maack GmbH	2015	Zaqatala	unbekannt	Ata Baba
			Deutschland		
DE001	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Baden-Württemberg	Baienfurt	Mix
DE002	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Aiglsbach	Mix
DE003	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Aiglsbach	Mix
DE004	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Bonstetten	Mix
DE005	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Burgheim	Mix
DE006	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Hilgertshausen-Tandern	Mix
DE007	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Oettingen	Mix
DE008	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Pförring	Mix
DE009	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Rennertshofen	Mix
DE010	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Schmiechen	Mix
DE011	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Bayern	Taufkirchen (Vils)	Mix
DE012	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2014	Saxony	Grimma	Mix
DE013	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Baden-Württemberg	Baienfurt	Mix
DE014	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Aiglsbach	Mix
DE015	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Aiglsbach	Mix
DE016	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Aiglsbach	Mix
DE017	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Dorfen	Mix
DE018	EO1 Deutscher Haselnussanbauer UG	2015	Bayern	Nandlstadt	Mix
DE019	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Pförring	Mix
DE020	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Bayern	Taufkirchen (Vils)	Mix
DE021	EO1 Deutscher Haselnussanbauer UG	2015	Bayern	Wolfersdorf	Mix
DE022	EO <sup>1</sup> Deutscher Haselnussanbauer UG	2015	Rheinland-Pfalz	Nieder-Wiesen	Mix
DE023	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Corabel
DE024	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Tonda Romana
DE025	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Tonda di Giffoni

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
DE026	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Istrische Runde
DE027	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Daria
DE028	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Torino
DE029	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Truchsessa
DE030	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Hallesche Riesen
DE031	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Emoa 3
DE032	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Emoa 1
DE033	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Riccia di Talanico
DE034	Sortenversuchsanstalt Gonnersdorf	2014	Bayern	Gonnersdorf	Lange Zellernuss
DE035	Franken Genuss UG & Co.KG	2015	Bayern	Gonnersdorf	Mix
			Frankreich		
FR001	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Bonneville, La Jarthe	Corabel
FR002	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Bonneville	Pauetet
FR003	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Razac D'Eymet	Tonda di Giffoni
FR004	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Grignols	Butler
FR005	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Dieulivol, Grande rivière	Corabel
FR006	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Grignols	Ennis
FR007	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Levignac	Barcelona
FR008	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Grignols, Auzac	Pauetet
FR009	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Dieulivol, Grande rivière	Segorbe
FR010	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Amou, Marlat	Ennis
FR011	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Eyre Moncube	Pauetet
FR012	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Tournon D'Agenaus, Pont de Marty	Butler
FR013	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Haute Vignes, Bourdil	Corabel
FR014	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Beaugas et Moulinet	Ennis
FR015	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Taillecavat, Piquetuile	Barcelona
FR016	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Cancon, Pradou	Lewis
FR017	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Moulinet, Allemans	Pauetet
FR018	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Monflanquin	Segorbe
FR019	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Galapian, Rousselet	Tonda di Giffoni
FR020	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Monein	Butler
FR021	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin	Corabel

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
FR022	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin, Goudandaou	Ennis
FR023	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin, Couren	Pauetet
FR024	SCA Unicoque	2014	Centre-Val de Loire	Maron	Ennis
FR025	SCA Unicoque	2014	Centre-Val de Loire	Maron	Segorbe
FR026	SCA Unicoque	2014	Centre-Val de Loire	Roche, Les Champs Bézard	Corabel
FR027	SCA Unicoque	2014	Centre-Val de Loire	Roche, Les Champs Bézard	Ennis
FR028	SCA Unicoque	2014	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Corabel
FR029	SCA Unicoque	2014	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Ennis
FR030	SCA Unicoque	2014	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Pauetet
FR031	SCA Unicoque	2014	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Segorbe
FR032	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Pessoulens	Corabel
FR033	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Pessoulens, Lagraulas	Ennis
FR034	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne, Ramier de Lasserre	Corabel
FR035	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne, Ramier de Lasserre	Pauetet
FR036	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne	Segorbe
FR037	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Bazordan, Micoulaous	Corabel
FR038	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Bazordan	Ennis
FR039	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Bazordan, Médérac	Pauetet
FR040	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Bazordan, Mourère	Segorbe
FR041	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Butler
FR042	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Corabel
FR043	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Ennis
FR044	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Pauetet
FR045	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Saint Sardos, Cornac	Butler
FR046	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Caumont	Corabel
FR047	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Saint Arroumex	Ennis
FR048	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Caumont, Rende Haut	Feriale
FR049	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Larrazet, Catet	Barcelona
FR050	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Lavit de Lomagne	Pauetet
FR051	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Praderes	Segorbe
FR052	SCA Unicoque	2014	Languedoc-Roussillon-Midi-Pyrénées	Asques, Aubernes	Tonda di Giffoni
FR053	SCA Unicoque	2014	Pays de la Loire	St. Julien, Les Haies	Barcelona

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
FR054	SCA Unicoque	2014	Pays de la Loire	Saint Martin D'Arce	Ennis
FR055	SCA Unicoque	2014	Pays de la Loire	Bazouges/Loir	Ennis
FR056	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Jarnac	Ennis
FR057	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Jarnac, La Chapelle	Pauetet
FR058	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Lathus, La font Villars	Corabel
FR059	SCA Unicoque	2014	Aquitaine-Limousin-Poitou-Charentes	Lathus, La font Villars	Tonda di Giffoni
FR060	SCA Unicoque	2014	Auvergne-Rhône-Alpes	Marsaz	Segorbe
FR061	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Bonneville, La Jarthe	Corabel
FR062	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Bonneville	Pauetet
FR063	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Razac D'Eymet	Tonda di Giffoni
FR064	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Grignols	Butler
FR065	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Dieulivol, Grande rivière	Corabel
FR066	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Grignols	Ennis
FR067	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Levignac	Barcelona
FR068	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Grignols	Pauetet
FR069	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Dieulivol, Grande rivière	Segorbe
FR070	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Amou, Marlat	Ennis
FR071	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Eyre Moncube	Pauetet
FR072	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Tournon D'Agenaus, Pont de Marty	Butler
FR073	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Haute Vignes, Bourdil	Corabel
FR074	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Beaugas et Moulinet	Ennis
FR075	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Taillecavat, Piquetuile	Barcelona
FR076	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Cancon, Pradou	Lewis
FR077	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Moulinet, Allemans	Pauetet
FR078	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Monflanquin	Segorbe
FR079	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Galapian, Rousselet	Tonda di Giffoni
FR080	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Monein	Butler
FR081	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin	Corabel
FR082	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin, Goudandaou	Ennis
FR083	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Hagetaubin, Couren	Pauetet
FR084	SCA Unicoque	2015	Centre-Val de Loire	Maron	Ennis
FR085	SCA Unicoque	2015	Centre-Val de Loire	Roche, Les Champs Bézard	Corabel

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
FR086	SCA Unicoque	2015	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Corabel
FR087	SCA Unicoque	2015	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Ennis
FR088	SCA Unicoque	2015	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Pauetet
FR089	SCA Unicoque	2015	Centre-Val de Loire	Saint-Hilaire-Saint-Mesmin	Segorbe
FR090	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Pessoulens	Corabel
FR091	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Pessoulens, Lagraulas	Pauetet
FR092	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne, Ramier de Lasserre	Corabel
FR093	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne, Ramier de Lasserre	Pauetet
FR094	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Verdun/Garonne	Segorbe
FR095	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Bazordan	Ennis
FR096	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Bazordan, Médérac	Pauetet
FR097	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Butler
FR098	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Corabel
FR099	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Ennis
FR100	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Montlauzun, Couty	Pauetet
FR101	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Saint Sardos, Cornac	Butler
FR102	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Caumont	Corabel
FR103	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Saint Arroumex	Ennis
FR104	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Larrazet, Catet	Barcelona
FR105	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Lavit de Lomagne	Pauetet
FR106	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Praderes	Segorbe
FR107	SCA Unicoque	2015	Languedoc-Roussillon-Midi-Pyrénées	Asques, Aubernes	Tonda di Giffoni
FR108	SCA Unicoque	2015	Pays de la Loire	St. Julien, Les Haies	Barcelona
FR109	SCA Unicoque	2015	Pays de la Loire	Saint Martin D'Arce	Ennis
FR110	SCA Unicoque	2015	Pays de la Loire	Bazouges/Loir	Ennis
FR111	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Jarnac	Ennis
FR112	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Jarnac, La Chapelle	Pauetet
FR113	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Lathus, La font Villars	Corabel
FR114	SCA Unicoque	2015	Aquitaine-Limousin-Poitou-Charentes	Lathus, La font Villars	Tonda di Giffoni
FR115	SCA Unicoque	2015	Auvergne-Rhône-Alpes	Marsaz	Segorbe
			Georgien		
GE001	Schlüter & Maack GmbH	2014	Gurien	unbekannt	Mix

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
GE002	Schlüter & Maack GmbH	2014	Imeretien	Vani	Mix
GE003	Schlüter & Maack GmbH	2014	Imeretien	Vani	Mix
GE004	Schlüter & Maack GmbH	2014	Imeretien	Vani	Mix
GE005	Schlüter & Maack GmbH	2014	Samegrelo	unbekannt	Mix
GE006	August Storck KG	2014	Gurien	unbekannt	Mix
GE007	August Storck KG	2014	Imeretien	Vani	Mix
GE008	August Storck KG	2014	Samegrelo	unbekannt	Mix
GE009	Eganut LLC	2015	Samegrelo	unbekannt	Mix
GE010	Schlüter & Maack GmbH	2015	Samegrelo	Zugdidi	Mix
			Italien		
IT001	Seeberger GmbH	2014	Latium	Viterbo	Tonda Gentile Romana
IT002	Seeberger GmbH	2014	Latium	Viterbo	Tonda Gentile Romana
IT003	Stelma SRL Unipersonale	2014	Kampanien	Avellino	Mortarella
IT004	Stelma SRL Unipersonale	2014	Kampanien	Salerno	Tonda di Giffoni
IT005	Stelma SRL Unipersonale	2014	Latium	Viterbo	Tonda Gentile Romana
IT006	Stelma SRL Unipersonale	2014	Piedmont	Alba	Tonda Gentile Trilobata
IT007	August Storck KG	2014	unbekannt	unbekannt	Mortarella
IT008	August Storck KG	2014	unbekannt	unbekannt	Romane
IT009	Alta Langa Azienda Agricola	2015	Piedmont	Langhe	Tonda Gentile Trilobata
IT010	Azienda Agricola Cascina Valcrosa	2015	Piedmont	Langhe	Tonda Gentile Trilobata
IT011	Azienda Agricola Cascina Valcrosa	2015	Piedmont	Langhe	Tonda Gentile Trilobata
IT012	Corilu Societa Cooperativa Agricola	2015	Piedmont	Lu Monferrato	Tonda Gentile Trilobata
IT013	Seeberger GmbH	2015	Latium	Viterbo	Tonda Gentile Romana
IT014	Stelma SRL Unipersonale	2015	Kampanien	Avellino	Mortarella
IT015	Stelma SRL Unipersonale	2015	Latium	Caprarola	Tonda Gentile Romana
IT016	Stelma SRL Unipersonale	2015	Piedmont	Alba	Tonda Gentile Trilobata
IT017	AgroTeamConsulting	2015	Kampanien	Teano	San Giovanni
IT018	AgroTeamConsulting	2015	Kampanien	Teano	Camponica
IT019	AgroTeamConsulting	2015	Kampanien	Teano	Mortarella
IT020	AgroTeamConsulting	2015	Piedmont	Cortemilia	Tonda Gentile Trilobata
IT021	AgroTeamConsulting	2015	Kampanien	Teano	Tonda di Giffoni
IT022	AgroTeamConsulting	2015	Kampanien	Vairano	Tonda Gentile Romana

			Anhang		
Name	Lieferant	Erntejahr	Region	Stadt/Bezirk	Sorte
			Kanada		
CA001	Grimo Nut Nersery	2014	Ontario	Niagara-On-The-Lake	Mix
			Spanien		
ES001	Coselva SCCL	2014	Katalonien	Tarragona	Negret
ES002	Crisol de Frutos Secos	2014	Katalonien	Tarragona	Mix
ES003	Crisol de Frutos Secos	2014	Katalonien	Tarragona	Mix
ES004	Coselva SCCL	2015	Katalonien	Tarragona	Negret
ES005	Crisol de Frutos Secos	2015	Katalonien	Tarragona	Mix
ES006	Crisol de Frutos Secos	2015	Katalonien	Tarragona	Mix
			Türkei		
TR001	Schlüter & Maack GmbH	2014	Ordu	Ordu	Tombul
TR002	Schlüter & Maack GmbH	2014	Samsun	Carsamba	Tombul
TR003	Schlüter & Maack GmbH	2014	Samsun	Samsun	Tombul
TR004	Schlüter & Maack GmbH	2014	Samsun	Samsun	Tombul
TR005	Basaran Entegre Gıda san. ve Tic. A.Ş	2015	Ordu	unbekannt	Mix
TR006	Basaran Entegre Gıda san. ve Tic. A.Ş	2015	Ordu	unbekannt	Mix
TR007	Universität Hamburg <sup>2</sup>	2015	Düzce	Akçakoca	Çakildak
TR008	Universität Hamburg <sup>2</sup>	2015	Düzce	Akçakoca	Çakildak
TR009	Universität Hamburg <sup>2</sup>	2015	Düzce	Akçakoca	Çakildak
TR010	Universität Hamburg <sup>2</sup>	2015	Düzce	Akçakoca	Çakildak
TR011	Universität Hamburg <sup>2</sup>	2015	Düzce	Akçakoca	Çakildak
TR012	Schlüter & Maack GmbH	2015	Ordu	unbekannt	Palaz
TR013	Schlüter & Maack GmbH	2015	Ordu	unbekannt	Çakildak
TR014	Schlüter & Maack GmbH	2015	Ordu	unbekannt	Yağli
			USA		
<b>US001</b>	Oregon Hazelnuts	2015	Oregon	unbekannt	Barcelona

<sup>1</sup> EO: Erzeugerorganisation
 <sup>2</sup> Institut f
 ür Lebensmittelmikrobiologie und Biotechnologie, Universit
 ät Hamburg

Name	Lieferant	Erntejahr	deklarierte Region	deklarierte Stadt/Bezirk	deklarierte Sorte	Verarbeitung
		Aserbaidscha	an			
AZ004	Universität Hamburg	2014	unbekannt	unbekannt	unbekannt	roh
AZ005	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
		Frankreich	l			
	Koki Haselnusskerne – natural – ganz – 200 g; Kaufland Omni-					
FR116	channel International GmbH & Co. KG, Stresemannstraße 300,	2014	unbekannt	unbekannt	Barcelona	roh
	22761 Hamburg; erworben am 30.06.2014					
		Georgien				
GE011	Carl Wilhelm Clasen GmbH	2014	Samegrelo	unbekannt	unbekannt	roh
GE012	August Töpfer & Co. (GmbH & Co.) KG	2014	unbekannt	unbekannt	unbekannt	roh
GE013	Heinrich Brüning GmbH	2014	unbekannt	unbekannt	unbekannt	roh
GE014	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
GE015	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
GE016	August Töpfer & Co. (GmbH & Co.) KG	2015	unbekannt	unbekannt	unbekannt	roh
		Italien				
IT023	August Töpfer & Co. (GmbH & Co.) KG	2014	Latium	Rom	Tonda Gentile Romana	roh
IT024	Horst Walberg Trockenfrucht Import GmbH	2014	Latium	Viterbo	Tonda Gentile Romana	roh
IT025	August Töpfer & Co. (GmbH & Co.) KG	2015	Kampanien	unbekannt	Tonda di Giffoni	roh
IT026	Heinrich Brüning GmbH	2015	Latium	unbekannt	Tonda Gentile Romana	roh
IT027	Stollwerck GmbH	2015	unbekannt	unbekannt	Tonda Romana	roh
IT028	Ferrero OHG mbH	2015	unbekannt	unbekannt	unbekannt	roh
IT029	Ferrero OHG mbH	2015	unbekannt	unbekannt	unbekannt	roh
IT030	Heinrich Brüning GmbH	2015	unbekannt	unbekannt	Tonda Gentile Romana	roh
IT031	Fratelli Caffa s.a.s.	2015	Piedmont	unbekannt	Tonda Gentile Trilobata	roh
IT032	Stollwerck GmbH	2015	unbekannt	unbekannt	Tonda Gentile Romana	geschält+geröstet
		Spanien				
ES007	August Töpfer & Co. (GmbH & Co.) KG	2014	Katalonien	Tarragona	Mix	roh
		Türkei		-		
TR015	Heinrich Brüning GmbH	2014	Düzce	Akçakoca	unbekannt	roh

#### Tabelle 8-16: Spezifikationen der nicht-authentischen Haselnussproben. (modifiziert nach Klockmann 2017)

		Anhang				
Name	Lieferant	Erntejahr	deklarierte Region	deklarierte Stadt/Bezirk	deklarierte Sorte	Verarbeitung
TR016	Carl Wilhelm Clasen GmbH	2014	unbekannt	unbekannt	unbekannt	roh
TR017	Heinrich Brüning GmbH	2014	Ordu	Ordu	unbekannt	roh
TR018	Heinrich Brüning GmbH	2014	Ordu	Ordu	unbekannt	roh
TR019	Universität Hamburg <sup>2</sup>	2014	Düzce	Akçakoca	unbekannt	roh
TR020	Universität Hamburg <sup>2</sup>	2014	Giresun	unbekannt	unbekannt	roh
TR021	Universität Hamburg <sup>2</sup>	2014	Ordu	unbekannt	unbekannt	roh
TR022	Universität Hamburg <sup>2</sup>	2014	Ordu	unbekannt	unbekannt	roh
TR023	Universität Hamburg <sup>2</sup>	2014	Ordu	unbekannt	unbekannt	roh
TR024	Universität Hamburg <sup>2</sup>	2014	Ordu	unbekannt	unbekannt	roh
TR025	Universität Hamburg <sup>2</sup>	2014	Samsun	unbekannt	unbekannt	roh
TR026	Universität Hamburg <sup>2</sup>	2014	unbekannt	unbekannt	unbekannt	roh
TR027	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR028	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR029	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR030	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR031	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR032	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR033	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR034	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR035	Ferrero OHG mbH	2015	Düzce	Akçakoca)	unbekannt	roh
TR036	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR037	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR038	Ferrero OHG mbH	2015	Düzce	Akçakoca	unbekannt	roh
TR039	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
TR040	August Töpfer & Co. (GmbH & Co.) KG	2015	unbekannt	unbekannt	unbekannt	roh
TR041	Rapunzel Naturkost GmbH	2015	unbekannt	unbekannt	unbekannt	roh
TR042	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
TR043	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	roh
TR044	Heinrich Brüning GmbH	2015	Ordu	unbekannt	unbekannt	roh
TR045	Ferrero OHG mbH	2015	Ordu	unbekannt	unbekannt	roh
TR046	Ferrero OHG mbH	2015	Ordu	unbekannt	unbekannt	roh

		Anhang				
Name	Lieferant	Erntejahr	deklarierte Region	deklarierte Stadt/Bezirk	deklarierte Sorte	Verarbeitung
TR047	Heinrich Brüning GmbH	2015	Trabzon	unbekannt	unbekannt	roh
TR048	Stollwerck GmbH	2015	unbekannt	unbekannt	unbekannt	geschält+geröstet
TR049	Lübecker Marzipan-Fabrik v. Minden & Bruhns GmbH & Co. KG	2015	unbekannt	unbekannt	unbekannt	geschält+geröstet
TR050	Alfred Ritter GmbH & Co. KG	2015	Ordu	unbekannt	unbekannt	geschält+geröstet
TR051	Ludwig Schokolade GmbH & Co. KG	2015	Ordu	unbekannt	unbekannt	geschält+geröstet
		USA				
US002	Koki Haselnusskerne – natural – ganz – 200 g; Kaufland Omni- channel International GmbH & Co. KG, Stresemannstraße 300, 22761 Hamburg; erworben am 30.06.2014	2014	unbekannt	unbekannt	Barcelona	roh
		unbekannt				
NB001	Gut&Günstig Haselnusskerne – natural – ganz – 200 g; Edeka e aktiv Markt Sven Anders, Grindelallee 126, 20146 Hamburg; erworben am 21.04.2014	unbekannt	unbekannt	unbekannt	unbekannt	roh

<sup>2</sup> Institut für Lebensmittelmikrobiologie und Biotechnologie, Universität Hamburg

# 9 Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde. In diese Arbeit eingebrachte Ergebnisse aus Praktika und Projektstudien entstammen meiner Anleitung.

Hamburg, den 24. Dezember 2017

Sven Klockmann