



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

**Mass Spectrometry based
Proteomic Profiling of Truffles**

By

Dennis Krösser, M. Sc.

From Hamburg

Dissertation

For the acquisition of the academic degree

Doctor rerum naturalium

Dr. rer. nat.

At the

University of Hamburg

Faculty of Mathematics, Informatics and Natural Sciences

Department of Chemistry

Hamburg, February 2021

Erstgutachter: Prof. Dr. Hartmut Schlüter

Zweitgutachter: Prof. Dr. Henning Tidow

Tag der Disputation und Druckfreigabe: 07.05.2021

Research for this thesis was carried out from September 2017 until February 2021 in the Mass Spectrometric Proteomics Group of Prof. Hartmut Schlüter, located at the University Medical Center Hamburg-Eppendorf.



With support from



by decision of the
German Bundestag

This thesis was done as part of the Competence Network Food Profiling (CNFP). CNFP was a joint project of universities and commercial enterprises. The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

Table of content

| | |
|---|----|
| List of abbreviations | 4 |
| 1. Zusammenfassung / Abstract | 6 |
| 1.1 Zusammenfassung | 6 |
| 1.2 Abstract | 8 |
| 2. Introduction | 10 |
| 2.1 What are truffles? | 10 |
| 2.1.1 Truffles biology and life cycle | 10 |
| 2.1.2 Truffles in economy | 12 |
| 2.1.3 Food Fraud | 14 |
| 2.2 Proteomics | 15 |
| 2.2.1 Bottom-up proteomics by LC-MS/MS | 16 |
| 2.2.3 Label-free quantification | 19 |
| 2.2.4 Data-independent acquisition (DIA) mass spectrometry | 20 |
| 2.2 Matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF MS) | 22 |
| 2.2.1 MALDI-TOF biotyping | 23 |
| 3. Aim of the thesis | 25 |
| 4. Materials | 26 |
| 5. Methods | 29 |
| 5.1 MALDI TOF analysis of intact proteins from truffles | 29 |
| 5.1.1 MALDI sample preparation | 29 |
| 5.1.2 MALDI-TOF MS acquisition | 29 |
| 5.1.3 MALDI-TOF data processing and clustering analysis | 30 |
| 5.2 Bottom-up LC-MS/MS based proteomics | 30 |
| 5.2.1 Truffle samples | 30 |
| 5.2.2 Protein extraction from truffle powder | 31 |

| | | |
|-------|--|----|
| 5.2.3 | Tryptic digestion of extracted proteins | 31 |
| 5.2.4 | Testing sample preparation by weight-to-volume ratio of extraction buffer..... | 32 |
| 5.2.5 | Testing different incubation times and enzyme-to-protein ratio for tryptic proteolysis | 33 |
| 5.3 | LC-MS/MS by data dependent acquisition (DDA) mass spectrometry | 33 |
| 5.3.1 | LC-MS/MS measurements | 33 |
| 5.3.2 | Data processing | 35 |
| 5.4 | LC-MS/MS by data independent (DIA) mass spectrometry | 35 |
| 5.4.1 | High pH reversed phase fractionation for spectral library generation..... | 35 |
| 5.4.2 | Spectral library generation by LC-MS/MS measurements in DDA mode | 37 |
| 5.4.3 | Data processing for spectral library generation | 38 |
| 5.4.4 | LC-MS/MS measurements in DIA mode | 39 |
| 5.4.5 | Data processing of DIA measurements | 42 |
| 5.4.6 | Functional annotation of differentially regulated proteins | 42 |
| 6. | Results | 43 |
| 6.1 | Mass spectrometry based profiling of truffles..... | 43 |
| 6.2 | Fast profiling by top-down MALDI-TOF analysis of truffles | 43 |
| 6.2.1 | Testing sample preparation for MALDI-TOF analysis | 43 |
| 6.2.2 | MALDI-TOF measurement of truffle samples..... | 48 |
| 6.3 | Deep profiling and characterization by bottom-up proteomics based on LC-MS/MS | 53 |
| 6.3.1 | Testing of sample preparation | 53 |
| 6.3.2 | Simplification of sample preparation | 56 |
| 6.3.3 | Data-dependent acquisition (DDA) approach for profiling of truffles | 61 |
| 6.3.4 | Data-independent acquisition (DIA) approach for profiling of truffles | 64 |
| 6.3.5 | Building the spectral library | 64 |
| 6.3.6 | Profiling truffles regarding the difference between species | 64 |
| 6.3.7 | Introduction of a smaller protein panel for distinction of truffles | 67 |
| 6.3.8 | Gene ontology enrichment analysis..... | 74 |
| 6.3.9 | Profiling truffles regarding their geographical origin..... | 77 |

| | |
|---|-----|
| 6.4 Summary of results..... | 78 |
| 7. Discussion | 79 |
| 9. Outlook..... | 91 |
| 10. References | 93 |
| 11. Appendix | 101 |
| 11.1 Supplemental figures | 101 |
| 11.2 List of figures | 104 |
| 11.3 List of tables | 106 |
| 11.4 List of publications and conference contributions..... | 107 |
| 11.5 Safety and disposal..... | 108 |
| 11.6 CMR list | 110 |
| 12. Danksagung..... | 111 |
| 13. Eidesstattliche Erklärung..... | 112 |

List of abbreviations

| | |
|----------|--|
| % | Percent |
| °C | Degree Celsius |
| µg | Microgram |
| µl | Microliter |
| µm | Micrometer |
| Å | Angstrom |
| ACN | Acetonitrile |
| ANOVA | Analysis of variance |
| BCA | Bicinchoninic acid |
| DDA | Data-dependent acquisition |
| DIA | Data-independent acquisition |
| DTT | Dithiothreitol |
| ESI | Electrospray ionization |
| FA | Formic acid |
| FASP | Filter aided sample preparation |
| FDR | False discovery rate |
| g | Relative centrifugal force |
| HCD | High collision dissociation |
| HPLC | High performance liquid chromatography |
| IAA | Iodoacetamide |
| LC | Liquid chromatography |
| LC-MS/MS | Liquid chromatography tandem mass spectrometry |
| M | Molar |
| m/z | Mass-to-charge ratio |
| MALDI | Matrix-assisted laser desorption-ionization |
| mg | Milligram |
| min | Minutes |
| mm | Millimeter |

| | |
|-------|---|
| mM | Millimolar |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| MS1 | Full scan spectra |
| MS2 | Fragment spectra |
| PCA | Principal Component analysis |
| pH | pH value |
| s | Seconds |
| SDC | Sodium deoxycholate |
| SDS | Sodium dodecyl sulfate |
| TEAB | Triethylammonium bicarbonate |
| TOF | Time of flight |
| UPLC | Ultra performance liquid chromatography |
| VOC | Volatile organic compound |

1. Zusammenfassung / Abstract

1.1 Zusammenfassung

Trüffel als teures und exklusives Lebensmittel sind ein lohnendes Ziel für Lebensmittelbetrug. Trotz ihres hohen Bekanntheitsgrades standen sie bisher wenig im Fokus der Forschung, insbesondere nicht auf Proteomebene. Ziel dieser Arbeit war es, einen schnellen Profiling-Ansatz zur Identifizierung von Trüffelarten zu entwickeln. Darüber hinaus sollte eine tiefere Analyse zur umfassenden Charakterisierung von Trüffelarten durchgeführt werden.

Eine schnelle Identifizierung verschiedener Trüffelarten konnte durch die Matrix-unterstützte Laser-Desorptions-Ionisations-Flugzeitanalyse (MALDI-TOF) intakter Proteine erfolgreich durchgeführt werden. Dafür wurden Probenvorbereitung und Messungen nach etablierten Protokollen in der Mikrobiologie getestet und auf Trüffel übertragen. Aus den aufgenommenen Spektren konnten eindeutige MALDI-Profile für die verschiedenen Arten erstellt werden, die eine Identifizierung und Differenzierung ermöglichen. Mit dem generierten Datensatz wurden bei einer anschließenden Klassifikationsanalyse bis zu 95 % der Instanzen korrekt klassifiziert.

Nach der schnellen Identifizierung durch die Analyse intakter Proteine erfolgte eine tiefere Charakterisierung der Trüffelarten mittels markierungsfreier, differentieller und quantitativer Flüssigkeitschromatographie-Tandem-Massenspektrometrie (LC-MS/MS) basierter Bottom-up-Proteomik. Verschiedene, etablierte Probenvorbereitungsprotokolle wurden auf ihre Anwendbarkeit für Trüffel getestet. Die Extraktion von Proteinen mit dem ionischen Detergens Natriumdeoxycholat gefolgt von einer tryptischen Proteolyse in Lösung, stellte sich dabei als am besten geeignet heraus. Eine weitere Vereinfachung der Probenvorbereitung war durch die Reduktion der enzymatischen Inkubationszeit sowie die Verwendung eines definierten Extraktionsvolumens-zu-Gewichts-Verhältnisses anstatt eines kolorimetrischen Assays zur Schätzung der Proteinkonzentration möglich.

Durch LC-MS/MS-Messungen von 40 Proben im datenabhängigen Aufnahmemodus (DDA) konnten Trüffelarten bereits differenziert werden. Aus den generierten quantitativen Daten wurden Protein-Fingerabdrücke für *T. aestivum*, *T. melanosporum* und *T. magnatum* erstellt.

Obwohl bereits im DDA-Modus eine Differenzierung der Trüffel möglich war, fehlte es an gewünschter Analysetiefe. Daher wurden Messungen im datenunabhängigen Aufnahmemodus (DIA) durchgeführt. Die hierfür erforderliche Spektrenbibliothek wurde mittels zweidimensionaler Chromatographie vor der massenspektrometrischen Analyse generiert. Mit 9,170 Proteinen und 51,628 Peptiden von drei Trüffelarten enthält sie die höchste Anzahl an Trüffelproteinen, die bisher mit Massenspektrometrie identifiziert werden konnte. Durch die Messungen im DIA-Modus wurden 2,715 Proteine mit quantitativen Werten in allen 72 Proben identifiziert. Es wurden wieder Protein-Fingerabdrücke für die verschiedene Trüffelarten erstellt, die nun 2,066 Proteine umfassten. Mit ihnen war eine Differenzierung der Trüffelarten *T. aestivum*, *T. melanosporum*, *T. magnatum*, *T. albidum Pico* und *T. indicum* möglich. Die Protein-Fingerabdrücke konnten anschließend auf ein kleineres Panel mit 15 Protein-Markerkandidaten reduziert werden, um eine einfachere Differenzierung der verschiedenen Trüffelarten zu ermöglichen. Nach der globalen Analyse wurden Trüffel durch Spezies-gegen-Spezies-Vergleiche detaillierter miteinander verglichen. Um einen Einblick in die funktionelle Kategorisierung der Proteine mit unterschiedlicher Abundanz zu erhalten, wurde eine Gen-Ontologie-Anreicherungsanalyse durchgeführt. Sowohl im globalen als auch im Spezies-zu-Spezies-Vergleich wurde hierbei eine Anreicherung von Proteinen für verschiedene Stoffwechsel- und Reduktions-Oxidationsprozesse festgestellt. Weiterhin wurde eine Anreicherung von Stoffwechselprozessen für Schwefelverbindungen in *T. magnatum* und *T. melanosporum* festgestellt, beides Arten die für ihr starkes Aroma bekannt sind.

Diese Arbeit umfasst somit eine schnelle Methode zur Identifizierung von Trüffelarten auf intakter Proteinebene und eine umfassende proteomische Charakterisierung verschiedener Trüffelarten durch Bottom-up-Proteomik durchgeführt, die eine deutliche Differenzierung der Trüffelarten ermöglichte.

1.2 Abstract

Truffles as expensive food are predestined for food fraud. Although widely known, they are scarcely studied, especially at proteome level. Aim of this thesis was to develop a fast profiling approach for identification of truffle species and to perform a deep profiling approach for comprehensive characterization.

For fast identification, matrix assisted laser desorption-ionization time of flight (MALDI-TOF) analysis of intact proteins was successfully performed on truffle samples. Sample preparation and measurements according to established protocols in microbiology were tested and transferred to truffles. Then 72 samples of different truffle species were prepared and measured. Distinct MALDI profiles for different species could be generated from acquired spectra. An identification and differentiation of species was enabled. Using the generated data set for classification analysis, up to 95 % of instances were classified correctly.

After fast profiling with analysis of intact proteins, in-depth characterization was done by label-free differential quantitative liquid chromatography tandem mass spectrometry (LC-MS/MS) based bottom-up proteomics. Different proteomic sample preparation protocols were tested. Protein extraction by the ionic detergent sodium deoxycholate followed by subsequent tryptic proteolysis in solution was most suitable. Further simplification of sample processing by reducing enzymatic incubation time as well as using a defined extraction volume-to-weight ratio instead of performing a colorimetric assay for estimation of protein concentration was shown to be feasible.

By LC-MS/MS measurements of 40 samples in data-dependent acquisition mode (DDA) truffle species could be differentiated. A protein fingerprint for *T. aestivum*, *T. melanosporum* and *T. magnatum* was created from obtained quantitative data. Although analysis in DDA mode already allowed for differentiation of truffles, it lacked the desired depth in analysis. Therefore, measurements in data-independent acquisition mode (DIA) were performed. A required spectral library was generated by two-dimensional chromatography prior to mass spectrometric analysis. Containing 9,170 proteins and 51,628 peptides of three truffle species, the highest reported number of truffle proteins identified by mass spectrometry to date. Within the DIA measurements, 2,715 proteins were identified with quantitative values over all 78 samples. Protein fingerprints for different truffle species were generated and consisted of 2,066 proteins. Differentiation of samples from truffle species *T. aestivum*, *T. melanosporum*, *T. magnatum*, *T. albidum* Pico and *T. indicum* was possible. Obtained

quantitative protein fingerprints could be reduced to a smaller protein panel of 15 marker candidates for a simplified differentiation of truffle species. Then truffles were compared in more detail by species-against-species comparisons. For gaining insight into functional categorization of differentially abundant proteins, gene ontology enrichment analysis was performed. In both, global and species-to-species comparison, enrichment was seen for various metabolic and reduction-oxidation processes. Further, enrichment of sulfur-compound metabolic processes was seen for *T. magnatum* and *T. melanosporum*, both known for their strong aroma.

In summary, within this thesis a fast method for identification of truffle species on intact protein level was developed and a comprehensive proteomic characterization of different species by bottom-up proteomics was performed and allowed for distinct differentiation of truffle species.

2. Introduction

2.1 What are truffles?

2.1.1 Truffles biology and life cycle

Truffles are famously renowned as ingredients in the haute cuisine, the fine dining experience, sometimes even called black diamonds of the kitchen. They are highly sought for their characteristic, hard to describe smell and taste. Some descriptions contain the words musky, garlick-y, sulphurous and funky, with a nutty, earthy taste. One of the compounds making up the truffles smell is adrostenone, a pheromone found in the saliva of male pigs¹, explaining why female pigs make excellent truffle hunters. Truffles should be eaten freshly. Most common practice in the kitchen is to shave them very thinly over pasta, risotto or scrambled eggs. Adding truffle pieces to oils, achieving transfers of aroma, is found in high quality products only. Most commercially available cheap truffle oils are aromatized with synthetic compounds of molecules which were identified to give truffles their distinct aroma.

In scientific perspective, truffles are the edible fruiting bodies of subterranean growing fungi. They belong to the division of ascomycetes, named after their characteristic ascus (from Greek: ἀσκός (*askos*), meaning "sac" or "wineskin"), a microscopic sexual structure in which spores are formed. Truffles mostly originate from the *Tuber* genus (class Pezizomycetes, order Pezizales, family Tuberaceae), which consists of 180 to 220 species by recent estimation². By vegetative, meaning asexual, growth truffles form long, branching and filamentous structures, called hyphae. The network of hyphae is called mycelium. Through the mycelium nutrients and water are absorbed from the environment. Truffles are ectomycorrhizal fungi³. A mycorrhiza (from Greek μύκης *mýkēs*, "fungus", and ῥίζα *rhiza*, "root") is a symbiotic relationship between a fungus and the roots of a plant. In this mycorrhizal association the hyphae of a truffle connects with the plant's roots and grows a dense sheath on the root's surface, called mantle⁴. Then the hyphae only penetrate the plant's roots epidermis (rhizodermis) and forms a network of inwards growing hyphae between the plant root's cells, the so called Hartig net (after Theodor Hartig, a german forestry biologist and botanist). They do not penetrate or damage the cell walls of plant cells. By this symbiotic association an exchange of nutrients between truffle and host plant takes place. Truffles receive carbohydrates from their host plant and provide water in exchange, as well as macronutrients (e.g. potassium, nitrogen) and micronutrients (e.g. zinc, iron)⁵. They are dependent on host plants as a carbon source, because they lack the ability to degrade cell

walls themselves, which limits their capability to decompose plant litter⁶. Same dependency is present for many ectomycorrhizal fungi which have lost enzymes meant for carbon metabolism. A truffle's life cycle consists of vegetative growth of the mycelium, sexual phase for production of spores as well as asexual spore production⁷. Spores in the sexual phase are produced in the asci and are therefore called ascospores. For this sexual production of spores, truffles form a sporocarp, the fruiting body. This fruiting body is a multicellular structure in which the truffles asci are located for sexual spore production. Many truffles need a symbiotic relationship with a host plant to form a fruiting body. Growing beneath the earth, truffle's fruiting bodies cannot simply release spores to be spread by wind or water. They are therefore in need of animal vectors to spread their spores.

To attract fungi eating animals, the truffle will begin to form and release volatile compounds when the ascospores in the fruiting body are fully developed. These volatile compounds are vital for the truffle's aroma, not only necessary to attract animals, but mainly for which it is a highly sought and prized food for us humans. The volatile constituents of truffle aroma can be sulfur volatiles, metabolites of non-sulfur amino acid constituents, fatty acid derived volatiles or others. Some volatiles can be characteristic for a species, in some species they vary highly in quantity⁸, which explains why some truffle sorts are famous for their strong and characteristic aroma and therefore expensive. White truffles of *Tuber borchii* and *Tuber magnatum* are exuding high amounts of sulfuric volatiles and can have quite a pungent aroma. Additionally, the host plant, interacting mammals and insects as well as bacteria inside the fruiting body can have an influence on the smell of a truffle.

After animals like deer, pigs, or rodents are attracted to the truffle's fruiting body by its smell, they dig it up and eat it. While the fruiting body is being eaten, first spores will be set free and spread in the process. Swallowed ascospores can survive the passage through the gastrointestinal tract because of their thick chitin walls. After being excreted, the spores can germinate under appropriate conditions, such as close proximity to a suitable host plant. This dispersal system holds the advantage of spores being concentrated in the feces and that they are more likely to be deposited in a similar area to where the animal roamed for the truffles. This results in a higher chance of availability of suitable host plants. In contrast, the distribution and scattering of airborne spores transported by wind is more random and does not result in a high chance of submitting the spores to an environment suited for germination³. The truffles life cycle is summarized and pictured in Figure 1.

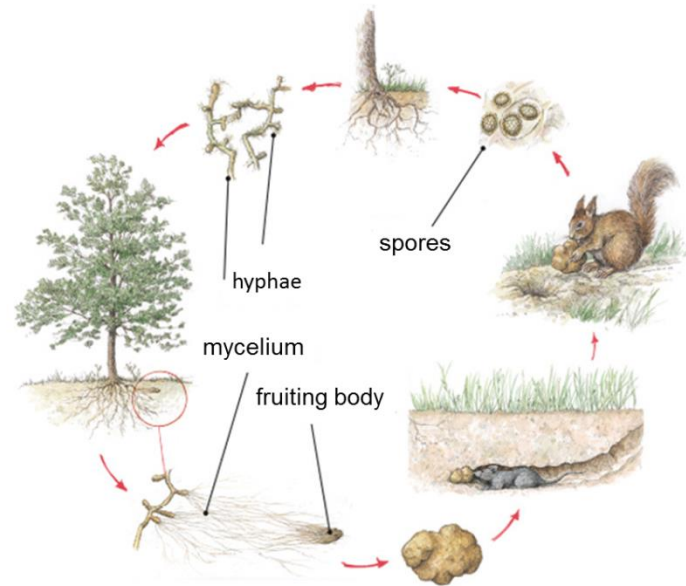


Figure 1: The truffles life cycle (Modified after Centro nazionale studi tartufo⁹).

To summarize, truffles grow hyphae in a vegetative manner. The branched network of hyphae forms the mycelium. In symbiotic relationship with the roots of a host plant the truffle enters the sexual phase and forms a fruiting body. When the fruiting body is mature it releases volatile compounds to attract animal vectors for digging it up and spreading the spores. Spores will start germinating under preferred conditions after being deposited.

2.1.2 Truffles in economy

Only some truffle species of the *Tuber* genus are of high interest within the economy. Most sought after truffles are the Alba white truffle (*Tuber magnatum* Pico) and the Périgord black truffle (*Tuber melanosporum* Vittad.), varying between 3000-5000 € in France in 2018¹⁰⁻¹². The white truffle *Tuber magnatum* or *trifola d'Alba Madonna* ("Truffle of the White Madonna" in Italian) displays to be the most valuable truffle originating from Italy. Harvest is predominantly in October and November. Whitish truffle *Tuber borchii* is also mostly located in Italy, but known to be not as aromatic as *Tuber magnatum*. Black truffles of *Tuber melanosporum* grow in association with oaks and hazelnuts and are harvested in late autumn to winter. The black truffles' genome was sequenced in 2010, representing the first truffle genome¹³. Next are summer truffle *Tuber aestivum* and burgundy truffles, *Tuber uncinatum*. While summer truffles are harvested earlier from May to August, burgundy truffles harvest is taking place from September to December. Both of them are found over most parts of Europe. The burgundy truffle is said to have a stronger aroma than the summer truffle, while both of them are used as substitute for the even more aromatic black truffle.

Alongside with these mostly European truffles, Asian, especially Chinese truffles are recently flooding the markets. These truffles mostly belong to the species *Tuber indicum* and are found in big areas all over China. They are similar in appearance to the black truffle *Tuber melanosporum* and for a layman difficult to distinguish by eye, but aroma-wise the Chinese truffle is far inferior.



Figure 2: Picture of different economical relevant truffle species¹⁴.

Originally truffles were looked for with the help of truffle pigs. But since difficulties arose while trying to keep the pigs from destroying valuable goods, truffle hunters switched to specially trained dogs. By this approach even nowadays big parts of the truffles on the markets are acquired. Another way is to look for the characteristic truffle brûlé caused by some species. Brûlé is French and translates to burnt ground. It describes the area of scant vegetation around host trees associated with some *Tuber* species in their symbiotic phase^{15,16}.



Figure 3: Truffle hunt with the help of a dog on the left¹⁷ and the characteristic brûlé around trees in symbiotic association with truffles¹⁸ on the right.

Besides the traditional hunt for wild growing truffles, cultivation of truffles exhibits an ever-growing economical field. *T. melanosporum* was first cultivated in France in the 19th century, followed by other species like *T. aestivum*, *T. uncinatum*, *T. borchii* and *T. indicum*. Despite research efforts, cultivation is not fully domesticated. Cultivation of the most priced truffle, *T. magnatum*, is not successful to date¹⁹.

Biggest breakthrough in truffle cultivation was the development of methods to inoculate roots of potential host trees under controlled conditions with the fungus. Inoculated trees are then grown and cultivated in orchards, the same as for fruits and nuts. After inoculation first truffles appear several years after planting the host plant seedlings. Onset and duration of truffle growth is depending on the used species and host plant. Yield in Europe is typically between 25 to 35 pounds per acre but can vary dramatically²⁰. By this approach *T. melanosporum* was successfully cultivated in Australia and Northern America. Especially in the US more and more truffle orchards are forming. But cultivation is troublesome. *T. melanosporum*, the most widely cultivated species, needs summers without extreme heat and winters lacking extreme cold. In summer damage may be caused by heated soil, in winter by freezing grounds. They prefer well drained soils²¹. The soils pH should be slightly alkaline, fertilizers shouldn't be overused and trees should be planted relatively isolated²². Also inoculated host plants need to be free of rivaling fungi before plantation.

As truffle hunting employs labor-intensive work and cultivation is difficult and takes a lot of time, this contributes strongly to concluding high prized truffles. For economic gain it is most tempting to cheat using the declaration of truffles, labeling them as expensive species with false origin. This especially applies for *T. indicum*, very similar to *T. melanosporum* in appearance. Economic gain is enormous when only small amounts of *T. indicum* are mixed into a batch of black truffles, because the black truffle is at least four times the impersonator's value. False declaration of food products with the motivation of economic gain is called food fraud.

2.1.3 Food Fraud

Truffles as expensive food are especially prone to be the target of food fraud. Food fraud was defined in 2011 by Spink and Moyer as the “deliberate and intentional substitution, addition, tampering, or misrepresentation of food, food ingredients, or food packing; or false or misleading statements about a product for economic gain”²³. Meaning, the motivation behind food fraud concludes in economic profit. But there can be severe effects on product- and process-quality as well as effects on the consumer of the given food product. Lifestyle, dietary or religiously motivated eating habits can be affected. Monitoring the sugar intake with diabetes is a dietary example, vegetarian and vegan diets are lifestyles, and only eating kosher or halal foods are religious constrictions. Food fraud can even threaten a significant public health risk when products are mislabeled or mixed into the original product. Allergenic

products like nuts can cause severe allergic reactions, worst case scenarios leading to the death of consumers. Supposing hazelnuts would be mixed into peanuts, serious health issues for people with the corresponding allergy could arise.

Food fraud is only one category of the main food risks. Other main types of risks are defined by Spink and Moyer²³ as:

1. Food quality, where an unintentional act leads to a food product that doesn't reach stated or required attributes or standards. An example is fruits or vegetables accidentally being bruised by mishandling.
2. Food safety, where an unintentional act leads to a food product that can pose health concerns if consumed. An example is raw vegetables contaminated with bacteria due to limited protection and control of the food product.
3. Food defense, where the food product is intentionally altered to pose health concerns. An example would be poisoning of food to harm consumers.

For selling truffles, especially their origin and species are important because of the widely assumed linkage to higher quality and taste of in particular European truffles. Omics approaches have gained strong attention for authentication regarding geographical origin, biological identity and production process of food²⁴. Proteomics for food authentication, especially quantitative proteomics, can be utilized because protein abundance is not only linked to the endogenously derived genotype but also to exogenous influences of either natural or man-made origin. Examples for natural influences are climate, soil composition and stress conditions, man-made influences can be storage conditions, processing and harvest of food. Proteomics approaches in the field of food authentication were successfully used for determining the biological identity (genotyping)²⁵⁻²⁹ or geographical origin³⁰, type of cultivation³¹ and storage processes³²⁻³⁴ for different classes of food.

2.2 Proteomics

Cells contain thousands of proteins, which inherit numerous functions. Proteins give cells their structure, allow biochemical reactions to take place and have part in many other processes in a cell and organism. The term “proteome” was introduced by Marc R Wilkins³⁵ in analogy to the genome in 1994. A proteome is defined as the entirety of proteins in a cell, tissue or organism over its lifetime. Analysis of proteomes under defined conditions at a given time point is referred to as proteomics^{35,36}. Smallest units of a proteome are the protein

species. Protein species are different functional proteins deriving from one gene product by post-translational modifications, alternative splicing, post-translational processing and other processes. The number of human protein species for example is estimated with one billion different species³⁷. Furthermore, the abundance of different proteins in one cell or tissue can span over twelve orders of magnitude³⁸. Mass spectrometry (MS) has proven to be an excellent technique to expand research of the proteome. Mainly, two different approaches are applied. Top-Down approaches focus on the separation and analysis of intact proteins³⁹. Bottom-Up proteomics is the most common approach for proteome analysis. Here, liquid chromatography tandem mass spectrometry (LC-MS/MS)^{40,41} is used to analyze small peptides derived from the most often tryptic proteolysis of intact proteins.

2.2.1 Bottom-up proteomics by LC-MS/MS

In bottom-up proteomic workflows, utilizing LC-MS/MS, a peptide mixture generated by enzymatic proteolysis with trypsin is separated on a chromatographic column. Eluting peptides are ionized in an electrospray source by an electrospray ionization process (ESI) and subjected into the mass spectrometer. The mass spectrometer functions as detector for the chromatography and can further fragment peptides for identification.

The protease of choice for generating peptides is trypsin. Tryptic peptides have in average a size of 6-25 amino acids, which is very suitable for separation by liquid chromatography and the mass spectrometric analysis. Furthermore trypsin has the advantage to specifically cleave peptide bonds C-terminal after the basic amino acids arginine and lysine. When the generated peptides are separated and ionized under acidic conditions, this will lead to the addition of at least two charges by protonation for the majority of tryptic peptides. One charge is added at the C-terminal basic residue; the other at the peptides N-terminal amino group. If a histidine is present or cleavage sites are missed by trypsin, peptide charges of +3 and more are possible.

Separation of peptides is achieved using high- or ultra-performance liquid chromatography (HPLC/UPLC) instruments. This separation step of peptides prior to mass spectrometric analysis is required because of the proteomes complexity and limited instrumental capability of the mass spectrometer. Reversed phase chromatography with C18 based stationary phases are commonly used to separate peptides according to their hydrophobicity. A small portion of peptides is strongly hydrophilic, a small portion strongly hydrophobic and most of them somewhere in between. Regarding their typical size, C18 material works best for peptides, while bigger molecules, like proteins, would require material less hydrophobic (e.g. C4

material) to allow separation and prevent irreversible binding to the stationary phase. The mobile phase consists of aqueous solvent to bind peptides on the column and an increasing gradient of organic solvent to elute them. Subsequent to HPLC application, eluting peptides are ionized by electrospray ionization (ESI). ESI is an ionization technique used in mass spectrometry to produce ions in the gas phase by applying a high voltage to a liquid, creating an aerosol. ESI is a soft ionization technique causing only very little fragmentation and therefore is especially suited for the analysis of biomolecules. Electrospray ionization consists of two processes, both desolvation and ionization of molecules. The process is depicted in Figure 4.

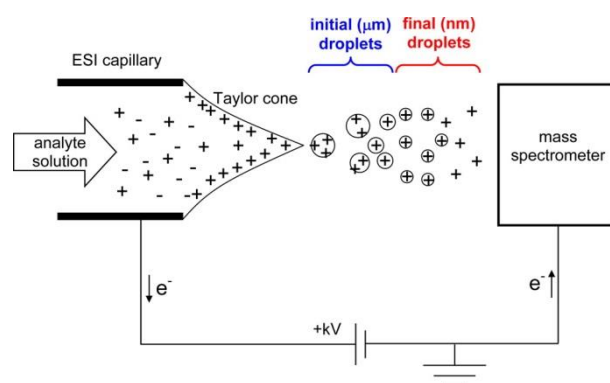


Figure 4: Schematic representation of the electrospray ionization (ESI) process⁴². At the tip of the capillary a Taylor cone forms and an aerosol is generated. Liquid from the aerosol's droplets is evaporating. When the Rayleigh limit of charges for droplets is reached, the droplet will burst through the Coulomb explosion in smaller droplets until ionized molecules are transferred into the gas phase.

After ESI, ionized peptides are infused in a mass spectrometer. A mass spectrometer measures the mass-to-charge ratio (m/z) of ionized molecules in the gas phase. It generally consists of three parts, an ion source (e.g. beforehand mentioned ESI source), a mass analyzer and a detector. Once the ion source has generated charged molecules in the gas phase, the mass analyzer separates the ions according to their m/z ratio. Subsequently, the detector monitors the income of ionized molecules. Modern mass spectrometers often consist of two or more mass analyzers. An example is the Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer shown in Figure 5.

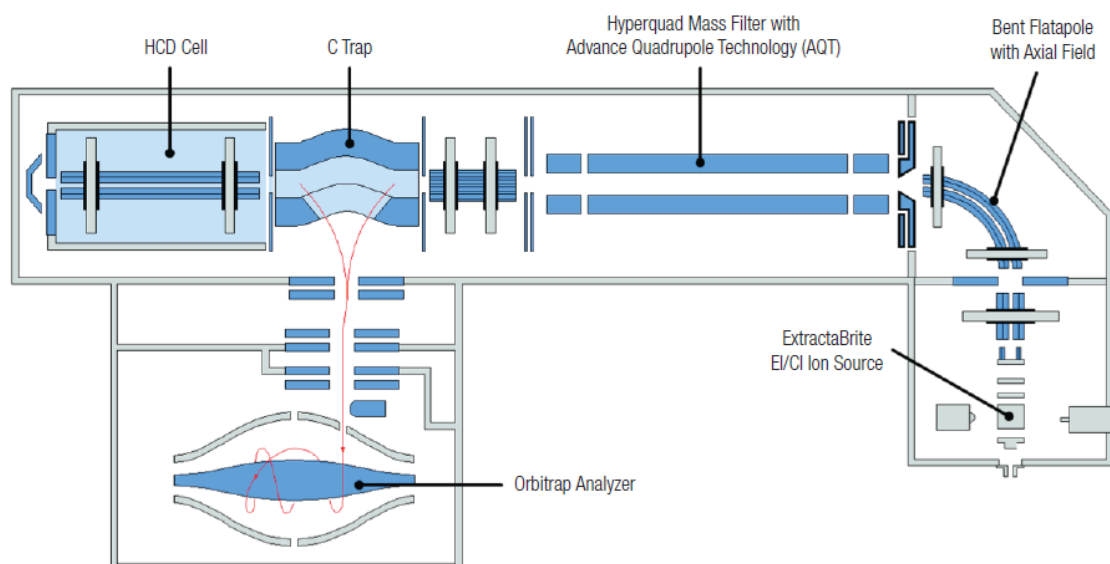


Figure 5: Schematic construction of the Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific).

In this hybrid instrument a quadrupole and orbitrap mass analyzer are utilized. The quadrupole is used for filtering molecules over a broad m/z range or selecting only molecules in a small m/z window. A typical broad scan range for peptides would be 400-1200 m/z . This will cover most of the generated tryptic peptides in the expected charge states. The Higher Collision Dissociation cell (HCD cell) can be used to fragment molecules and the orbitrap analyzer measures the m/z values of either intact or fragmented molecules. Orbitraps are high resolution mass spectrometers. Ions are trapped, simultaneously injected into the orbitrap and oscillating along a central electrode. This oscillation is only dependent on the ions m/z values. The acquired frequency of oscillating ions can be transformed by Fourier transformation into m/z values and then represented as mass spectra. The longer this oscillation of ions is measured, the higher the achieved resolution. Therefore the orbitrap needs a defined acquisition time to reach a certain resolution for a given m/z value.

2.2.2 Data-dependent acquisition (DDA) mass spectrometry

LC-MS/MS based bottom-up proteomics experiments are mostly done in data dependent acquisition mode (DDA). Using DDA mode, the peptide mixture, separated on the preceding LC system, is directed into the mass spectrometer via an ESI source. Peptide precursor ions are detected by scanning them while still intact in full scan mode of the MS instrument, referred to as MS1 scan. Then, precursor ions are isolated with a narrow mass-to-charge ratio (m/z) window. This selection of precursor ions for fragmentation is done according to their signal intensity. Only most intense signals are chosen for fragmentation, often incorporating

an intensity threshold value. In special analytical cases this precursor selection can be based on other parameters like charge state. Then these selected and isolated precursor ions of single peptides are fragmented, i.e. with Higher-Collision Dissociation (HCD). Resulting fragment spectra are named MS2 spectra in regard to the MS1 spectrum as intact spectrum. From one MS1 scan several precursors are selected for fragmentation. Instruments can either fragment the ten or more most intense precursors or can be told to select and fragment precursors for a specified duration, e.g. 2-3 seconds. This process of MS1 scan, selection of precursors and their fragmentation is called duty cycle and repeated over the whole duration of a LC-MS/MS run. Peptide identification is done by comparison of the experimentally derived MS2 spectra with theoretical tandem mass spectra⁴³. Theoretical spectra are generated from the *in-silico* proteolysis of a corresponding protein database^{44,45}. This procedure is possible because of peptides predictable fragmentation behavior depending on the type of activation. With HCD, peptides are fragmented at the weakest bond, the peptide bond. This will be taken into account for peptide *in-silico* fragmentation with database-driven identification. For tryptic peptides with charges added through protonation at C- and N-termini, distinct fragment ion series can be observed by HCD. Fragment ion series where the charge is retained on the N-terminus are named b ions, fragment ion series with the charge retained on the N-terminus y ions⁴³. Peptides are scored by the search engine according to the number of matches for b- and y-ions. One downside of bottom-up proteomics is the loss of information about different protein species. Peptides can only be matched to their corresponding gene products, not individual protein species. Therefore, bottom-up is an indirect measurement of gene products by their corresponding peptides⁴¹.

2.2.3 Label-free quantification

Often the aim of a proteomic study or experiment is not only to identify proteins, but additionally quantify them. Bottom-up proteomics can be used for label-free quantification. Here, peptide and corresponding protein abundance can be compared. This comparison is relative between samples. Results will be a fold change in peptide or protein abundance. For an absolute quantification with a concentration or amount as result, an isotope labelled internal standard of known concentration has to be applied. Label-free quantification is mostly based on the area under the chromatographic peak. In LC-MS/MS based bottom-up proteomics, eluting peptides are selected as precursor ions in the MS1 scan and fragmented. If a peptide is identified, the chromatographic MS1 trace of the peptide precursor ion can be plotted as an extracted ion chromatogram. Quantification is done by integration of the

resulting chromatographic area under the curve for this precursor ion. Relative comparison now is enabled by comparing the peak areas of a given peptide in two or more samples. The quantified peptides of a protein can be summed up to the corresponding protein abundance⁴⁶. Because the quantification relies on the chromatographic area under the curve, signal traces need to be monitored well enough to represent the according analyte. For a reliable quantification 8-12 points across a chromatographic peak are considered sufficient. For this, duty cycle duration of the MS instrument and peak width from chromatographic separation has to be matched.

Relative quantification of proteins in different samples can also be achieved by labeling approaches. These heavy isotopes containing labels are either attached to (e.g. tandem mass tags, TMT) or metabolically incorporated into the samples (stable isotopic labeling of amino acids in cell culture, SILAC). While offering the advantage of multiplexing at least two samples by usage of labels, labeling approaches with heavy isotopes are expensive. This is especially the case for incorporation of labels by SILAC, e.g. the SILAC mouse⁴⁷. A further problem of metabolic labeling is that it is not possible in case of human samples. While still not as precise as quantification based on reporter ions, label-free quantification has the advantage of incorporating as many samples as wanted in the experiment. Most important, label-free quantification can be done for any type of sample at hands.

2.2.4 Data-independent acquisition (DIA) mass spectrometry

With data dependent acquisition problems arise regarding reproducibility in peptide and protein identification and therefore quantification. First, the number of eluting peptides can exceed the analytical capacity of the mass spectrometer at a given time point. In this case only most intense precursor ions are chosen for fragmentation. Second, eluting peptides can be under the instruments detection limit. Therefore only high abundant proteins are reproducibly identified. This results in missing values for lower abundant proteins through a set of samples and carries over to reliable quantification as well. To overcome the randomness of data dependent acquisition, the more elaborate technique of data independent acquisition mode (DIA) was developed⁴⁸. In DIA mode the instrument is repeatedly cycling through consecutive precursor isolation windows over the chromatographic time range. All precursor ions within the small predefined isolation windows are fragmented. Fragment spectra generated with this approach are therefore systematic and unbiased. They are independent from the content of the optionally done MS1 scan. No longer are single precursor ions chosen

according to their intensity. To optimize the acquisition scheme in DIA mode, different parameters have to be taken into account. This is further depending on the instrumentation used. Cycle time and chromatographic peak width have to be matched appropriately for enough points across the chromatographic peak to perform a reliable quantification. In case of data acquisition with the orbitrap analyzer and its need for acquisition time to reach a certain resolution, the resolution setting, DIA isolation window size and number of isolation windows have to be carefully chosen and coordinated. Because not only single precursor ions but small m/z ranges containing many precursor ions are fragmented, a database-driven identification of peptides and proteins is not possible. Generated MS2 spectra are highly complex and not suited for such an approach. To match fragment ions to the corresponding precursors a spectral library needs to be generated. In this spectral library fragment ions and their assigned peptides are deposited. Additional information including peptide retention time and isotopic patterns may additionally enhance the quality of identification through matching with the library. Spectral libraries are usually generated by measuring a subset of samples or pooled samples in DDA mode. To increase number of identifications and therefore depth of the library often a more-dimensional separation and fractionation is performed for these samples. This approach strongly reduces the sample complexity and therefore increases the number of identified proteins, which enhances depth in analysis. This is crucial for an adequate quality spectral library, because only deposited peptides and their assigned fragments can be matched in later DIA experiments. An example for two-dimensional chromatographic separation is the often used high pH reversed phase separation with concatenated fraction pooling prior to low pH reversed phase chromatography of the generated fractions^{49,50}.

For successful and reliable fragment matching with subsequent quantification, at least 4 or more y -ions of one peptide are recommended. The y_1 ion, the c -terminal arginine or lysine of a tryptic peptide, should be excluded, as it is not specific enough for reliable matching.

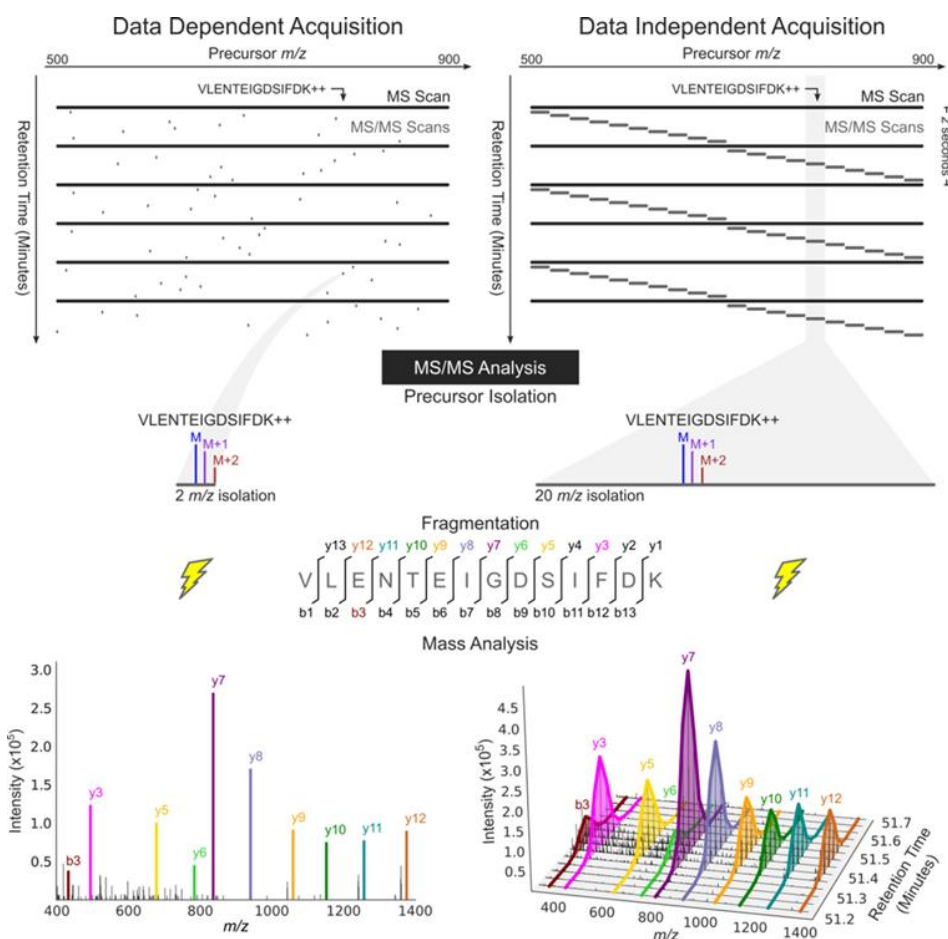


Figure 6: Graphical comparison of data dependent acquisition and data independent acquisition.⁵¹

When matching the spectra generated by DIA measurements against the spectral library, the assigned fragment ions of peptides are monitored. The area under the curve of extracted ion chromatograms for the monitored fragment ions can then be used to calculate peptide intensity. By comparison of peptide identity relative quantification is enabled.

2.2 Matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF MS)

Matrix-assisted laser ionization-desorption (MALDI) is an ionization technique used for mass spectrometry. For MALDI a matrix is used that absorbs the lasers energy to generate ions with minimal fragmentation. Like ESI, MALDI is a soft ionization technique and therefore well suited for analysis of biomolecules^{52–54}. In contrast to ESI, which mostly generates multiple charged ions, MALDI generally leads to singly charged molecules.

MALDI is a three step process. The sample is mixed with a suitable matrix and then irradiated by a pulsed laser with a suitable wavelength for the used matrix. Matrix and sample are

ablated and desorbed, ionized by protonation or deprotonation and then, from the ablation plume, led into the mass spectrometer.

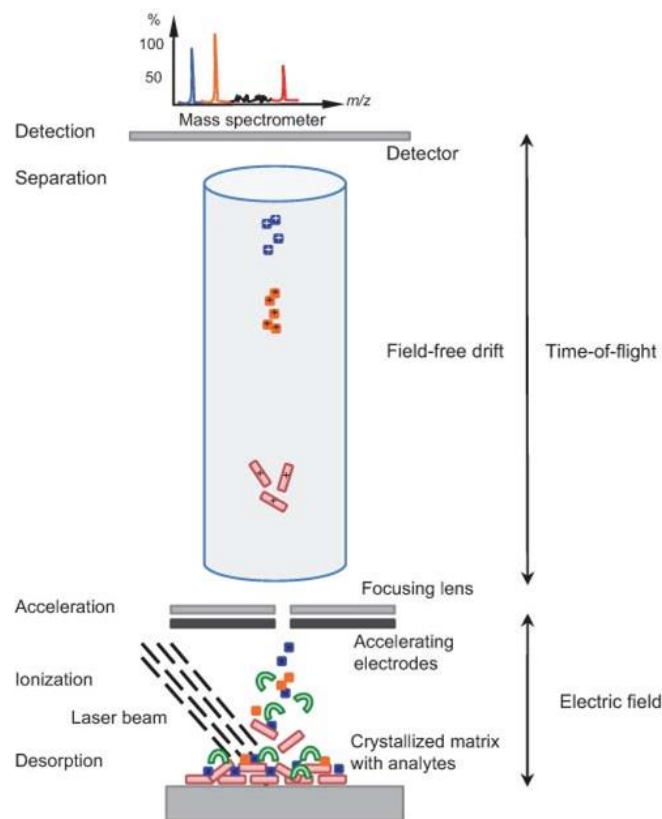


Figure 7: Principle of Matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS)⁵⁵.

Once inside the mass spectrometer, ions are accelerated by an electrical field and led in a field-free drift tube before they hit the detector. The time after the ions enter the drift tube until they hit the detector is measured. This is name-giving for MALDI time of flight analysis (MALDI-TOF). Because all ions are accelerated with the same electrical field strength, the velocity and time of flight of them is only dependent on their m/z ratio. Here, ions with small m/z ratio are accelerated stronger and reach the detector faster as ions with a high m/z ratio.

2.2.1 MALDI-TOF biotyping

MALDI-TOF MS is routinely used in clinical diagnostics to identify pathogenic microorganisms and bacteria⁵⁶. The food industry uses MALDI-TOF biotyping to identify food-associated bacteria for accessing food processing and product quality⁵⁷. Both fields highly profit from the quick and easy sample preparation as well as the rapid measurement of hundreds of samples in an automated manner. In MALDI-TOF biotyping reference spectra for

bacterial species are generated. These spectra are highly reproducible and mostly independent from culture medium^{58–60}. After measurement they can be deposited in a database. This is only possible if there are characteristic profiles by which species can be distinguished, as seen in Figure 8.

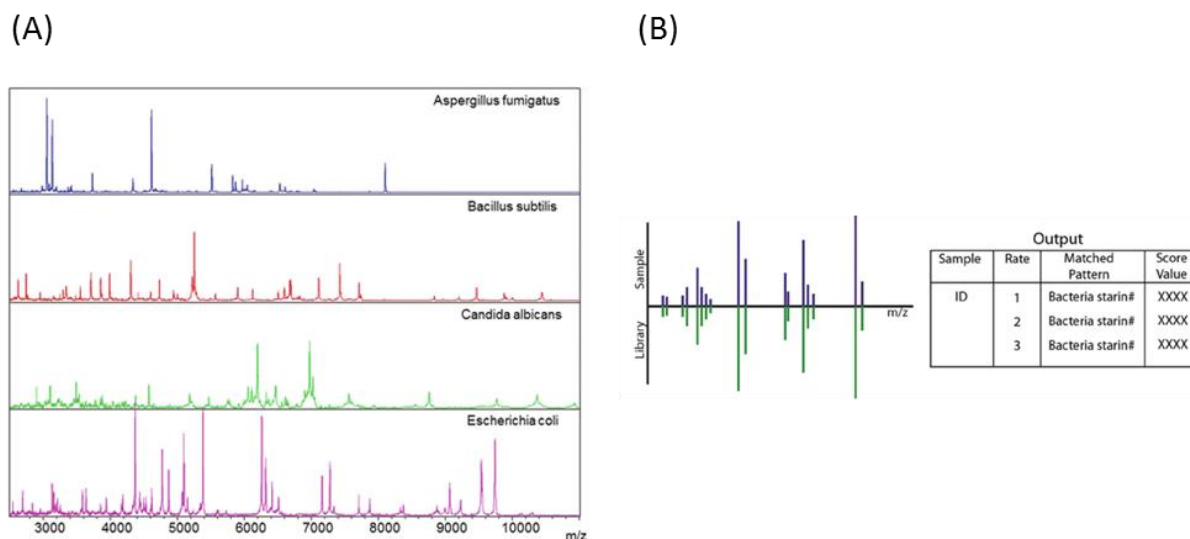


Figure 8: Principle of MALDI-TOF biotyping for bacteria. (A) MALDI-TOF MS profiles acquired for four microorganisms. The microorganisms are belonging to different phylogenetic branches: the mold *Aspergillus fumigatus*, the Gram-positive bacterium *Bacillus subtilis*, the yeast *Candida albicans*, and the Gram-negative bacterium *Escherichia coli* by Pranada et al.⁵⁶ (B) Experimental spectrum of unknown sample is searched against reference spectrum in database. The output shows the best hits with reference spectra from the database. Modified after Luzzatto-Knaan et al.⁶¹

For unknown samples experimental spectra are generated and matched against reference spectra for each species deposited in the database. Identification is done by pattern matching of unknown and reference spectra. A score value indicates how good the sample spectrum matches against deposited spectra. A match with the highest score value shows the best hit and the most likely identification of an unknown sample.

3. Aim of the thesis

Aim of this thesis was to develop a fast method for identification of truffle species and to perform a comprehensive proteomic characterization of them. Therefore, two different mass spectrometry-based profiling approaches on proteome level will be done.

Fast profiling should enable identification of truffle samples. This can be of great use as a first step in the assessment of possible food fraud. If a truffle samples seems suspicious and likely to be falsely declared by a fast approach, it can further be subjected to an in-depth analysis for verification or rejection of this suspicion. For fast profiling, an analysis of intact proteins will be performed. MALDI-TOF biotyping analysis for measuring intact proteins is routinely used in microbiology to identify bacteria. Established protocols will be tested and the approach transferred for the analysis of truffle samples.

In-depth profiling of truffles will be performed by a bottom-up proteomic approach. Here an untargeted, differential and label-free quantitative LC-MS/MS based analysis of tryptic peptides will be done. Generated peptides will be measured in data-dependent acquisition (DDA) mode and later in a more elaborate mass spectrometric approach, the data-independent acquisition (DIA). Further two-dimensional chromatographic fractionation prior to mass spectrometric analysis of peptides will be performed to build a reference spectral library. This library is needed for comprehensive characterization with depth in analysis by DIA. By this in-depth differential proteomics approach a differentiation of truffle species based on protein abundances should be enabled. Main goal will be the generation of a protein fingerprint for different truffle species. Further species-against-species comparisons will be done to highlight proteomic differences more in detail. Obtained results should then be interpreted biologically and linked to underlying biological processes responsible for differences in truffles.

4. Materials

Table 1: List of chemicals and enzymes used.

| Chemicals and enzymes | Manufacturer |
|------------------------------------|---|
| Acetonitrile, LC-MS grade | Merck (Darmstadt, Germany) |
| alpha-Cyano-4-hydroxycinnamic acid | Bruker (Billerica, Massachusetts, USA) |
| Ammonium hydrogen carbonate | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Dithiothreitol | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Formic acid | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Iodoacetamide | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Methanol | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Sequencing grade modified Trypsin | Promega (Madison, Wisconsin, USA) |
| Sodium deoxycholate | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Sodium dodecyl sulfate | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Triethylammonium bicarbonate | Thermo Fisher (Waltham, Massachusetts, USA) |
| Trifluoroacetic acid. | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Urea | Sigma-Aldrich (St. Louis, Missouri, USA) |
| Water, LC-MS grade | Merck (Darmstadt, Germany) |

Table 2: List of disposables used.

| Disposables | Manufacturer |
|--|---|
| Amicon Ultra 0.5 ml 10K centrifugal filters | Merck Millipore (Billerica, Massachusetts, USA) |
| Amicon Ultra 0.5 ml 30K centrifugal filters | Merck Millipore (Billerica, Massachusetts, USA) |
| Capturem™ Trypsin Miniprep Kit (Mass Spectrometry Grade) | TaKaRa (Kusatsu, Shiga, Japan) |
| Pipette Tips | Sarstedt (Nümbrecht, Germany) |
| Reaction Tubes | Sarstedt (Nümbrecht, Germany) |
| SepPak C18 1cc 100 mg | Waters (Milford, Massachusetts, USA) |

Table 3: List of devices used.

| Devices | Manufacturer |
|----------------------------|---|
| Analytical scale ALS 120-4 | Kern & Sohn GmbH (Balingen-Frommern, Germany) |
| Centrifuge 5424 | Eppendorf (Hamburg, Germany) |
| Microplate Reader | Tecan Life Sciences (Männedorf, Switzerland) |
| Probe sonicator Sonoplus | Bandelin electronic GmbH & Co. KG (Berlin, Germany) |
| Speedvac | Thermo Fisher (Waltham, Massachusetts, USA) |
| Thermomixer compact | Eppendorf (Hamburg, Germany) |

Table 4: List of chromatography instruments used.

| Chromatography Instruments | Manufacturer |
|-----------------------------------|---|
| Agilent 1200er series | Agilent (Santa Clara, California, USA) |
| Dionex Ultimate 3000 | Thermo Fisher (Waltham, Massachusetts, USA) |
| Äkta prime plus fractionator | GE Healthcare (Chicago, Illinois, USA) |
| nanoAcquity | Waters (Milford, Massachusetts, USA) |

Table 5: List of chromatography columns used.

| Chromatography columns | Manufacturer |
|--|---|
| Acquity UPLC® Peptide BEH C18 Column, 130 Å pore size, 1.7 µm particle diameters, 75 µm x 200 mm | Waters (Milford, Massachusetts, USA) |
| Acquity UPLC® Symmetry C18 Trap Column, 100 Å pore size, 5 µm particle diameters, 180 µm x 20 mm | Waters (Milford, Massachusetts, USA) |
| ProSwift™ RP-4H analytical, 1 x 250 mm | Thermo Fisher (Waltham, Massachusetts, USA) |

Table 6: List of mass spectrometry instruments used.

| Mass spectrometry instruments | Manufacturer |
|--|---|
| Orbitrap Fusion™ Tribrid™ Mass Spectrometer | Thermo Fisher (Waltham, Massachusetts, USA) |
| Q Exactive™ UHMR Hybrid Quadrupole-Orbitrap™ Mass Spectrometer | Thermo Fisher (Waltham, Massachusetts, USA) |
| rapifleX MALDI Tissue typer | Bruker (Billerica, Massachusetts, USA) |

Table 7: List of software used.

| Software | Manufacturer/Developer |
|--------------------------------------|--|
| FlexAnalysis 4.0 and FlexControl 4.0 | Bruker (Billerica, Massachusetts, USA) |
| Magellan | Tecan Life Sciences (Männedorf, Switzerland) |
| MassUp | López-Fernández et al. |
| MaxQuant 1.6.2.10 | Mann and Cox Lab Group |
| mMass | Martin Strohm |
| Perseus 1.5.8.5 and 1.6.2..3 | Mann and Cox Lab Group |
| ProteomeDiscoverer 2.0 | Thermo Fisher (Waltham, Massachusetts, USA) |
| Skyline 20.1 | MacCoss Lab Group |

5. Methods

5.1 MALDI TOF analysis of intact proteins from truffles

5.1.1 MALDI sample preparation

Formic acid-acetonitrile sample extraction method: Formic acid-acetonitrile (FA/ACN) sample extraction was performed based on the protocol from Sauer and Freiwild^{62,63}. 8-10 mg of lyophilized truffle powder was weighed in. Each sample was mixed with 200 μ l of 70 % formic acid and 200 μ l acetonitrile. Samples were vortexed for 30 s and shaken for 5 min. Then samples were centrifuged for 2 min at 13,000 G and 1 μ l of the supernatant was spotted on a 384 spot ground steel target. Spots were dried for 15 min at room temperature before 1 μ l of MALDI matrix solution (saturated α -Cyano-4-hydroxycinnamic acid, 50 % acetonitrile, 2.5 % trifluoroacetic acid and HPLC grade water) was overlaid onto each spot for crystallization. After layering matrix onto spots the target was again dried for 15 min at room temperature.

Combined solution preparation method: Combined solution extraction method was based on the work of Reeve et al⁶⁴. 8-10 mg lyophilized truffle powder was weighed in. Samples were mixed with 100 μ l of a combined extraction and matrix solution. This solution contained 12 mg/ml α -Cyano-4-hydroxycinnamic acid in 60 % acetonitrile and 2.5 % trifluoroacetic acid. Samples were vortexed for 30 s and shaken for 5 min, then centrifuged for 2 min at 13,000 G and 1 μ l of the supernatant was spotted on a 384 spot ground steel target. The spots were dried for 15 min at room temperature.

As positive control 1 μ l of the Bruker Protein Calibration Standard 1 was spotted and overlaid with 1 μ l of the matrix solution. Negative control was spotted matrix solution. Directly after drying the target was introduced into the MALDI-TOF system (rapifleXTM MALDI TissueTyperTM).

5.1.2 MALDI-TOF MS acquisition

For data acquisition, the rapifleXTM MALDI TissueTyperTM was controlled with Flex Control Software 4.0. Truffle profiles were acquired in positive mode with a method using linear mode detection and a scan range of 2,000-20,000 m/z. Each spectrum was the sum of 20,000 laser shots. Each spot was scanned by the “smart complete sample” function of the software with the diameter limited to 2,000 μ m, 50 shots a raster and a frequency of 2,000.

Detector voltage was set according to the proposed value after detector check in linear mode, laser power was set to 45 %.

5.1.3 MALDI-TOF data processing and clustering analysis

MALDI-TOF spectra were opened with FlexAnalysis software 4.0 and processed by integrated smoothing and baseline-subtraction functions before export as graphics. Unprocessed MALDI-TOF spectra were further exported as mzXML files with FlexAnalysis software 4.0. These mzXML files were processed with the MassUp software⁶⁵. Smoothing by Savitzky Golay algorithm and baseline correction by TopHat algorithm was done by implemented operators. Peak detection was done by the MALDIQuant feature with a signal-to-noise ratio of 3 and half window size of 60. Peaks had to be in at least 66 % of replicate samples. Minimum peak intensity was set to values of 500 (comparing all biological samples), 1,000 (measurement reproducibility test and method comparing tests) and 5,000 (applicability test for biological samples). Intra- and inter-sample peak matching was done by the MALDIQuant feature with a tolerance of 0.002. Settings for clustering analysis were a minimum variance of 0. Cluster reference value was set to average, the distance function set to Euclidean and the conversion values to presence. Principal component analysis was done with default settings and the resulting sample projection values were exported. For classification analysis naiveBayes and random forest were tested as classifier with a 66 % split of data in training and test set.

Un-processed MALDI-TOF spectra were exported as mzXML files and further imported in the mMass software⁶⁶. Here, spectra were baseline-corrected and smoothed by the batch processing operator. Then processed spectra were averaged into a final average spectrum including a barcode depiction of the obtained signal intensities in the average spectrum.

5.2 Bottom-up LC-MS/MS based proteomics

5.2.1 Truffle samples

All truffle samples were obtained as homogenized truffle powder from the group of Prof. Markus Fischer from the University of Hamburg in cooperation with the Trüffelkontor GmbH (Waldmünchen, Germany) and the La Bilancia Trüffelhandels GmbH (Munich, Germany). Truffles were washed with ultrapure water and stored at -80 °C. Homogenization was done with a GM 300 knife mill (Retsch, Haan, Germany), using one part truffle and one part dry ice. The obtained powder was lyophilized for 48 h and further stored at -80 °C.

5.2.2 Protein extraction from truffle powder

Samples containing powder of sample were mixed with one of the following extraction buffers in technical triplicates: 1% sodium deoxycholate (SDC)/ 100 mM triethylammonium bicarbonate (TEAB), in short SDC buffer, 8 M urea/50 mM ammonium hydrogen carbonate, in short urea buffer, or 5 % sodium dodecyl sulfate (SDS)/ 50 mM ammonium hydrogen carbonate, in short SDS buffer, in a ratio of 14 μ l buffer to 1 mg of truffle powder. Samples mixed with SDC and SDS buffers were boiled for 10 min at 99 °C before sonication with a sonication probe (Bandelin Sonoplus, 30 % energy for 30 s). Samples mixed with urea buffer were incubated on ice for 30 min before sonication on ice with the sonication probe. All samples were centrifuged for 10 min at 10,000 g and the supernatant transferred to a new reaction tube. Obtained supernatant was used for protein concentration estimation by Pierce™ Protein Assay Kit according to manufacturer's instructions, hereinafter referred to as BCA assay (bicinchoninic acid assay).

5.2.3 Tryptic digestion of extracted proteins

In solution tryptic digestion of sodium deoxycholate extracted proteins

20 μ g of SDC buffer extracted protein was used and diluted ad 100 μ l total volume with SDC buffer. For reduction of disulfide bonds 1 μ l 1 M dithiothreitol (DTT) was added (final concentration 10 mM) and samples incubated for 30 min at 56 °C in a heating block. For blocking reduced cysteines, 4 μ l 0.5 M iodoacetamide (IAA) was added (final concentration 20 mM). Samples were incubated for 30 min in the dark at 37 °C. For enzymatic cleavage 0.2 μ g trypsin was added in order to reach a trypsin-to-protein ratio of 1:100. Tryptic digestion was performed over night at 37 °C. To stop the reaction and precipitate SDC, formic acid (FA) was added to 2 % final concentration. Then samples were centrifuged for 5 min at 16,000 g and obtained supernatant transferred in a new tube. Before LC-MS/MS measurement collected supernatants were dried in a vacuum centrifuge and resuspended with a concentration of 1 μ g/ μ l in 0.1 % FA.

Filter aided sample preparation (FASP) for urea and sodium dodecyl sulfate extracted proteins

Protein extracts obtained by Urea and SDS containing buffers were further processed with the FASP method⁶⁷. For this, 20 μ g of protein extracts were taken and transferred into a centrifuge spin filter with a molecular weight cut-off of either 10 kDa or 30 kDa (Amicon Ultra 0.5 ml, Merck Millipore (Billerica, Massachusetts, USA)). Then, 200 μ l of 8 M urea in

50 mM ammonium hydrogen carbonate, further called UA solution, were added and samples inside the filters centrifuged 10 min at 14,000g. This step was repeated once. For reduction 50 μ l 10 mM DTT was added and the samples incubated for 30 min at 56 °C. Then samples were centrifuged for 5 min at 14,000 g. Then 50 μ l 20 mM IAA was added for alkylation. Incubation was done 30 min at 37 °C in the dark. Samples were again 5 min centrifuged at 14,000 g. Then samples were washed twice by adding 100 μ l UA solution and 10 min centrifugation at 14,000 g, followed by washing two times with 200 μ l 50 mM ammonium hydrogen carbonate and centrifugation for 10 min at 14,000 g. For enzymatic cleavage, trypsin was added in a 1 to 100 ratio and proteolysis was done at 37 °C overnight. Generated peptides were collected in a new reaction tube by 10 min centrifugation at 14,000 g and dried in a vacuum centrifuge. Prior to LC-MS/MS analysis, samples were resuspended in a concentration of 1 μ g/ μ l in 0.1 % FA.

Capturem™ Kit for tryptic proteolysis of Urea extracted proteins

20 μ g of proteins from urea extracted samples were adjusted to 50 μ l final volume by adding 8 M urea buffer. For reduction 0.5 μ l 1 M DTT was added (final concentration 10 mM) and the samples incubated for 30 min at 56 °C, followed by alkylation with 2 μ l 0.5 M IAA (final concentration 20 mM) for 30 min at 37 °C in the dark.

Capturem™ column membrane was activated with 200 μ l activation buffer and centrifugation at 500 g for 1 min. Then sample was loaded and digested by centrifugation at 500 g for 1 min. Obtained flow through containing the generated peptides was desalted by Sep-Pak 1cc (100 mg) C18 cartridges (Waters, Milford, Massachusetts, USA). For this, cartridges were activated by adding 1 ml methanol, followed by 1 ml 80 % ACN/0.1 % FA. Equilibration was done with two times 1 ml of 0.1 % FA. Samples were adjusted to 1 ml end volume by adding 0.1 % FA and loaded onto the cartridge. Loaded samples were washed two times with 1 ml 0.1 % FA before elution with 1 ml 80 % ACN/0.1 % FA. Eluted peptides were dried in a speedvac and resuspended in a concentration of 1 μ g/ μ l in 0.1 % FA before LC-MS/MS analysis.

5.2.4 Testing sample preparation by weight-to-volume ratio of extraction buffer

Proteins from ground truffle powder were extracted according to the SDC protocol. In short, 14 μ l buffer for 1 mg of sample was added, samples were boiled, sonicated and centrifuged before a BCA assay was performed. For this test four different *Tuber* species were chosen including two samples of the same species from geographically near origins see Table 8.

Table 8: Tuber species with corresponding origin used for testing the sample preparation by weight of the amount of truffle powder used.

| Species | Origin |
|----------------|---------|
| Tuber indicum | China |
| Tuber sinense | China |
| Tuber aestivum | Romania |
| Tuber aestivum | Hungary |
| Tuber magnatum | Italy |

After performing a BCA assay, a mean concentration value over all samples was calculated. For tryptic digestion, the obtained mean concentration value over all samples was used to take a volume according to 20 µg of sample amount. For comparison all samples were additionally processed according to the individual protein concentrations obtained by BCA assay and 20 µg were digested as well. All samples were processed in technical triplicates. LC-MS/MS measurements were done on a quadrupol-orbitrap (Q Exactive, Thermo Fisher) mass spectrometer coupled to a UPLC system (nanoAcquity, Waters).

5.2.5 Testing different incubation times and enzyme-to-protein ratio for tryptic proteolysis

Proteins from ground truffle powder of one truffle sample were extracted according to the SDC. In short, 14 µl buffer/mg sample was added, the samples were boiled, sonicated and centrifuged before a BCA test was performed. According to concentration by BCA test, 20 µg of were further processed. Samples were reduced and alkylated. Then, the tryptic proteolysis was performed for either 1 hour or overnight. For both proteolysis durations enzyme-to-protein ratios of 1:25 and 1:100 were tested. LC-MS/MS measurements were done on a quadrupol-orbitrap (Q Exactive, Thermo Fisher) mass spectrometer coupled to a UPLC system (nanoAcquity, Waters).

5.3 LC-MS/MS by data dependent acquisition (DDA) mass spectrometry

5.3.1 LC-MS/MS measurements

LC-MS/MS measurements were done on a quadrupol-orbitrap (Q Exactive, Thermo Fisher) mass spectrometer coupled to a UPLC system (nanoAcquity, Waters). For analysis, 1 µg of peptides were injected by the autosampler onto an reversed phase trapping column (Acquity UPLC® Symmetry C18; 100 Å pore size, 5 µm particle diameters, 180 µm x 20 mm) and

separated on a reversed phase separation column (Acquity UPLC® Peptide BEH C18; 130 Å pore size, 1.7 µm particle diameters, 75 µm x 200 mm). Trapping was done for 5 min at a flow rate of 15 µl/min with 99 % solvent A (0.1 % FA in water) and 1 % solvent B (0.1 % FA in ACN). Separation and elution of peptides was done by a linear gradient from 1 to 30 % solvent B in 60 min. Then solvent B was increased to 95 % in 1 min, hold for 2 min and decreased to 1 % in 1 min. Equilibration of the column was done for 15 min at 1 % solvent B.

Eluting peptides were infused in a quadrupole-orbitrap mass spectrometer (Q Exactive). Parameters for MS spectra acquisition are listed in Table 9.

Table 9: Parameters for MS1 spectra acquisition (DDA measurements).

| Parameter | Value |
|------------------------|--------------|
| Acquisition Mode | Positive |
| Scan Range | 400-1200 m/z |
| Orbitrap Resolution | 70 000 |
| AGC Target | 1E6 |
| Maximum Injection Time | 240 ms |

Parameters for MS/MS spectra acquisition are listed in Table 10.

Table 10: Parameters for MS/MS spectra acquisition (DDA measurements).

| Parameter | Value |
|-----------------------------|-----------------------------------|
| Acquisition Mode | Positive |
| Included Charge States | 2-5 |
| Data Dependent Acquisition | Top 15 |
| Precursor Priority | Highest Intensity (Threshold 1E4) |
| Isolation Window | 2 Da |
| Orbitrap Resolution | 17 500 |
| AGC Target | 1E5 |
| Maximum Injection Time | 50 ms |
| Normalized Collision Energy | 28 % |
| Dynamic Exclusion | 20 s |

5.3.2 Data processing

Raw files were processed with the MaxQuant⁶⁸ (Version 1.6.2.10) software. All samples were handled as individual experiments. Proteins were digested in silico with trypsin and up to two missed cleavages were allowed. A minimal peptide length of 6 amino acids and maximum peptide mass of 6,000 Da was defined. Allowed variable modifications were oxidation of methionine, acetylation of protein N-termini and the conversion of glutamine to pyroglutamic acid. Fixed modification was the carbamidomethylation of cysteines. Error tolerance for the first precursor search was 20 ppm, for the following main search 4.5 ppm. Fragment spectra were matched with a 20 ppm error tolerance. For identification the Perigord Black Truffle (Strain Mel28) protein database was used (downloaded from UniProt 13.07.2017, 4,113 entries). False discovery rate for peptide spectrum matches and proteins was set to 1 %. For protein quantification all identified razor and unique peptides were considered. Match between runs and second peptides functions were enabled. For statistical analysis the ProteinGroups result file from MaxQuant was loaded into Perseus⁶⁸ (Version 1.5.8.5). Quantitative LFQ Intensity values for protein groups were used as main columns. These quantitative values for all protein groups were transformed in log2 values and normalized by the median protein area and Anova Test, Student's T-Test, hierarchical clustering and principal component analysis was performed. Anova Test and Student's T-Test was done with a permutation based corrected p-value of 5 %.

5.4 LC-MS/MS by data independent (DIA) mass spectrometry

5.4.1 High pH reversed phase fractionation for spectral library generation

Pooled samples of *T. aestivum*, *T. magnatum* and *T. melanosporum* were used for high pH fractionation. For each species, 5 samples containing the same amount of peptides solved in 10 mM ammonium hydrogen carbonate with a pH of 8 were pooled. From each sample 15 µg were taken, mixed together and vortexed. From pooled samples for each species 50 µg were injected for high pH fractionation. High pH (HpH) Fractionation was done on an HPLC (Agilent 1200er series) connected to a fractionator (Äkta Prime). Peptides were separated on a polymer based monolithic reversed phase column (Thermo ProSwift™ RP-4H) with the dimensions 1 x 250 mm. Flow rate was set to 200 µl / min. Solvent A was 10 mM ammonium hydrogen carbonate, solvent B 80 % ACN/10 mM ammonium hydrogen carbonate. In both solvents, pH was checked to be around a value of 8. Listed in Table 11 is the linear gradient used for separation.

Table 11: Gradient for separation of peptides by high pH reversed phase liquid chromatography.

| Time [min] | % Solvent B |
|------------|-------------|
| 0 | 3.3 |
| 5 | 3.3 |
| 25 | 38.5 |
| 26 | 95 |
| 36 | 95 |
| 37 | 3.3 |
| 45 | 3.3 |

From the high pH reversed phase run, 29 fractions were collected. Each fraction consisted of 200 μ l, corresponding to 1 minute LC run time. First 9 minutes were pooled into three fractions while the rest of the fractions were pooled by a concatenated proceeding according to Table 12.

Table 12: Concatenated pooling scheme of fractions collected by offline high pH reversed phase fractionation on a HPLC.

| Collected Fractions | Final Fractions |
|---------------------|-----------------|
| 1 + 2 + 3 | Fraction -3 |
| 4 + 5 + 6 | Fraction -2 |
| 7 + 8 + 9 | Fraction -1 |
| 10 + 20 | Fraction 1 |
| 11 + 21 | Fraction 2 |
| 12 + 22 | Fraction 3 |
| 13 + 23 | Fraction 4 |
| 14 + 24 | Fraction 5 |
| 15 + 25 | Fraction 6 |
| 16 + 26 | Fraction 7 |
| 17 + 27 | Fraction 8 |
| 18 + 28 | Fraction 9 |
| 19 + 29 | Fraction 10 |

Pooled fractions were dried in a speedvac and resolved in 0.1 % FA before LC-MS/MS analysis.

5.4.2 Spectral library generation by LC-MS/MS measurements in DDA mode

LC-MS/MS measurements were done on a quadrupole-iontrap-orbitrap mass spectrometer (Orbitrap Fusion™ Tribrid™, Thermo Fisher) coupled to a UPLC system (Dionex Ultimate 3000). For analysis, 1 µg of peptides were injected by the autosampler onto a reversed phase trapping column (Acquity UPLC® Symmetry C18; 100 Å pore size, 5 µm particle diameters, 180 µm x 20 mm) and separated on a reversed phase separation column (Acquity UPLC® Peptide BEH C18; 130 Å pore size, 1.7 µm particle diameters, 75 µm x 200 mm). Trapping was done for 5 min at a flow rate of 10 µl/min with 99 % solvent A (0.1 % FA in water) and 1 % solvent B (0.1 % FA in ACN). Then solvent B was increased to 95 % in 1 min, hold for 2 min and decreased to 1 % in 1 min. Equilibration of the column was done for 15 min at 1 % solvent B.

Eluting peptides were infused in a quadrupole-iontrap-orbitrap mass spectrometer (Orbitrap Fusion™ Tribrid™) and measured in data-dependent acquisition mode. Parameters for MS spectra acquisition are listed in Table 13.

Table 13: Parameters for MS1 spectra acquisition (DDA measurements for spectral library).

| Parameter | Value |
|------------------------|--------------|
| Acquisition Mode | Positive |
| Scan Range | 390-1010 m/z |
| Detector Type | Orbitrap |
| Orbitrap Resolution | 120 000 |
| AGC Target | 2E% |
| Maximum Injection Time | 120 ms |

Parameters for MS/MS spectra acquisition are listed in Table 14.

Table 14: Parameters for MS/MS spectra acquisition (DDA measurements for spectral library).

| Parameter | Value |
|-----------------------------|-----------------------------------|
| Acquisition Mode | Positive |
| Included Charge States | 2-5 |
| Data Dependent Acquisition | Top Speed (Cycle Time 3 s) |
| Precursor Priority | Highest Intensity (Threshold 1E4) |
| Isolation Mode | Quadrupole |
| Isolation Window | 1.6 Da |
| Detector | Orbitrap |
| Orbitrap Resolution | 15 000 |
| AGC Target | 1E5 |
| Maximum Injection Time | 50 ms |
| Normalized Collision Energy | 30 % |
| Dynamic Exclusion | 15 s |

5.4.3 Data processing for spectral library generation

Raw files were processed with the ProteomeDiscoverer (Version 2.0) software. All three species were handled as single experiments. Collected fractions of each species were processed as fractions, consisting of 13 raw files. Proteins were digested in silico with trypsin and up to two missed cleavages were allowed. A minimal peptide length of 6 amino acids and maximal peptide length of 144 amino acids was defined. Allowed variable modifications on peptide level were oxidation of methionine and the conversion of glutamine to pyro-glutamic acid at the N-terminus. On protein level loss of starter methionine, acetylation and a combination of both on N-termini was allowed. Fixed modification was carbamidomethylation of cysteines. Error tolerance for precursor search was 10 ppm. Fragment spectra were matched with a 20 ppm error tolerance. For identification the corresponding protein database files for each *Tuber* species were used, *Tuber magnatum* (TrEMBL, 9,412 entries, downloaded on 06.04.20), *Tuber aestivum* (TrEMBL, 9,311 entries, downloaded on 06.04.2020) and *Tuber melanosporum* (SwissProt and TrEMBL, 7,494 entries, downloaded on 06.04.20). Proteome Discoverer Result files for the three individual species were merged to a final combined result file.

5.4.4 LC-MS/MS measurements in DIA mode

LC-MS/MS measurements in data-independent acquisition mode were done on a quadrupole-orbitrap-iontrap hybrid (Orbitrap Fusion, Thermo Fisher) mass spectrometer coupled to a UPLC system (Dionex Ultimate 3000). For analysis, 1 μg of peptides were injected by the autosampler onto an reversed phase trapping column (Acquity UPLC® Symmetry C18; 100 Å pore size, 5 μm particle diameters, 180 μm x 20 mm) and separated on a reversed phase separation column (Acquity UPLC® Peptide BEH C18; 130 Å pore size, 1.7 μm particle diameters, 75 μm x 200 mm). Trapping was done for 5 min at a flow rate of 10 $\mu\text{l}/\text{min}$ with 99 % solvent A (0.1 % FA in water) and 1 % solvent B (0.1 % FA in ACN). Then solvent B was increased to 95 % in 1 min, hold for 2 min and decreased to 1 % in 1 min. Equilibration of the column was done for 15 min at 1 % solvent B.

Eluting peptides were infused in a quadrupole-iontrap-orbitrap mass spectrometer (Orbitrap Fusion™ Tribrid™) and measured in data-independent acquisition mode. Parameters for MS spectra acquisition are listed in Table 15.

Table 15: Parameters for MS1 spectra acquisition (DIA measurements).

| Parameter | Value |
|------------------------|--------------|
| Acquisition Mode | Positive |
| Scan Range | 390-1010 m/z |
| Detector Type | Orbitrap |
| Orbitrap Resolution | 60 000 |
| AGC Target | 2E5 |
| Maximum Injection Time | 60 ms |

Parameters for MS/MS spectra acquisition are listed in Table 16.

Table 16: Parameters for MS/MS spectra acquisition (DIA measurements).

| Parameter | Value |
|-----------------------------|--------------|
| Acquisition Mode | Positive |
| Isolation Mode | Quadrupole |
| Isolation Window | 20 m/z |
| Loop Count | 15 |
| Detector | Orbitrap |
| Orbitrap Resolution | 30 000 |
| AGC Target | 1E5 |
| Maximum Injection Time | 50 ms |
| Normalized Collision Energy | 28 % |

30 DIA windows of 20 m/z were set and adjusted by the optimize window placement function in Skyline (version 20.1)⁶⁹. The corresponding start and end m/z values for these windows are listed in Table 17.

Table 17: DIA windows with corresponding start and end m/z values.

| Start m/z | End m/z |
|------------------|----------------|
| 400.4550 | 420.4550 |
| 420.4650 | 440.4650 |
| 440.4750 | 460.4750 |
| 460.4850 | 480.4850 |
| 480.4950 | 500.4950 |
| 500.5050 | 520.5050 |
| 520.5150 | 540.5150 |
| 540.5250 | 560.5250 |
| 560.5350 | 580.5350 |
| 580.5450 | 600.5450 |
| 600.5550 | 620.5550 |
| 620.5650 | 640.5650 |
| 640.5750 | 660.5750 |
| 660.5850 | 680.5850 |
| 680.5950 | 700.5950 |
| 700.6050 | 720.6050 |
| 720.6150 | 740.6150 |
| 740.6250 | 760.6250 |
| 760.6350 | 780.6350 |
| 780.6450 | 800.6450 |
| 800.6550 | 820.6550 |
| 820.6650 | 840.6650 |
| 840.6750 | 860.6750 |
| 860.6850 | 880.6850 |
| 880.6950 | 900.6950 |
| 900.7050 | 920.7050 |
| 920.7150 | 940.7150 |
| 940.7250 | 960.7250 |
| 960.7350 | 980.7350 |
| 980.7450 | 1,000.7450 |

5.4.5 Data processing of DIA measurements

Spectra library generation and data processing of DIA measurements was done with Skyline (version 20.1)⁶⁹. For the spectral library, the combined Proteome Discoverer Result file of the high pH fractionated truffle samples and a combined background fasta database were loaded into Skyline.

For extraction of the transitions, the transition settings were set to allowed precursor charges of 2, 3, 4 and 5. Product ions were set to b-ions, y-ions and precursor. Further, product ions were set to 5 with a minimum of 4 product ions. Window extraction was done according to the measurement method and with chromatogram matching in a 5 min window. Repeated peptides were removed. Exported results were filtered by a dotp value of at least 0.85 in the best sample. Next, the exported result file contained the peptides total area fragments and was consolidated to generate values for proteins. This consolidated result file from Skyline was loaded into Perseus (Version 1.6.2.3) for statistical analysis. Quantitative total area fragment values were used as main columns. Then quantitative values for all proteins were transformed in log₂ values and normalized by the median protein area. Anova Test, Student's T-Test, hierarchical clustering and principal component analysis was performed. Anova Test and Student's T-Test was done with a permutation based corrected p-value of 5 % or 1 % as indicated in the results.

5.4.6 Functional annotation of differentially regulated proteins

Tuber magnatum and *Tuber melanosporum* annotation files were downloaded from the Gene Ontology Annotation website (<https://www.ebi.ac.uk/GOA/proteomes>). *Tuber aestivum* sequences (9,312 entries from UniProt TrEMBL database) were annotated with an annotation tool from the Gene Ontology Annotation website (<https://www.ebi.ac.uk/QuickGO/annotations>). A combined annotation file for the three species was created. A gene ontology file was downloaded from the Gene Ontology website (<http://geneontology.org/docs/download-ontology/>) in the obo 1.2 format. Functional annotation of proteins was performed with BiNGO⁷⁰ version 3.0.4 as a plugin for the Cytoscape⁷¹ version 3.8.0 platform. Parameters for finding over-represented categories after correction of biological processes were hypergeometric testing with Benjamini & Hochberg false discovery rate correction with a significance level of 0.05.

6. Results

6.1 Mass spectrometry based profiling of truffles

For a fast identification and thorough characterization of truffle species, both fast and deep profiling on proteomic level was performed. Fast profiling was done by top-down MALDI-TOF analysis of intact proteins while deep profiling used a LC-MS/MS based bottom-up proteomics approach, analyzing enzymatically generated peptides. A schematic overview of the procedure is depicted in Figure 9.

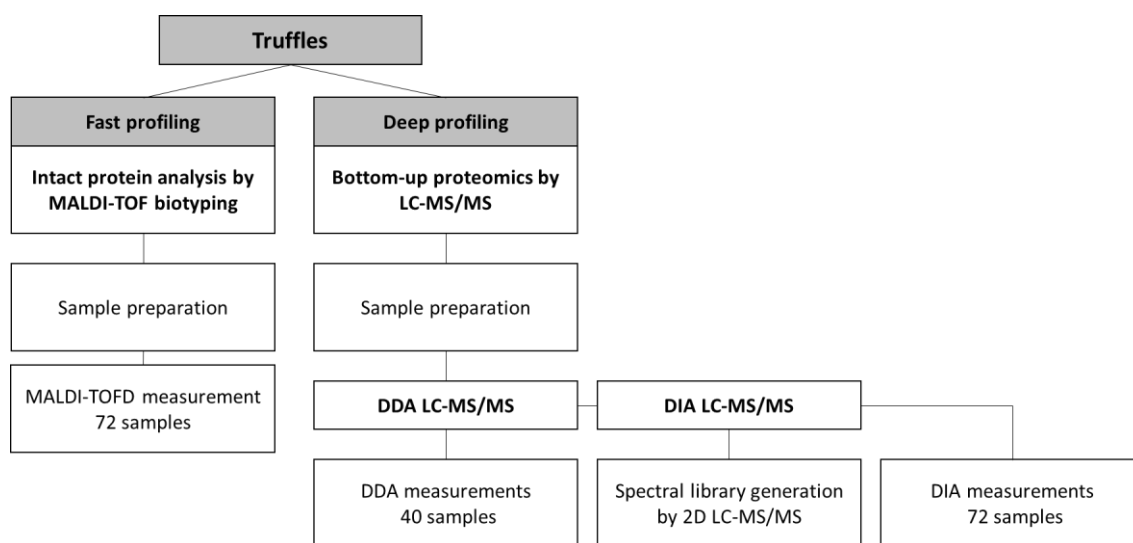


Figure 9: Schematic overview of the procedure for fast and deep profiling of truffle species.

Fast profiling should lead to a quick identification of truffles to uncover possible cases of food fraud. With deep profiling a comprehensive characterization of truffles and generation of quantitative protein profiles for a specific fingerprint should be done. Both combined will enable the identification of truffles and reveal possible cases of food fraud.

6.2 Fast profiling by top-down MALDI-TOF analysis of truffles

6.2.1 Testing sample preparation for MALDI-TOF analysis

First aim of this thesis was to develop a fast way to identify truffle samples. MALDI-TOF biotyping of intact proteins is a fast and reliable way to detect and identify bacteria in food products or in hospital laboratories. This approach should be transferred to truffles. First step before MALDI-TOF analysis was the extraction of proteins from the truffle samples. A widely used preparation method in microbiology is a formic acid and acetonitrile based protein extraction. Here proteins are extracted, then spotted on the MALDI target and overlaid with matrix solution. A second approach is a combined extraction and MALDI matrix

solution preparation. In the combined solution approach proteins are extracted in a solution already containing the MALDI matrix. Therefore an extract can be directly spotted and crystallized without a second step for matrix application. Both mentioned preparation methods for MALDI-TOF biotyping were tested on a ground truffle powder sample. Triplicates for extraction were done for both approaches. Resulting protein extracts were spotted in triplicates on a ground steel target and measured by a MALDI-TOF system. Generated MALDI-TOF spectra from extracted proteins are shown in Figure 10.

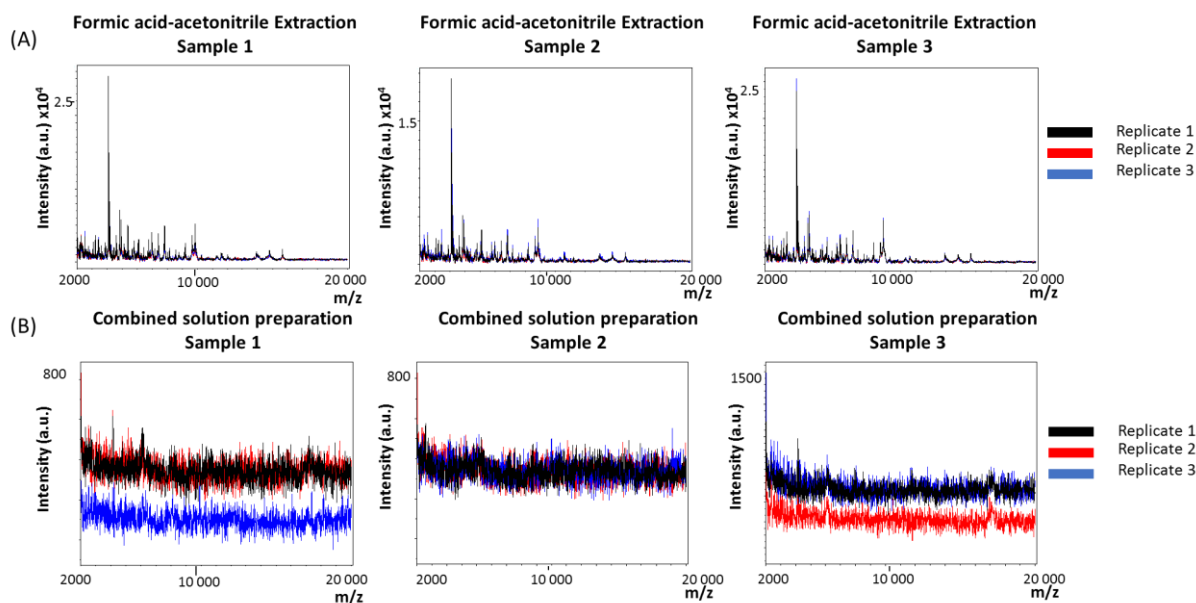


Figure 10: MALDI-TOF spectra for proteins extracted from a truffle sample by formic acid-acetonitrile extraction and combined solution preparation. Three replicates were extracted with both methods and spotted in triplicates. The figure is showing the overlaid spectra from the spotting triplicates for each extracted sample. (A) Shows the spectra for samples extracted by formic acid-acetonitrile and (B) spectra for samples extracted with the combined solution preparation method.

Samples prepared by formic acid-acetonitrile based protein extraction showed a high number of signals with sufficient intensity and signal-to-noise ratio in MALDI-TOF spectra. Repeatability in spotting replicates within each sample was high, indicated by the overlay of spectra. Further the repeatability of extraction was well, as the same characteristic signals are present in all three extraction samples. Examples for these characteristic signals are the highest signal in all spectra at 4,000 m/z, the signal at 10,000 m/z and the triplet of signals between 14,000 m/z and 16,000 m/z. In contrast, all samples prepared by the combined extraction approach didn't show any signals with sufficient signal intensity or sufficient signal-to-noise ratio that indicated proteins. Therefore this combined protein extraction and

crystallization approach was not suitable for intact protein analysis from truffle powder samples.

Next repeatability of extraction and repeatability of measurement on the MALDI-TOF instrument was assessed. Formic acid-acetonitrile extraction with subsequent MALDI TOF MS spectra acquisition was done in triplicates on three different days. On each day proteins from the same ground truffle powder sample were extracted three times. Each of these extraction replicates was spotted three times on a ground steel target. Acquired spectra are shown in Figure 11.

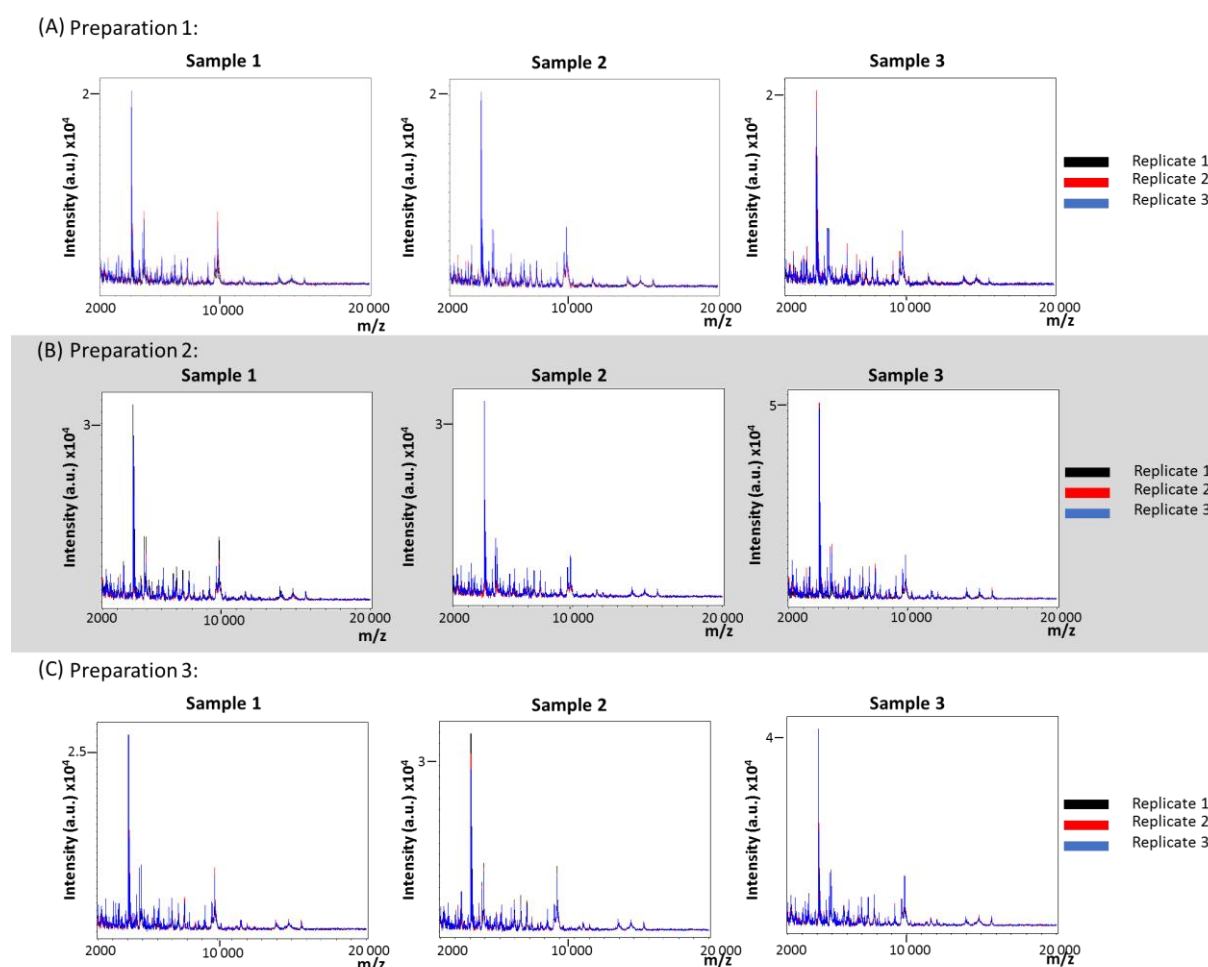


Figure 11: MALDI-TOF spectra for proteins extracted from a truffle sample by formic-acid/acetonitrile extraction on three different days in three individual preparations. Shown are the generated MALDI-TOF spectra for (A) preparation 1, (B) preparation 2 and (C) preparation 3.

All spectra for the individually handled samples extracted and measured on different days looked similar, indicated by their well-fitting overlay. Distinct signal patterns were shared between all samples. In all generated spectra signals at 4,000 m/z, 10,000 m/z and a signal

triplet between 14,000 m/z and 16,000 m/z were present. Further the signal intensity of all acquired spectra were in a similar range with highest intensity values of 2×10^4 to 4×10^4 a.u..

To further test if the formic acid-acetonitrile approach was suitable for the identification of different truffles, three different truffle species were prepared with it for MALDI-TOF measurements. From each of the truffle species three different biological samples were used. Each biological sample was extracted in three technical replicates. Resulting protein extracts were spotted on a ground steel target and measured on a MALDI-TOF system. Resulting spectra of different truffle species are displayed in Figure 12.

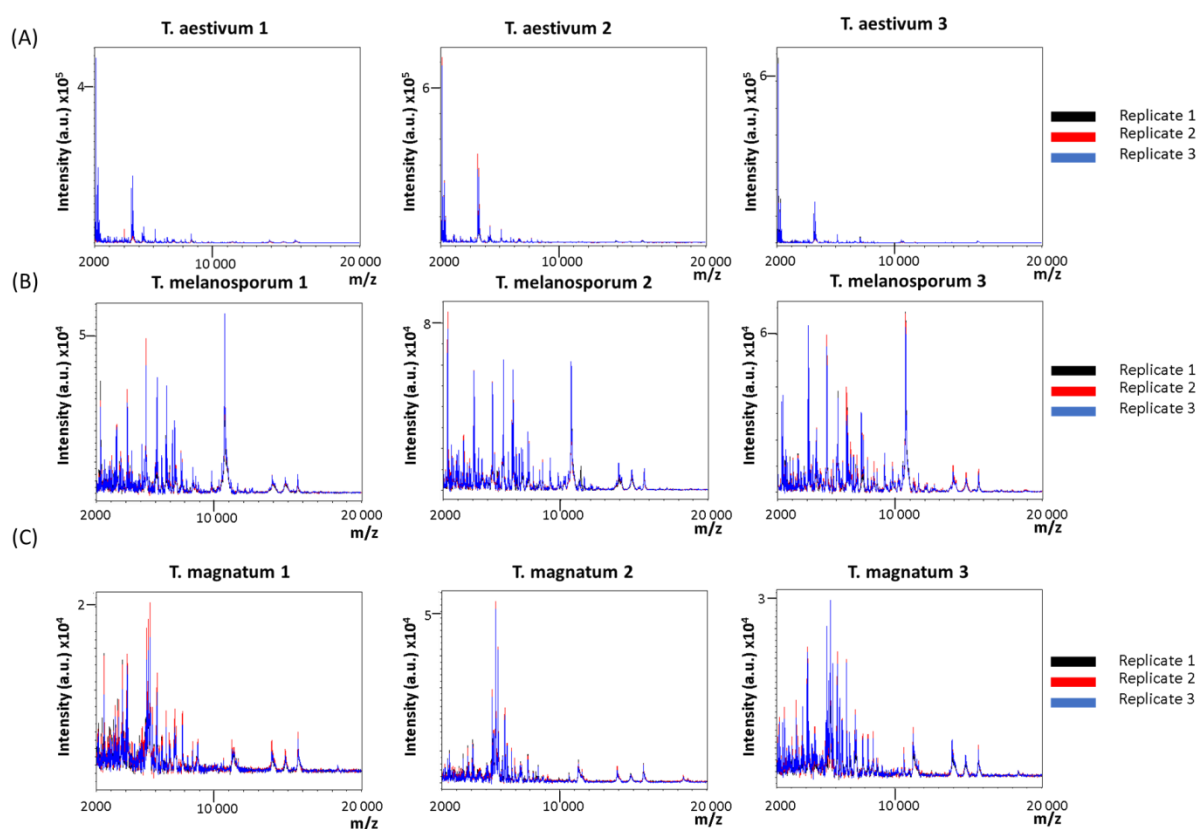


Figure 12: MALDI-TOF spectra for protein extracts from different truffle species. For each truffle species, three biological samples were used. The figure is showing the overlaid spectra from three separately extracted replicates of each biological sample. Species used were (A) *Tuber aestivum*, (B) *Tuber melanosporum* and (C) *Tuber magnatum*.

Acquired spectra of the three used truffle species differed greatly, while the spectra within a species showed high similarity. Each species showed reproducible and distinct signals in the spectra recorded for them. For *T. aestivum* a most prominent and specific signal was present at 2,200 m/z, for *T. melanosporum* at 10,900 m/z with many additional signals between 2,000 and 8,000 m/z. Further a signal triplet in the higher mass range between 14,000 and 16,000 m/z was observed. Similar to *T. melanosporum*, spectra for *T. magnatum* showed

many signals in the range of 2,000 to 8,000 m/z and again the same signal triplet in the higher mass range. Nevertheless *T. magnatum* lacked the characteristic signal of *T. melanosporum* at 10,900 m/z . Repeatability of protein extraction and spectra acquisition was high, indicated by the well-fitting overlay of all technical replicate spectra for each biological sample. Next, recorded spectra were processed and used for a principal component analysis and clustering analysis with integrated tools of the MassUp software. Sample projections from the principal component analysis and the dendrogram after hierarchical clustering analysis are shown in Figure 13.

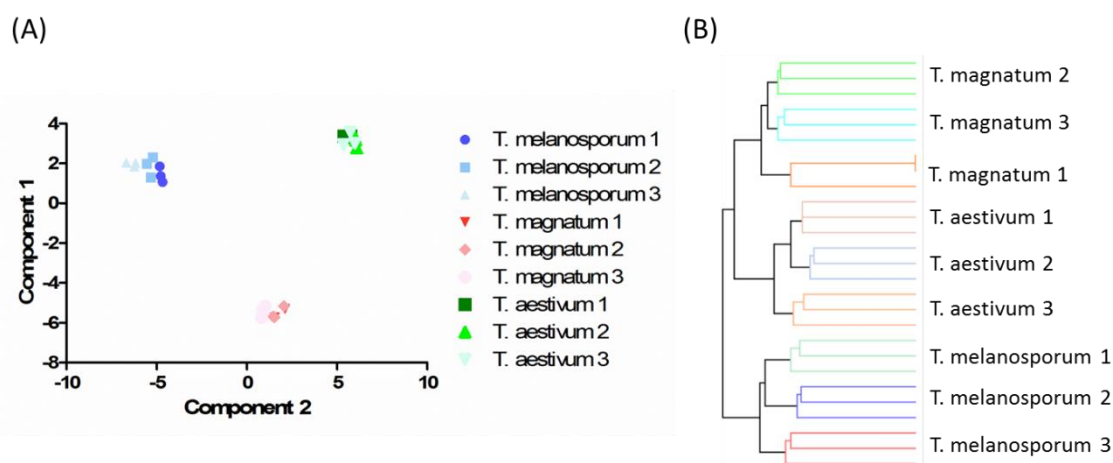


Figure 13: (A) Sample projection of the first two components based on principal component analysis (PCA) and (B) dendrogram showing the degree of similarity for different truffle species after hierarchical clustering analysis.

Sample projection of the first two components based on a principal component analysis showed a clear separation of different truffle species by their signal pattern and its intensity in MALDI-TOF spectra. All biological samples and corresponding technical replicates clustered together very closely. By this two-dimensional projection of principal components a further differentiation between biological and technical replicates was not possible, as they did overlay in the projection. For a more in depth differentiation, hierarchical clustering analysis by similarity in MALDI-TOF spectra was performed. Clustering of biological replicates from different species was shown in the resulting dendrogram. Further all technical replicates of each biological sample clustered accordingly. Therefore the chosen approach of MALDI-TOF for different truffle samples was a specific and reproducible way to differentiate species. Sample preparation by formic acid-acetonitrile extraction with chosen parameters for spectra acquisition on a MALDI-TOF instrument were suitable and further used in following experiments.

6.2.2 MALDI-TOF measurement of truffle samples

After testing for a suitable sample preparation and parameters for measurement, MALDI-TOF analysis was done for 72 biological truffle samples (28 *T. aestivum*, 4 *T. Albidum Pico*, 10 *T. indicum*, 19 *T. magnatum*, 11 *T. melanosporum*).

Generated MALDI-TOF spectra were imported in the MassUp software and further processed. After baseline correction and smoothing of spectra, intra-sample peak picking within all the different samples of a species was done. Peaks had to be present in at least 50 % of samples from each truffle species. First a hierarchical clustering analysis of samples based on the presence of signals in the spectra was done. Resulting dendrogram after hierarchical clustering is shown in Figure 14.

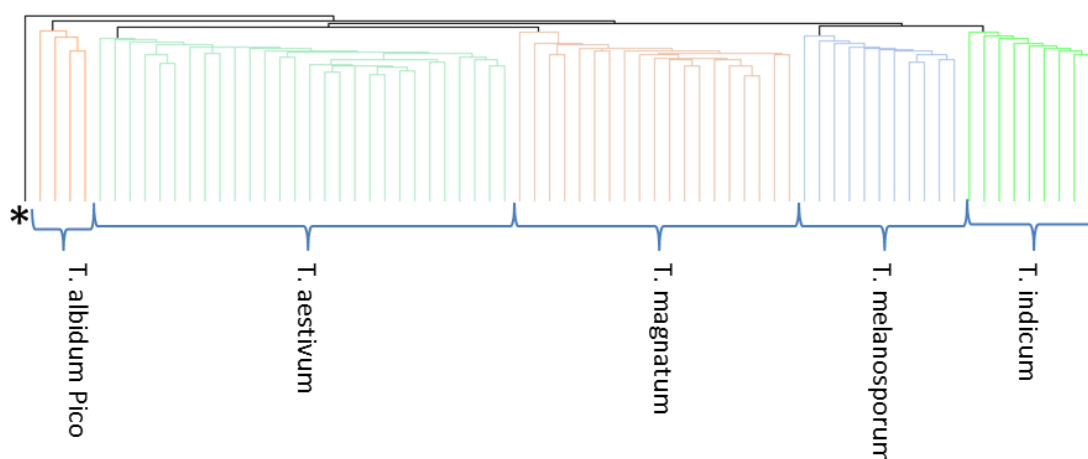


Figure 14: Dendrogram showing the degree of similarity for truffle samples of different species after hierarchical clustering analysis. An occurring outlier is marked by * - Outlier: *T. indicum*.

Generated dendrogram after hierarchical clustering analysis showed, that samples within a species are most similar. All samples for one species are grouped together and each species is separated in the dendrogram from others. One outlier was occurring. A sample of *T. indicum* didn't have similarity to other samples of its species and did not cluster within any of the other ones. Processed spectra with corresponding signals were then used for a principal component analysis. The sample projection is depicted in Figure 15.

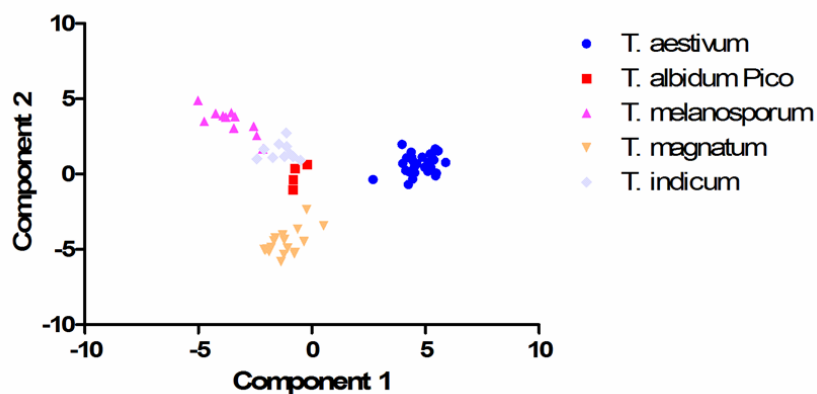


Figure 15: MALDI-TOF sample projection of the first two components based on principal component analysis (PCA).

Similar to the previous dendrogram, separation of samples from different truffle species could be observed in the sample projection after principal component analysis. All samples within a species clustered closely together while different species were separated in the projection. Here, the outlier of *T. indicum* is not visible. Nevertheless, it was removed from further analysis.

After removal of the outlier, acquired spectra for samples of each species were imported into the mMass software. Here, obtained signal intensities were averaged to generate consensus spectra. These consensus spectra could be used as MALDI-TOF profiles for truffle species. Profiles are shown in Figure 16.

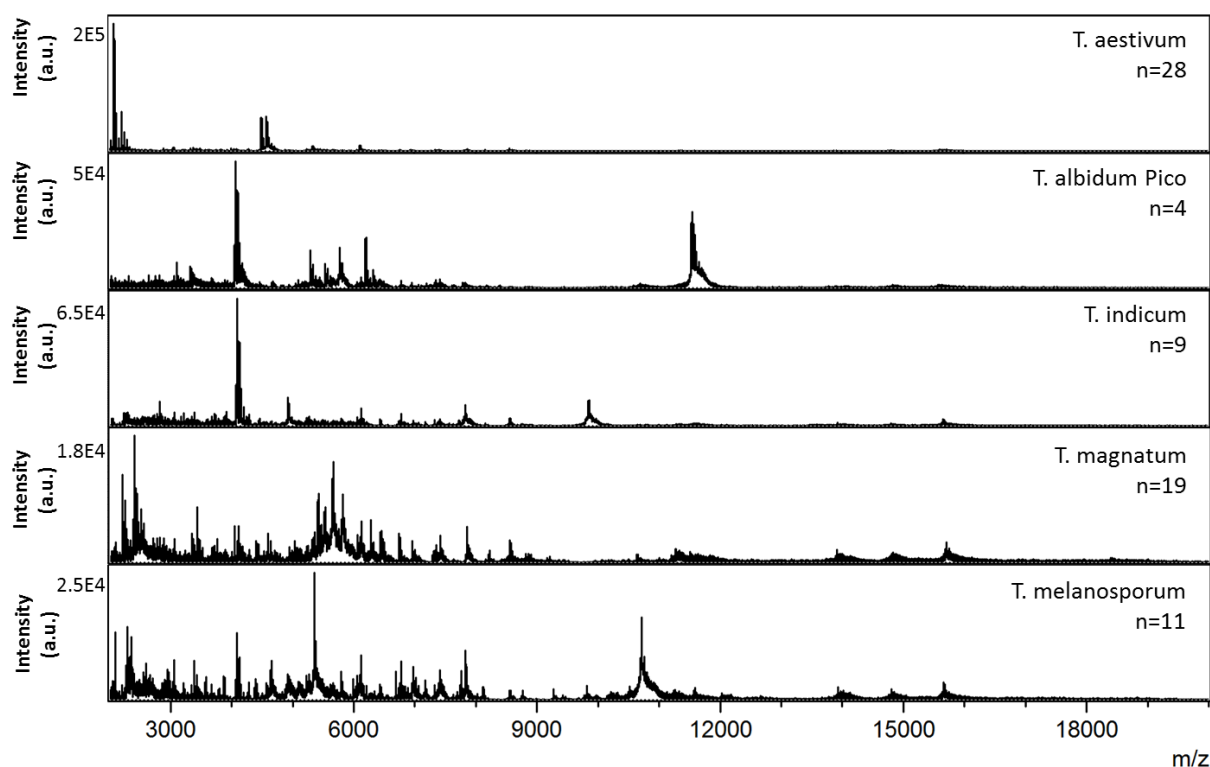


Figure 16: MALDI-TOF profiles of different truffle species. Profiles were generated by averaging signal intensity of all acquired MALDI-TOF spectra of samples within a species. Number of samples averaged within a species for generating the profile is indicated by the number n .

MALDI-TOF profiles of different truffle species differed strongly in recorded signal patterns. Highest general signal intensity was observed in *T. aestivum* with the strongest signal at 2E5 a.u. instead of 1.8-6.5E4 a.u. in other truffles. This strongest signal at 2,200 m/z in *T. aestivum* was also the most distinct signal for the species. Further some smaller peaks around 4,500 m/z could be only seen in *T. aestivum*. In contrast, *T. albidum Pico* had the most intense and distinct signal at 11,800 m/z. A second strong signal at 4,000 m/z was shared with *T. indicum*. Next, *T. indicum* had an additional and specific signal at 10,000 m/z. Both *T. magnatum* and *T. melanosporum* showed many signals in the scan range of 2,000 m/z to 8,000 m/z. These signals had areas of higher intensity peak assembly and were difficult to further distinguish. But the maxima of these peak assemblies were shifted to a slightly higher m/z range for *T. magnatum*. Here the strongest signals within peak assemblies were sitting at 2,500 m/z instead of 2,300 m/z and 5,800 m/z instead of 5,500 m/z. Additionally, both species shared a signal triplet in the higher scan range of 14,000 to 16,000 m/z. Specific for only *T. melanosporum* was a signal at 10,900 m/z.

Another way of displaying MALDI-TOF profiles for different truffle species is in form of a barcode. This barcode was derived from the averaged MALDI-TOF profile of each species.

High signal intensity in the profile will be transformed into a dark and thick line; lower intense signals will appear in lighter shades of grey. Absence of signals with sufficient intensity will appear as blank space in the barcode. Generated barcodes for the analyzed truffle species are displayed in Figure 17.

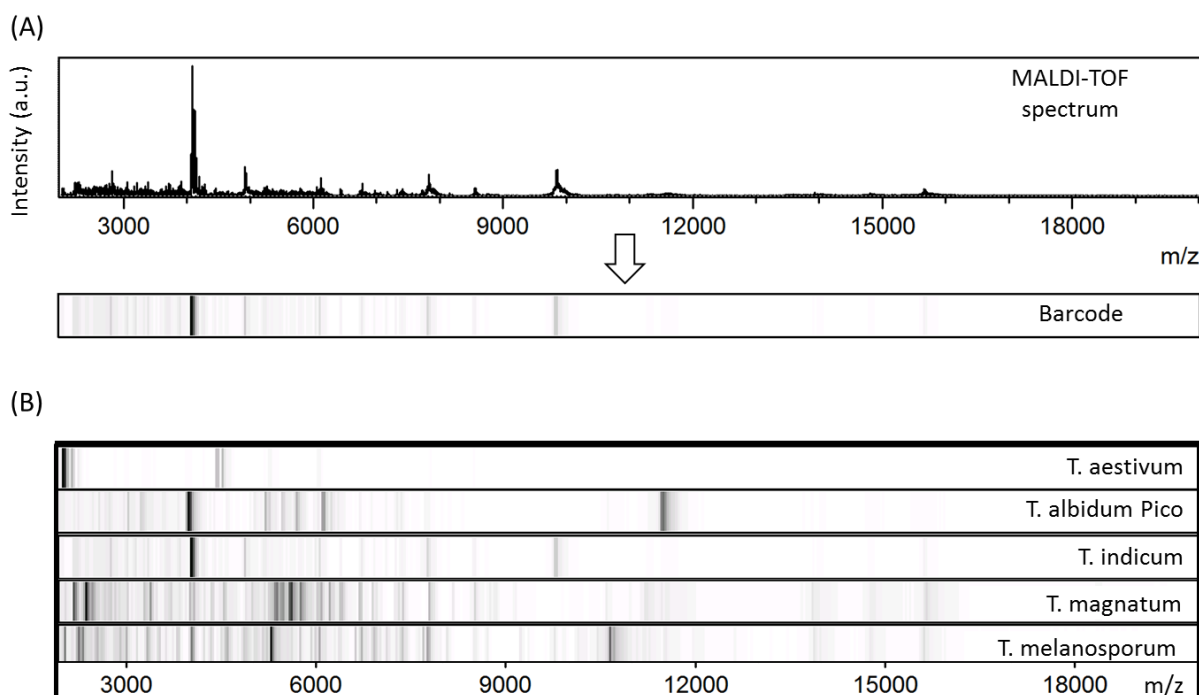


Figure 17: Signal intensity barcodes derived from MALDI-TOF spectra of different truffle species. Partition (A) shows how the signals and their intensity in the MALDI-TOF spectra are converted in a signal intensity barcode. Highly intense signals in the mass spectrum will be transferred into a darker and thicker line in the barcode than low intensity signals. In partition (B) the MALDI-TOF derived barcodes of different truffle species are shown.

In accordance to the generated MALDI-TOF profiles, differences between truffle species remained to be observable by transforming them into barcodes. By this barcode form of display it was even easier to spot differences between truffles. This is especially the case for comparison of *T. magnatum* with *T. melanosporum*. Here the shift of peak assemblies with higher intensity mentioned for the profiles is better visible.

Both generated MALDI-TOF profiles and barcodes clearly showed differences between truffles. Therefore identification of truffles was enabled. Next a classification analysis of truffle samples should be tested. Here, MALDI-TOF spectra of truffles were tried to be matched to a species. Classification analysis utilizes machine learning. Machine learning provides the ability to learn from examples and apply the gained knowledge to new instances. It's often used as a tool for predictions. In classification analysis, the obtained data set is split into trainings set and classification set. With the trainings set as used examples, the algorithm

tries to match the remaining classification set of data. Classification analysis of truffle samples based on the acquired and processed MALDI-TOF spectra is shown in Table 18.

Table 18: Classification analysis of truffle samples.

| | |
|----------------------------------|---|
| Instances in data set | 71 |
| Split percentage | 66% |
| Instances for training | 47 |
| Instances for classification | 24 |
| Correctly classified instances | 23 (96 %) NaiveBayes 21 (88 %) Random Forest |
| Incorrectly classified instances | 1 (4 %) NaiveBayes 3 (12 %) Random Forest |
| Classifier | NaiveBayes, Random Forest |

Confusion Matrix NaiveBayes:

| a | b | c | d | e | <-- Classified as |
|----|---|---|---|---|----------------------------|
| 12 | 0 | 0 | 0 | 0 | a = <i>T. aestivum</i> |
| 0 | 1 | 0 | 1 | 0 | b = <i>T. albidum Pico</i> |
| 0 | 0 | 4 | 0 | 0 | c = <i>T. melanosporum</i> |
| 0 | 0 | 0 | 4 | 0 | d = <i>T. magnatum</i> |
| 0 | 0 | 0 | 0 | 2 | e = <i>T. indicum</i> |

Confusion Matrix RandomForest:

| a | b | c | d | e | <-- Classified as |
|----|---|---|---|---|----------------------------|
| 12 | 0 | 0 | 0 | 0 | a = <i>T. aestivum</i> |
| 0 | 0 | 0 | 1 | 1 | b = <i>T. albidum Pico</i> |
| 0 | 0 | 3 | 1 | 0 | c = <i>T. melanosporum</i> |
| 0 | 0 | 0 | 4 | 0 | d = <i>T. magnatum</i> |
| 0 | 0 | 0 | 0 | 2 | e = <i>T. indicum</i> |

For classification analysis the set data set was split by percentage. From the 71 instances in total, 47 (66 %) were used for training and 24 (34 %) for classification. Classification was done successfully with naïve Bayes classifier in 23 of the 24 instances, corresponding to a correct matching in 95 % of instances. For *T. aestivum* all 12 instances were correctly matched, for *T. melanosporum* and *T. magnatum* all 4 and for *T. indicum* both instances. Mismatching was done for one instance of *T. albidum Pico*. From two classified samples, only one was matched correct. Another classifier was tested, the random forest classification, and led to 21 instances correctly (88 %) and 3 incorrectly (12 %) classified. Both

classification instances of *T. albidum Pico* and one instance of *T. melanosporum* were classified incorrectly.

6.3 Deep profiling and characterization by bottom-up proteomics based on LC-MS/MS

6.3.1 Testing of sample preparation

After establishing a fast way for a quick and easy identification of truffles by MALDI-TOF analysis of intact proteins, a more in-depth approach should be performed. For this, a bottom-up proteomic was chosen. In bottom-up proteomics extracted proteins are cleaved enzymatically into peptides and subsequently analyzed by LC-MS/MS. Established sample preparation protocols for extraction and proteolysis were tested and further simplified. Afterwards LC-MS/MS based analysis approaches for label-free quantification were performed to generate quantitative protein profiles, distinguish truffle species and do a comprehensive characterization.

Starting point for a bottom-up approach on truffles was the process of sample preparation. Sample preparation of truffles for LC-MS/MS based bottom-up proteomics consisted of two steps, protein extraction and subsequent tryptic proteolysis of extracted proteins into peptides.

To maximize surface area and increase protein yield, bead mill-ground truffle powder was used as material for protein extraction. For the extraction, buffers containing chaotropes, surfactants and detergents are widely used to disrupt membranes and setting proteins free. A ground powder truffle sample was mixed with three different extraction buffers in a 14 μ l buffer to 1 mg of truffle powder ratio. Used extraction buffers contained 1 % sodium deoxycholate (SDC), 8 M urea or 5 % sodium dodecyl sulfate (SDS). After extraction, the yielded protein amount was measured by a bicinchoninic acid (BCA) assay and is plotted in Figure 18.

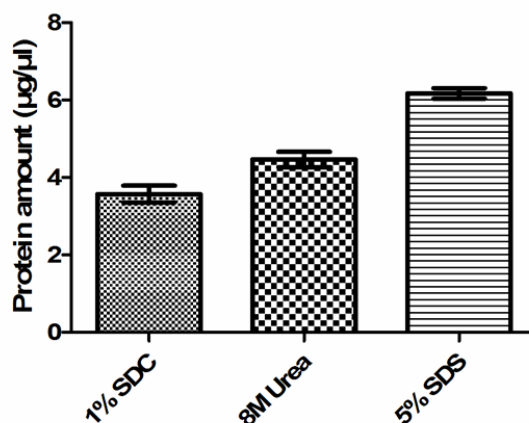


Figure 18: Bar plot comparing the extracted protein amount using different chaotropes/surfactants/detergents according to a performed BCA assay. Error bars indicate standard deviation of technical triplicates.

With buffer containing 1 % SDC measured protein amount after extraction was 3.6 µg/µl (± 0.3). Buffer containing 8 M urea yielded in 4.5 µg/µl (± 0.3) and with 5 % SDS in a protein concentration of 6.2 µg/µl (± 0.2). Therefore, extraction with 5 % SDS worked best in regard to extracted protein amount. Nevertheless, all three extraction buffers yielded in a protein concentration high enough for further processing of samples without any kind of limitation in sample material.

Drawback of using extraction agents is the need for removal of them prior to LC-MS/MS measurement. They show interference with the electrospray ionization of the mass spectrometer or interference in chromatographic separation of peptides. Typical bottom-up proteomic workflows take this into account and have steps included to remove these agents. Two typical workflows, the SDC protocol and FASP (filter aided sample preparation) approach were tested. Furthermore a kit utilizing immobilized trypsin for proteolysis with subsequent desalting was tested (Capturem Kit). For a comparison of the different tested sample preparation protocols, the corresponding protein identification rates were plotted in Figure 19.

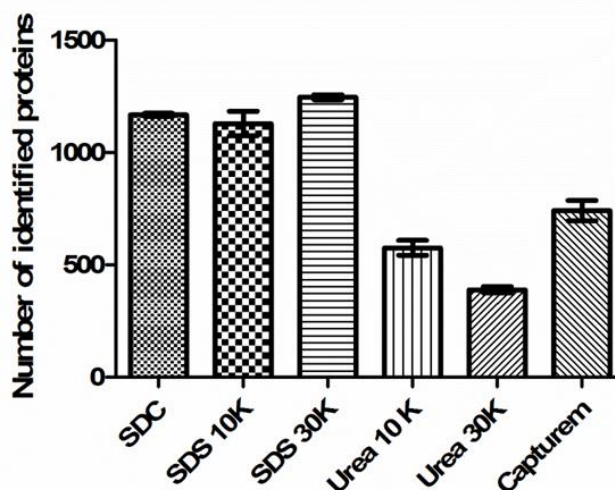


Figure 19: Bar plot comparing the number of identified protein using different sample preparation protocols for truffle powder. Protein had to be identified with at least 2 unique peptides. Error bars indicate standard deviation of technical triplicates.

1,169 (± 6) proteins were identified using the SDC protocol. By the FASP approach for preparing SDS extracted samples with 10 kDa cutoff filters 1,129 (± 45) proteins were identified, with a 30 kDa cutoff filter 1,247 (± 9) proteins. Urea extracted samples processed by FASP led to the identification of 576 (± 27) proteins with 10 kDa cutoff filters and 389 (± 11) proteins with 30 kDa cutoff filters. By the Capturem™ sample preparation kit 742 (± 37) proteins were identified. Therefore the FASP approach using SDS extracted samples and 30 kDa cutoff filters worked best for preparation of truffle powder. It had the highest and most repeatable protein identification rate. This was closely followed by the SDC protocol and the FASP approach with 10 kDa cutoff filters for SDS extracted proteins.

All urea extracted samples were checked for possible carbamylation of proteins. For these samples, the database search was repeated with inclusion of peptide carbamylation as variable modification. But only around 1 % of identified proteins were carbamylated. Therefore, it was excluded as cause for a low identification rate in urea extracted samples.

Due to their low protein identification rates, urea extracted samples processed with the FASP approach (both 10 kDa and 30 kDa cutoff filters) and processed with the Capturem™ sample preparation kit were excluded from further evaluation.

After comparing protein identification rates, a quantitative comparison of the remaining sample processing protocols was done. Only proteins identified and quantified over all samples prepared with the SDC protocol and the FASP approach for SDS extracted samples with both 10 kDa and 30 kDa cutoff filters were included. 922 proteins remained after

filtering. Coefficients of variation (CVs) within the different sample preparation methods were calculated for the abundance of these proteins. Protein abundance after preparation with the SDC protocol had a mean CV of 15.9 % and median CV of 12.1 %. With the SDS extracted samples and FASP approach using 10 kDa cutoff filters the CVs for protein abundance had a mean of 27.9 % and median of 23.8 %. Using the 30 kDa cutoff filter the mean CV was 16.3 % and the median CV 13.3 %.

Combining a high and repeatable identification rate with highest quantitative repeatability the SDC protocol was chosen for further sample preparation in following experiments.

6.3.2 Simplification of sample preparation

With a large number of samples possible simplification of the sample preparation process should be tested. Best target for adjustments would be the BCA test, done for the estimation of protein amount in given sample. A weight-volume ratio-based sample preparation protocol was compared with the original tryptic proteolysis approach based on the BCA assay. Five different truffle samples were used. One sample for each *T. indicum*, *T. sinense* and *T. magnatum* and two for *T. aestivum* with different geographical origins (Romania and Hungary) were used. Extraction buffer was added in a defined ratio to the amount weighed in. Then a BCA assay was performed and the result is shown in Figure 20.

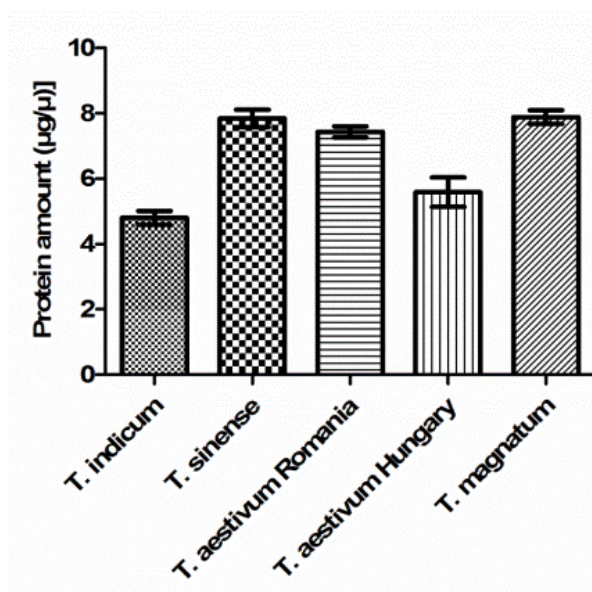


Figure 20: Bar plot showing the protein amount extracted with SDC buffer from different truffle species and origins. Error bars indicate standard deviation of technical triplicates.

Extraction for the different truffle species yielded in different protein amounts. For *T. indicum* 4.8 µg/µl (±0.3) were extracted, 7.8 µg/µl (±0.4) for *T. sinense* 7.8 µg/µl (±0.4) and for

T. magnatum 7.9 $\mu\text{g}/\mu\text{l}$ (± 0.3). Extraction for the two samples of *T. aestivum* with different geographical origin resulted in 7.4 $\mu\text{g}/\mu\text{l}$ (± 0.2) for *T. aestivum* from Romania and 5.6 $\mu\text{g}/\mu\text{l}$ (± 0.6) for *T. aestivum* from Hungary. With 39 % difference in amount of extracted protein *T. indicum* and *T. magnatum* were most different. Next, the average value of yielded protein amount in $\mu\text{g}/\mu\text{l}$ over all samples was calculated. According to this average protein amount value, an aliquot of each sample with the same volume was taken, corresponding to 20 μg of protein. For comparison, a second aliquot of each sample corresponding to 20 μg by the individual results of each sample from the performed BCA assay was taken.

All samples were processed by the SDC protocol and measured on a LC-MS/MS system. After data processing, quantitative values for identified proteins were used for hierarchical clustering analysis. Resulting heat map and dendrogram are plotted in Figure 21.

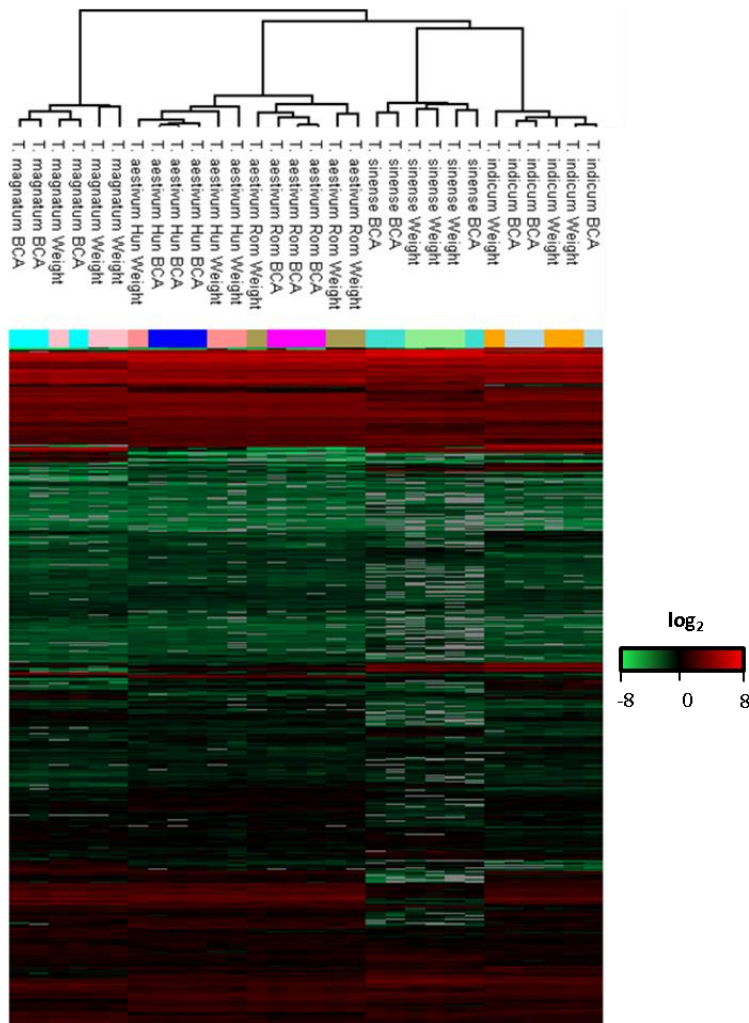


Figure 21: Heat map displaying protein abundance in different truffle species after hierarchical clustering. The heat map was generated by Perseus and displays \log_2 protein areas for 620 proteins identified in at least one of the three technical replicates for all groups. Each column in the heat map represents a different sample. Each line represents a protein. Red lines are high abundant proteins, green lines low abundant proteins. Grey lines represent missing values.

After hierarchical clustering analysis samples within different truffle species are most similar and clustering together in the dendrogram. All species are separated from each other. Within the species, a further clustering according to the chosen preparation approach is not visible. Samples prepared by both protein amounts according to the BCA assay and according to the adjusted weight-volume ratio are not separated. Therefore a quantitative difference between the different preparations approaches is not indicated while separation by biological identity of samples is still given.

To further test for a possible difference between tryptic proteolysis according to weight-volume ratio and by BCA test, results were filtered further. Only proteins identified with quantitative values in all samples were taken into account. This led to 373 proteins remaining

after filtering. For each sample, mean protein intensity values for these 373 proteins were calculated for both preparations by weight-volume ratio and by BCA test. Then these mean protein intensity values for samples processed by weight-volume ratio were plotted against corresponding mean protein intensity values for samples processed by BCA assay. Results are shown as multi scatter plot with Pearson correlation values in Figure 22.

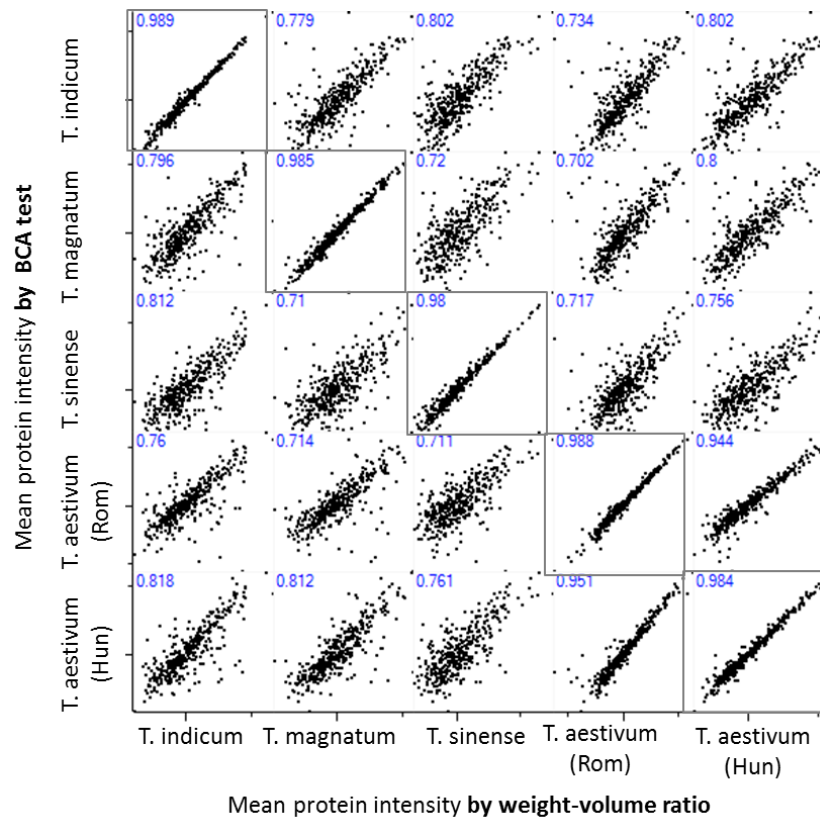


Figure 22: Multi scatter plot comparing the protein intensities obtained by sample preparation according to weight-volume ratio with sample preparation according to BCA test. Each dot represents a protein. The mean protein intensity quantified by sample preparation according to the BCA test is plotted against the mean protein intensity quantified by sample preparation by weight-volume ratio for the different truffle samples. Plotted are the 373 proteins identified in all samples. The depicted number in each of the individual scatter plots is the Pearson correlation value.

Resulting multi-scatter plot shows that related samples of the same species are most similar, regardless of the approach for sample preparation. All r square values comparing related samples processed by weight-volume ratio to BCA assay results are above values of 0.95. It also shows that both biological samples of *T. aestivum* with different geographical origins are very similar (Pearson correlation values of 0.95). Similarity between all other biological samples is lower, regardless of the preparation method chosen. These differences can be attributed to biological differences. Therefore it can be concluded, that protein extraction and

processing according to the BCA test as well as extraction and processing by weight with an adjusted weight-volume ratio are suitable ways for preparation of truffle samples.

After testing whether the BCA test was a necessary step for further sample processing, the focus was placed on time required for enzymatic proteolysis and trypsin-protein ratio. First, tryptic proteolysis for a time of 1 h was compared to overnight proteolysis. Additionally, trypsin-protein ratios of 1:100 and 1:25 were tested for both these durations. Generated tryptic peptides were measured on a LC-MS/MS system and protein identification rates are shown in Figure 23.

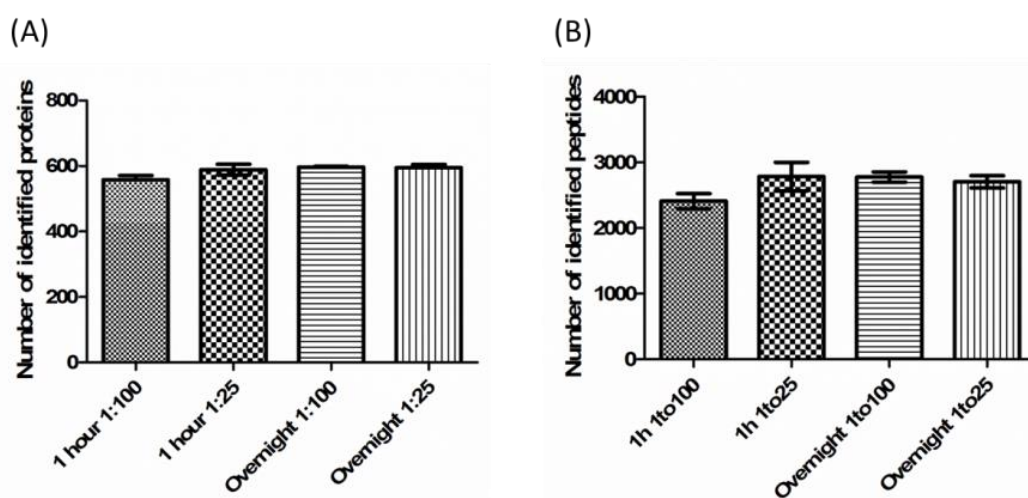


Figure 23: Box plots comparing identification rates using different digestion time and trypsin-to-protein ratios. Proteins were either digested for one hour or overnight with a trypsin-protein ratio of 1:100 or 1:25. Columns in (A) show the number of protein groups identified with at least 2 unique peptides, columns in (B) the number of identified peptides. Error bars indicate standard deviation of technical triplicates.

Proteolysis for 1 h with a trypsin-protein ratio of 1:100 proteins yielded in 2,411 (± 94) identified peptides from 558 (± 10) proteins. By 1 h proteolysis and 1:25 trypsin-protein ratio 2,785 (± 177) peptides from 589 (± 13) proteins were identified. Overnight proteolysis with 1:100 trypsin-protein ratio resulted in 2,777 (± 64) peptides and 597 (± 2) proteins being identified. Last, overnight proteolysis with 1:25 trypsin-protein ratio led to identification of 2,705 (± 78) peptides and 595 (± 8) proteins. Therefore, the highest and most reproducible number of proteins was identified with an overnight digestion and 1:100 trypsin-protein ratio.

Next, rates of tryptic missed cleavages within the generated peptides were compared for the different protocols. Observed missed cleavage rates are displayed in Figure 24.

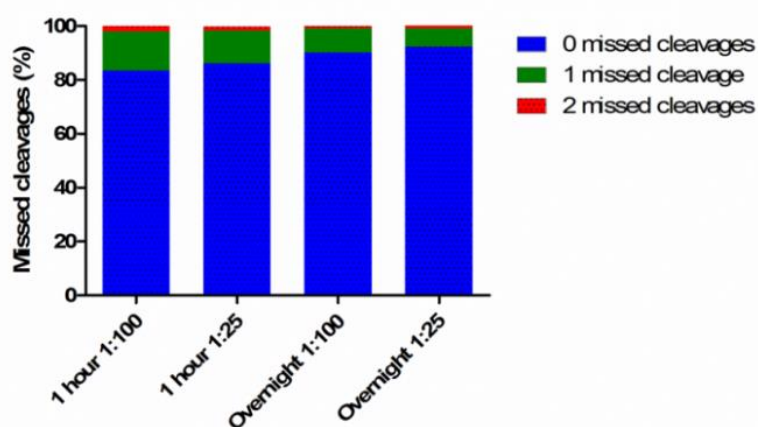


Figure 24: Grouped box plots comparing the rate of missed cleavages using different digestion times and enzyme-to-protein ratios.

Enzymatic proteolysis over 1 h with a trypsin-protein ratio of 1:100 generated peptides from which 83.5 % had no missed cleavages, 14.5 % with 1 missed cleavage and 2.0 % with 2 missed cleavages. An increase to a 1:25 trypsin-protein ratio and 1 hour proteolysis resulted in 86.2 % of peptides generated without missed cleavages, 12.2 % with 1 missed cleavage and 1.4 % with 2 missed cleavages. For overnight proteolysis with 1:100 trypsin-protein ratio, 90.2 % of peptides showed no missed cleavages, 9.0 % 1 missed cleavage and 0.8 % 2 missed cleavages. Lastly tested, overnight proteolysis with a 1:25 trypsin-protein ratio generated 92.3% of peptides without missed cleavages, 6.9 % with 1 missed cleavage and 0.8 % with 2 missed cleavages.

6.3.3 Data-dependent acquisition (DDA) approach for profiling of truffles

After testing and simplification of a suitable sample preparation, deep profiling of truffles was performed. Here, a bottom-up proteomics approach was chosen to generate a protein abundance fingerprint for different truffle species. By this protein abundance fingerprint differentiation of truffles should be enabled. Proteins from 40 truffle samples of the species *Tuber aestivum* (16 samples), *Tuber magnatum* (13 samples), *Tuber melanosporum* (7 samples) and *Tuber uncinatum* (4 samples) were prepared by the SDC protocol. Generated peptides were subsequently measured on a LC-MS/MS system. Bioinformatics processing yielded protein profiles including their relative quantities in analyzed different species.

Abundance values for 222 proteins identified in all samples were used to perform a principal component analysis (PCA). Sample projection of the first two principal components is shown in Figure 25.

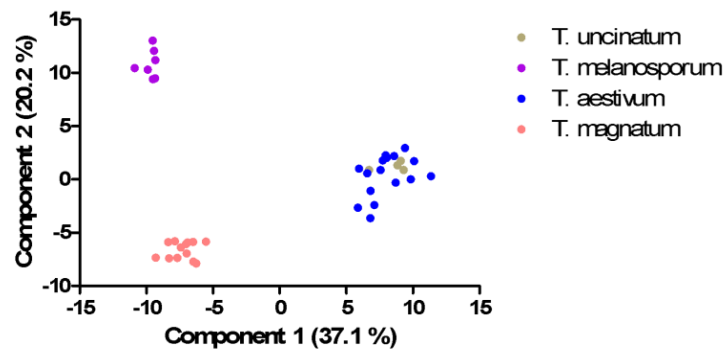


Figure 25: Sample projection of the first two components based on principal component analysis (PCA) for different truffle species. 222 proteins identified in all samples were used for the PCA.

Sample projection of the first two principal components showed that all *T. melanosporum* samples are highly similar. Same was true for all samples of *T. magnatum*. Both *T. melanosporum* and *T. magnatum* are clearly separated from each other but also from the *T. aestivum* and *T. uncinatum* samples. *T. aestivum* and *T. uncinatum* could not be separated from each other. There was a long and still ongoing debate whether *T. aestivum* and *T. uncinatum* are different species or not. Morphologic and phylogenetic analysis resulted in the conclusion they are conspecific^{72,73}. This is conclusive with the result obtained, also showing a co-specific relation between *T. aestivum* and *T. uncinatum* on proteomic level. Therefore *T. aestivum* and *T. uncinatum* will be used as synonyms, samples summarized as one species and further on addressed under the name *T. aestivum*.

To generate protein abundance profiles of truffle species, an analysis of variance (Anova) test was performed. As mentioned, *T. aestivum* and *T. uncinatum* samples were summarized under *T. aestivum* because of their co-specificity. By this approach profiles containing quantitative values for 166 Anova significant proteins were generated and displayed in a heat map after hierarchical clustering analysis in Figure 26.

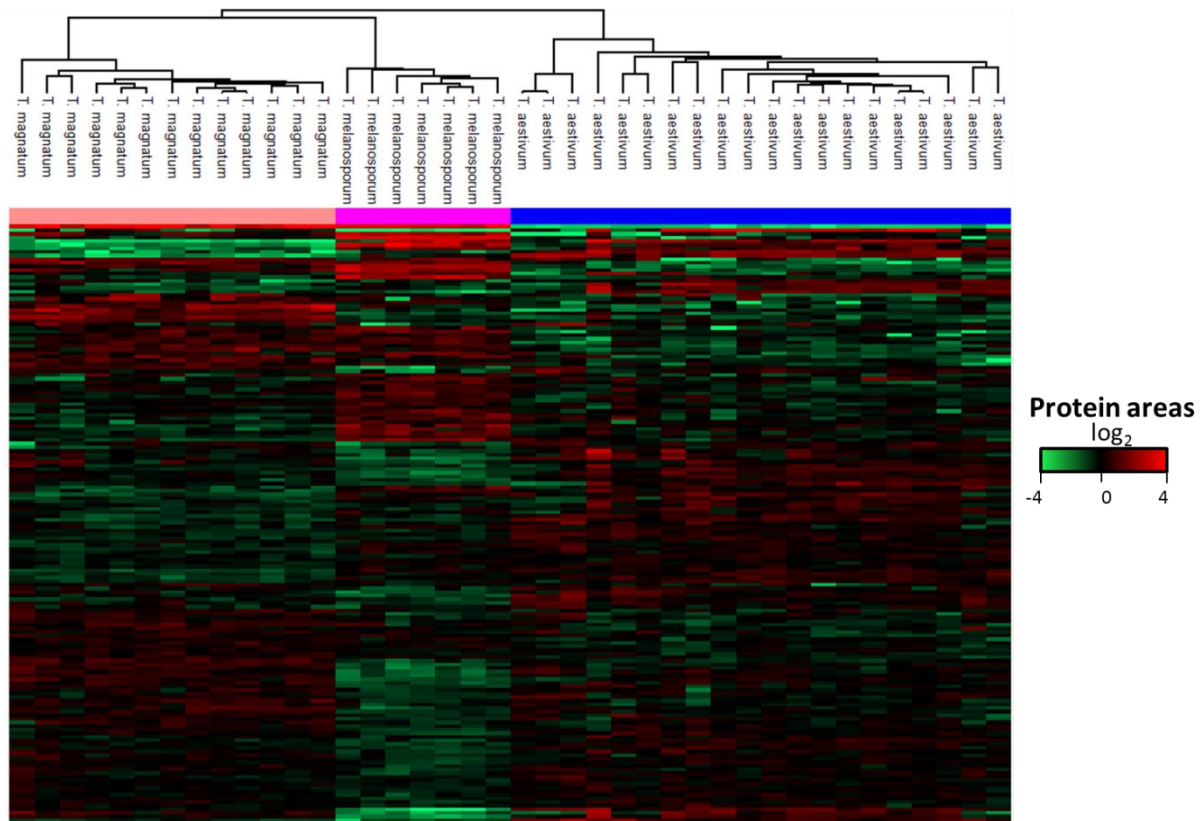


Figure 26: Heat map using Anova test with 5 % FDR displaying significant protein abundance in different truffle species after hierarchical clustering. The heat map was generated by Perseus and displays \log_2 protein areas for the 161 Anova significant proteins identified in all samples. Each column in the heat map represents a different sample. Each line represents a protein. Red lines are high abundant proteins, green lines low abundant proteins.

Taking up the results of the PCA analysis and its projection in Figure 25, the dendrogram after hierarchical clustering analysis of Anova significant proteins also showed a good separation of different species. All samples within a species were highly similar while different species were separated.

Additional to the fingerprint by Anova significant proteins, 222 proteins with quantitative values in all samples were compared species against species using a Student's T-Test. Tests were performed with a 5 % false discovery rate (FDR). First comparison of protein abundance levels was between *T. aestivum* and *T. magnatum*. 120 proteins were significantly different in their abundance. 38 proteins had an at least two-fold change in abundance, 23 were higher abundant in *T. aestivum*, 15 in *T. magnatum*. Then *T. aestivum* and *T. melanosporum* were compared. 123 proteins were significant different in their abundance. 74 of them had an at least twofold change, 41 were higher abundant in *T. aestivum*, the other 33 in *T. melanosporum*. Last *T. aestivum* and *T. melanosporum* were compared. 144 proteins were significant different in their abundance. 64 of them had an at least twofold change, 35 were

higher abundant in *T. magnatum* and 29 in *T. melanosporum*. As this results lacked in required depth, a biological interpretation was not purposeful.

6.3.4 Data-independent acquisition (DIA) approach for profiling of truffles

Results generated by DDA lacked wanted depth in analysis. While obtained quantitative protein fingerprint already enabled identification of truffles, the number of identified and quantified proteins was low. To overcome the problems of DDA and reach another level of depth, data-independent acquisition mode (DIA) measurements were performed. With DIA the analysis should improve highly in terms of identified and quantified proteins for a more thorough characterization of the different truffle species.

6.3.5 Building the spectral library

Measurements in data-independent acquisition mode require a spectral library for identification and quantification of proteins. For building a truffle spectral library, offline reversed phase high pH fractionation was done for samples of *T. aestivum*, *T. magnatum* and *T. melanosporum*. For each of the three species, 5 samples were pooled, separated on a HPLC and obtained fractions pooled in a concatenated way, resulting in 13 fractions. These 13 fractions for each species were measured on a LC-MS/MS system and searched against a corresponding protein database. The generated result file for *T. aestivum* contained 3,095 proteins and 18,234 peptides, for *T. magnatum* 3,544 proteins and 24,564 peptides and for *T. melanosporum* 3,519 proteins with 26,686 peptides. Then, result files were combined. Finally, the combined result file contained 9,170 proteins and 51,628 peptides from three different truffle species and was used as spectral library.

6.3.6 Profiling truffles regarding the difference between species

DIA measurements were done for 72 samples (28 *T. aestivum*, 4 *T. Albidum Pico*, 10 *T. indicum*, 19 *T. magnatum*, 11 *T. melanosporum*).

Data quality for reliable quantification was controlled by checking points across the chromatographic peaks for monitored fragment ions. For this, the median points across chromatographic peaks for each individual sample were calculated. From these values the average value over all samples was calculated. With an average of 11 points across chromatographic peaks a reliable quantification was enabled.

First, abundance values for 2,715 proteins identified in all samples were used to perform a principal component analysis (PCA). Sample projection of the first two principal components is shown in Figure 27.

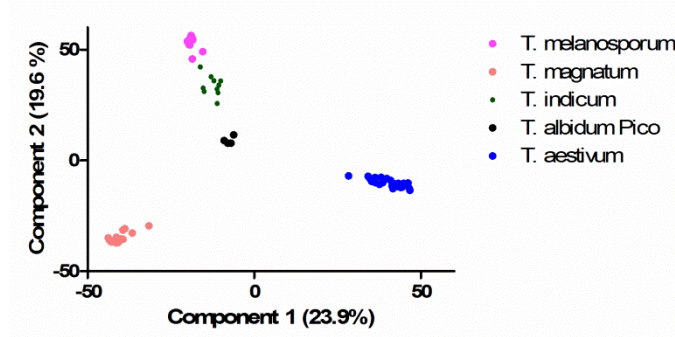


Figure 27: Sample projection of the first two components based on principal component analysis (PCA) for different truffle species. 2,715 proteins identified in all samples were used for the PCA.

Projection of the first two principal components showed samples for each *T. melanosporum*, *T. magnatum*, *T. indicum*, *T. albidum Pico* and *T. aestivum* were highly similar and clustering together. This is conclusive similar with results from DDA mode analysis in Figure 25. Both *T. magnatum* and *T. aestivum* were clearly separated from each other and the other three species. *T. melanosporum*, *T. indicum* and *T. albidum Pico* were still separated from each other, but appeared closer together in the projection. They didn't differ strongly within the first principal component, but were separated by the second principal component. Closest in the projection were *T. melanosporum* and *T. indicum*. Separation in the projection for both these species was done only by the second principal component.

After analysis and plotting of samples prior to statistical analysis, an Anova-test was performed to find significant differences in protein abundance values between truffle species. From 2,715 proteins identified in all samples a total of 2,066 were Anova significant. Depicted in Figure 28 is the obtained heat map displaying Anova significant proteins after hierarchical clustering analysis.

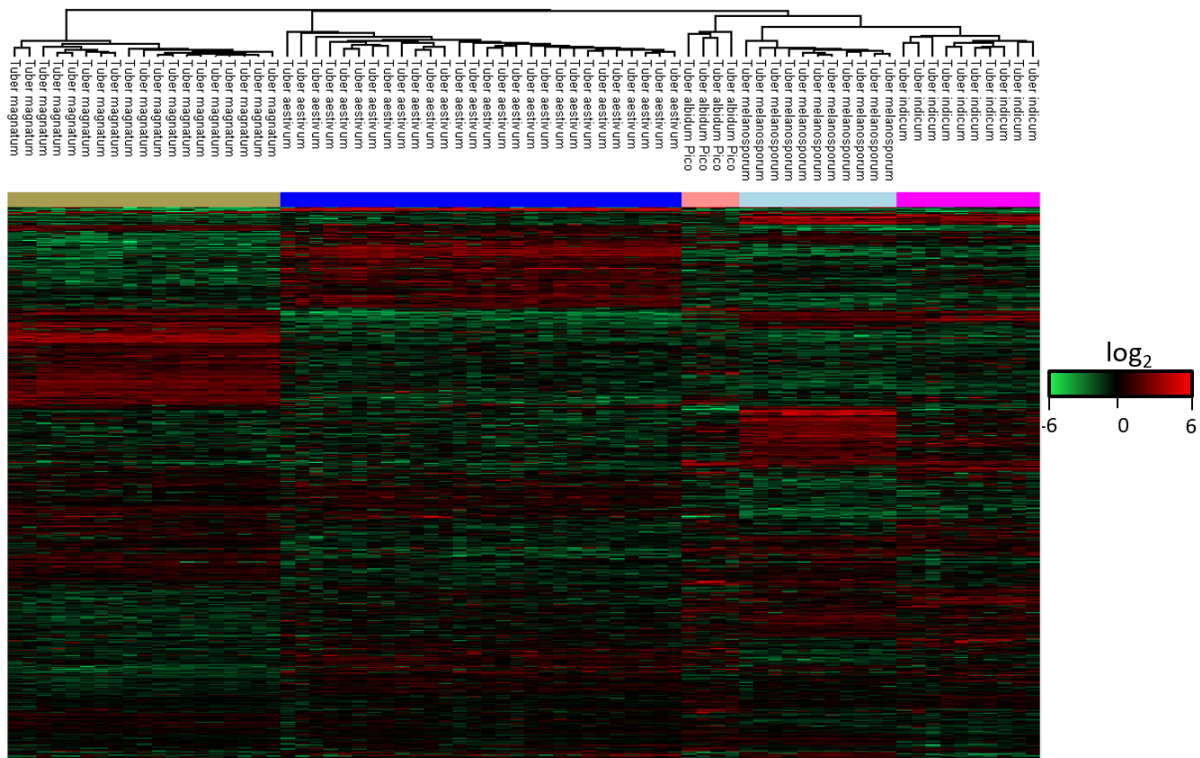


Figure 28: Heat map using Anova test with 5 % FDR displaying significant protein abundance in different truffle species after hierarchical clustering. The heat map was generated by Perseus and displays \log_2 protein areas for the 2,066 Anova significant proteins identified in all samples. Each column in the heat map represents a different sample. Each line represents a protein. Red lines are high abundant proteins, green lines low abundant proteins.

After hierarchical clustering analysis all samples within the same species were highly similar and clustered together in the dendrogram, while different species were separated. This was expected, as a separation was already seen prior to statistical analysis and filtering by sample projections after principal component analysis. Similar to Figure 27 *T. melanosporum* and *T. indicum* clustered together most closely and with a small distance to *T. albidum Pico*. Nevertheless, all three species were still separated. *T. aestivum* and *T. magnatum* were most distant from both each other as well as from the other species, which is conclusive with previous results.

With establishment of a method to differentiate truffles, a more comprehensive characterization was the next step. After Anova testing to generate truffle species specific barcodes and showing global differences, differences between species should be further worked out in more detail. Therefore two-sample t-tests between species with a 1 % FDR were performed. Significant regulated proteins were filtered for an at least two-fold change. Resulting numbers of proteins from these species-to-species comparisons are listed in Table 19.

Table 19: Results of the species-to-species comparison listing the number of T-test significant proteins upregulated with an at least two-fold change. Each species in this table is assigned with a color. For each species-against-species comparison, the number of upregulated proteins is listed. Therefore, two numbers are given for each comparison. The number of proteins upregulated in the corresponding species is indicated by the same color-coding.

| Species-to-species comparison | Number of significantly upregulated proteins with an at least two-fold change | | | | | | | | |
|-------------------------------|---|----------------|-----|--------------------|-----|----------------|-----|--------------------|-----|
| | Tuber indicum | Tuber magnatum | | Tuber melanosporum | | Tuber aestivum | | Tuber albidum Pico | |
| Tuber indicum | | 389 | 483 | 237 | 334 | 372 | 412 | 119 | 113 |
| Tuber magnatum | | | | 589 | 585 | 541 | 503 | 337 | 277 |
| Tuber melanosporum | | | | | | 489 | 475 | 349 | 288 |
| Tuber aestivum | | | | | | | | 232 | 189 |
| Tuber albidum Pico | | | | | | | | | |

For better insight in differences between species, the 2,715 proteins identified in all samples were used to perform species-against-species comparisons by Student's T-test. Strongest differences in abundance of proteins were observed between *T. melanosporum* and *T. magnatum*. In total 1,174 proteins were significantly different in abundance with an at least two-fold change. These total of 1,174 proteins consisted of 589 proteins which were upregulated in *T. magnatum* and 585 proteins upregulated in *T. melanosporum*. Using the same criteria, the second strongest differences were between *T. magnatum* and *T. aestivum* with 1,044 proteins and *T. melanosporum* compared to *T. aestivum* with 964 proteins. Smallest observed differences were between *T. indicum* and *T. albidum Pico* with only 232 proteins significant different in abundance and an at least two-fold change.

6.3.7 Introduction of a smaller protein panel for distinction of truffles

Next a smaller barcode in form of a protein panel should be generated. This protein panel can be used to assign unknown samples of these species. For generating this panel a reference was needed. This reference consisted of 5 proteins with an unchanged abundance over all the samples. Proteins were selected by the projection of the principal component analysis (Protein 1: A0A292Q353_9PEZI, Protein 2: A0A292Q9K6_9PEZI, Protein 3: A0A317SEE9_9PEZI, Protein 4: D5G4I5_TUBMM and Protein 5: A0A292Q6G0_9PEZI). These 5 proteins didn't contribute to the separation of species in the PCA projection. To additionally check for an unchanged abundance, intensity values of these 5 proteins were plotted for all individual samples. This plot is depicted in Figure 29.

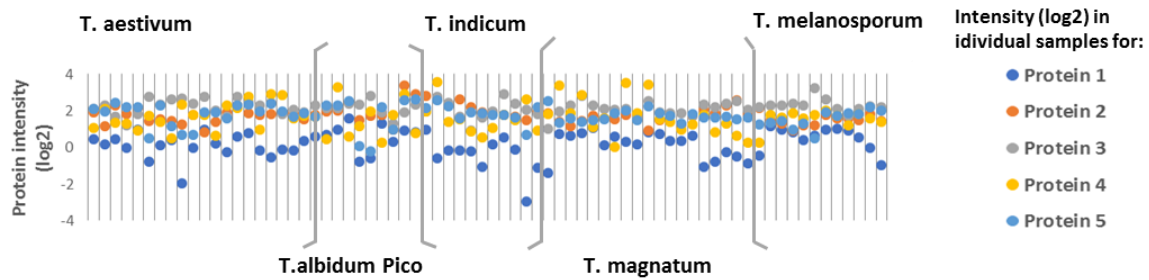


Figure 29: Profile plot showing the intensity of 5 proteins with small changes in abundance over the different truffle species. Each section in the graph corresponds to one sample with dots for the 5 proteins of unchanged abundance inside. Samples are ordered and grouped by grey bars into the different species.

After controlling for unchanged abundance of the 5 chosen proteins, a reference value was created. For the reference value, the mean abundance value (from the log₂ intensity values normalized by median) over all samples for each of the 5 reference proteins was calculated. These 5 mean abundance values were then again averaged into one final reference value.

After establishment of this reference, the protein barcode was generated. To generate the barcode, 15 proteins were selected. These proteins needed to be carefully selected. Figure 30 shows exemplary how the proteins were chosen.

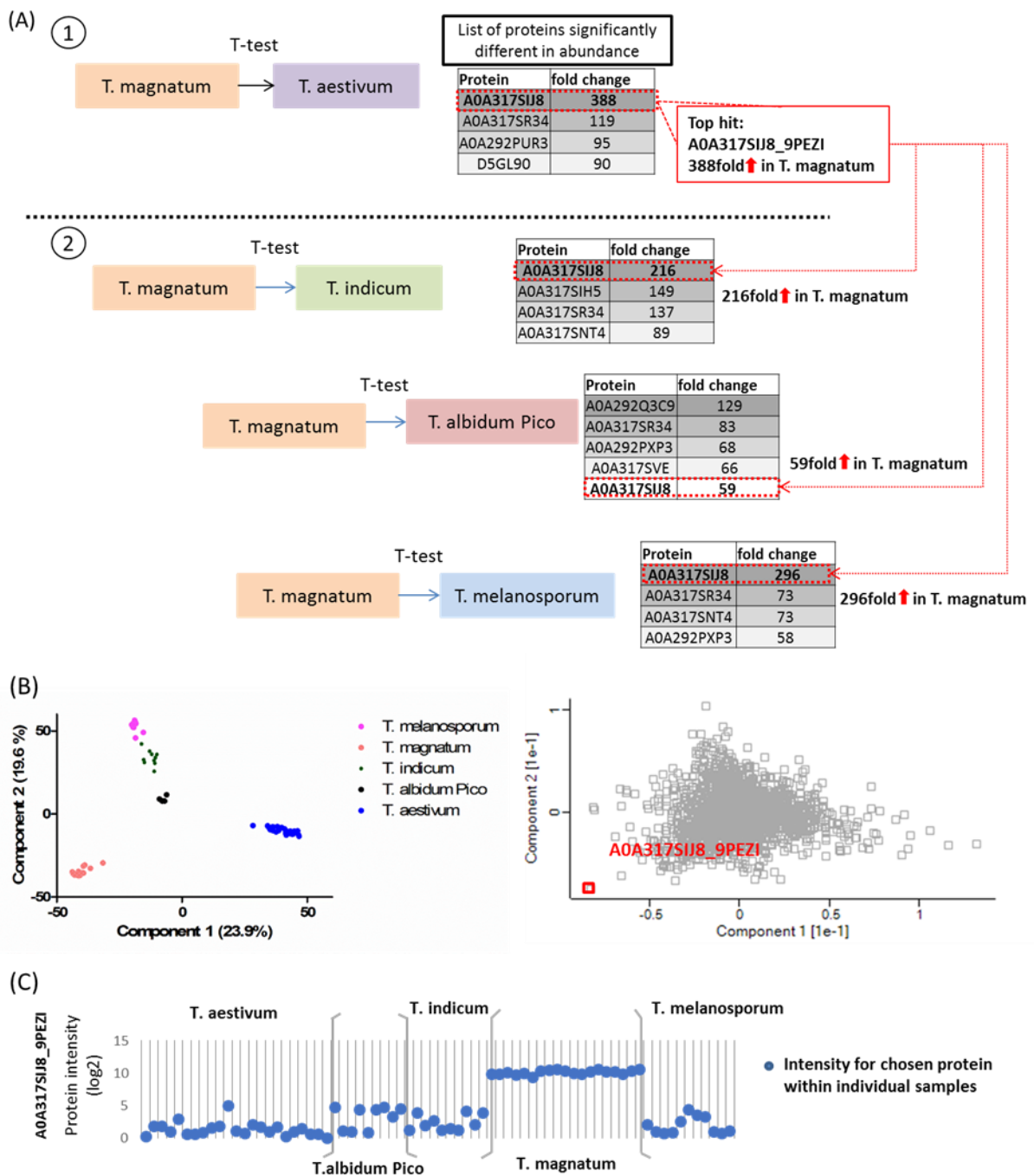


Figure 30: Overview of the process to choose proteins for the protein panel on the example of A0A317SIJ8_9PEZI. (A) First a T-test to compare protein abundances within *T. magnatum* and *T. aestivum* was done. From the list of proteins with significantly different abundance the Top hit with a 288fold higher abundance in *T. magnatum* was chosen. This candidate was searched for in the protein lists after T-test comparisons of *T. magnatum* with all other species. In all of the lists it showed higher abundance as well. (B) Sample projection of the first two principal components for different truffle species on the left side was compared with the protein loading plot shown on the right. The protein loading plot shows the proteins driving separation of species in the sample projection on the left. Chosen candidate protein for the protein panel is marked red in the loading plot and shows high contribution to separation for *T. magnatum*. (C) Lastly the proteins intensity was checked over all samples. This was done by plotting the protein's intensity in all samples ordered and grouped by their species, indicated by the grey bars. An overall higher intensity for the protein in *T. magnatum* only was observed. Therefore the protein A0A317SIJ8_9PEZI is fitting to be included in the panel.

First, obtained lists of proteins with significantly different abundance observed by T-tests between different species were investigated. Top hits of proteins with the highest fold-change in abundance were selected for further assessment. These top hits from one comparison were searched in the other T-test deriving lists of significantly different proteins. When a protein was consistently presented as higher abundant for one species in comparison with all other species, it was further assessed. Next, the assessed protein was controlled in the loading plot for the principal component analysis of the different species. When it contributed highly to the separation of one species, it was further included. Last step was the plotting of this proteins intensity within all individual samples. If there was a consistently higher abundance for this protein in one or two species only, it was chosen to be part of the protein panel. After selection, the mean abundance values (from the log₂ intensity values normalized by median) for these proteins were calculated over all samples of one species. For normalization, the previously established reference value (value of 1.46) of reference proteins with unchanged abundance was subtracted from each of the selected proteins in the panel. A list of normalized signal intensity values for the 15 chosen proteins is listed in Table 20.

Table 20: Protein Panel for the differentiation of truffle species. The panel consists of 15 proteins with normalized intensity values and their standard deviation.

| | Protein identifier | Tuber aestivum | Tuber albidum Pico | Tuber indicum | Tuber magnatum | Tuber melanosporum |
|-------------------|---------------------------|-----------------------|---------------------------|----------------------|-----------------------|---------------------------|
| Protein 1 | D5G725 _TUBMM | 0.2 (±0.5) | 2.7 (±1.0) | 0.6 (±0.5) | 0.2 (±0.6) | 7.6 (±0.4) |
| Protein 2 | A0A317SIM4 _9PEZI | -4.2 (±1.1) | -4.2 (±0.7) | 0.6 (±0.9) | 1.4 (±0.4) | -4.1 (±0.6) |
| Protein 3 | D5G6W9 _TUBMM | -2.9 (±1.1) | -4.1 (±1.2) | 4.4 (±0.2) | -3.9 (±1.4) | 4.0 (±0.3) |
| Protein 4 | A0A317SIJ8 _9PEZI | 0.1 (±1.3) | 2.8 (±0.6) | 0.9 (±1.1) | 8.7 (±0.3) | 0.4 (±1.3) |
| Protein 5 | A0A292PIL3 _9PEZI | 8.4 (±0.6) | 5.8 (±4.9) | 0.0 (±2.0) | -1.7 (±0.9) | -1.7 (±1.3) |
| Protein 6 | A0A292PQL2 _9PEZI | 6.0 (±0.6) | -2.7 (±1.0) | -2.5 (±0.9) | -2.7 (±1.1) | -2.4 (±0.9) |
| Protein 7 | D5GM64 _TUBMM | -6.8 (±1.5) | 0.9 (±1.8) | -3.8 (±1.4) | -6.8 (±1.9) | -1.3 (±0.2) |
| Protein 8 | A0A292Q3C9 _9PEZI | -3.4 (±1.7) | -4.7 (±0.4) | -4.1 (± 1.8) | -2.1 (±0.4) | 0.9 (±0.3) |
| Protein 9 | A0A317SR34 _9PEZI | -3.8 (±1.3) | -3.3 (±1.0) | -4.0 (±1.0) | 3.1 (±0.4) | -3.1 (±1.0) |
| Protein 10 | D5G506 _TUBMM | -4.6 (±1.3) | -4.0 (±2.2) | -5.9 (±1.1) | -5.4 (±1.8) | -0.3 (±0.4) |
| Protein 11 | A0A292Q736 _9PEZI | -2.3 (±0.7) | -0.5 (±1.1) | 2.7 (±1.9) | -2.3 (±0.7) | -1.6 (±0.7) |
| Protein 12 | A0A317SSY1 _9PEZI | -3.4 (±0.9) | -4.7 (±0.3) | -4.1 (±1.2) | -2.1 (±1.1) | 0.9 (±0.4) |
| Protein 13 | A0A292PSE4 _9PEZI | 7.2 (±0.4) | 0.6 (±0.2) | 0.4 (±0.8) | 0.9 (±0.8) | -0.6 (±0.8) |
| Protein 14 | D5GKK3 _TUBMM | -3.7 (±1.3) | -3.8 (±1.5) | 2.4 (±0.5) | -3.5 (±1.1) | 0.3 (±0.3) |
| Protein 15 | D5GG88 _TUBMM | 0.1 (±0.8) | -0.9 (±0.3) | 0.6 (±0.8) | -0.6 (±0.9) | 5.0 (±0.3) |

Based on the normalized intensity values from the protein panel, profiles for the different truffle species were generated. For all species the lowest signal intensity value (Protein 7 in *T. magnatum*) was set as ground level (0 % signal intensity). Further, highest signal within each individual species was set as maximum level for signal intensity (100 % signal intensity). Profiles show the relative ratio of signal intensity for proteins within the different species and are displayed as bar plots in Figure 31.

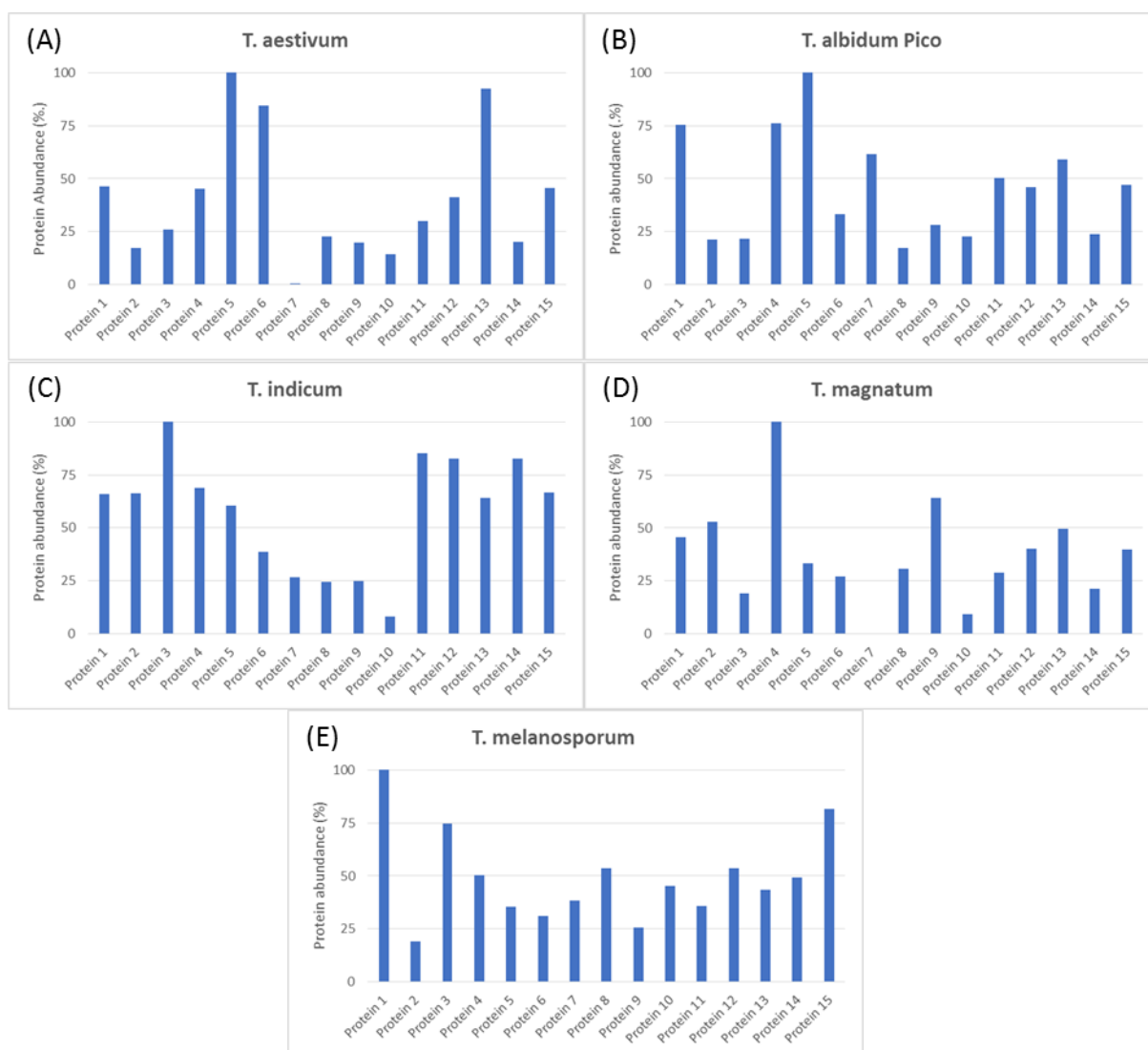


Figure 31: Bar plots displaying the protein abundance based profiles for the different truffle species. The bar plots display the protein abundance in percent of the highest signal for each truffle species. Plotted are (A) *T. aestivum*, (B) *T. albidum Pico*, (C) *T. indicum*, (D) *T. magnatum* and (E) *T. melanosporum*.

With the profiles it is possible to distinguish truffles by the ratio of protein abundances within a given sample. For example, in *T. aestivum* Protein 5 will show the strongest signal, with Protein 13 as second strongest and Protein 6 on third place. Additionally, protein 7 would be nearly absent. For *T. magnatum* Protein 4 has the highest abundance, followed by Protein 9 and 2. Similar to *T. aestivum* protein 7 is nearly absent too. Ratios of protein abundance are shown to be specific for each species. It is possible to distinguish different truffle species in a side by side comparison. For example Protein 5 shows the highest intensity in *T. aestivum* and *T. albidum Pico*, but they differ in Protein 7, which is nearly absent in *T. aestivum*. Protein 7 is also very low abundant in *T. magnatum*, but here the most prominent signal belongs to Protein 4 and so on. These comparisons can be done between all the species. To make them visually easier, a form of presentation similar to MALDI-TOF biotyping was chosen. Here,

two species were chosen for comparison and the lowest and highest values set as ground and maximum level, similar as for the profiles. Protein abundances, corresponding to signal intensities for the 15 proteins within the panel, were then bidirectional plotted in a grouped bar chart for comparison. A bidirectional plot comparing black truffles *T. melanosporum* and *T. indicum* is seen in Figure 32.

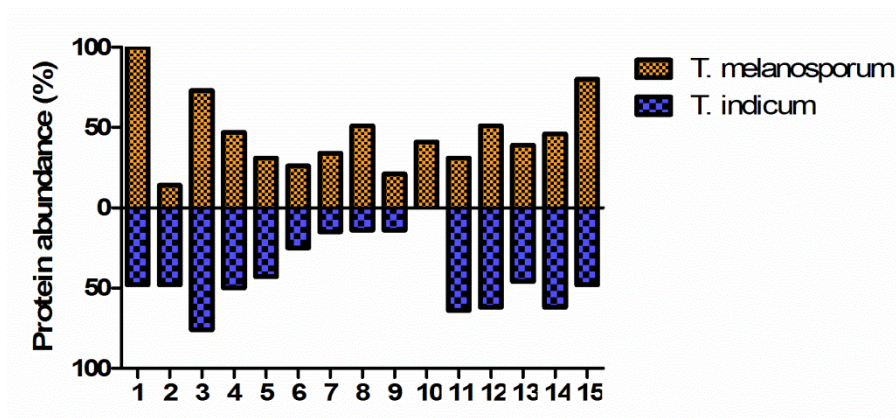


Figure 32: Bidirectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. melanosporum* with *T. indicum*.

By this bidirectional plotting differences between two species can be visualized easily. Comparing *T. melanosporum* and *T. indicum*, biggest differences are seen for protein 1 and protein 10. Both have a higher abundance in *T. melanosporum*. This continues with protein 7, 8, 9 and 15. Proteins 2 and 11 to 14 are more abundant in *T. indicum*. Therefore a quick distinction of both species is achievable with the help of the protein panel and bidirectional plotting of protein abundances. In a second example this approach was used to compare *T. magnatum* with *T. albidum Pico*. The resulting bidirectional bar plot is shown in Figure 33.

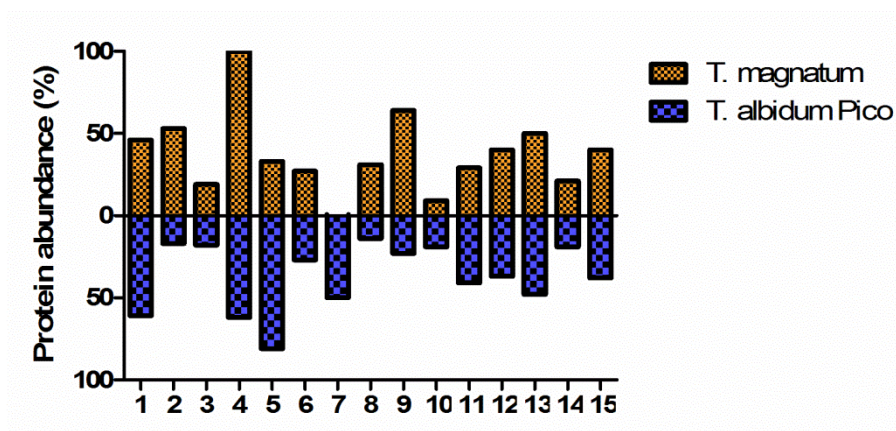


Figure 33: Bidirectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. magnatum* with *T. albidum Pico*.

In the comparison of *T. magnatum* with *T. albidum Pico* differences in protein abundance are quickly visible again. Protein 7 has a much higher abundance in *T. albidum Pico* and therefore a lower abundance in *T. magnatum*. Same behavior is seen for protein 5, with a high abundance in *T. albidum Pico*. In *T. magnatum* proteins 2, 4, 8 and 9 are of higher abundance. Rest of the proteins is around the same abundance level in both species. A bidirectional plotting approach therefore brings together both profiles and side-to-side comparison. It is possible to take the ratio for the signal intensity of given 15 proteins within each species into account, as well as to perform a side-to-side comparison of signal intensity for the proteins in two different species.

6.3.8 Gene ontology enrichment analysis

To work out underlying biological processes for seen differentiation of truffles, first Anova-significant proteins should be analyzed for functional annotation. Previously obtained list of Anova-significant proteins was subjected to a gene ontology enrichment analysis. From 2,066 proteins which were Anova-significant, 1,325 (64 %) had a functional annotation and could be matched to biological processes. A list of the ten most enriched biological processes within the Anova-significant proteins is listed in Table 21.

Table 21: Top 10 hits of enriched biological processes in Anova significant proteins from the comparison of all truffles.

| GO ID | GO Description | p-value | Corrected p-value | Cluster frequency | Total frequency |
|---------|--|----------|-------------------|--------------------|---------------------|
| 44281 | small molecule metabolic process | 5.66E-70 | 9.92E-67 | 368/1325 27.7% | 1195/10365 11.5% |
| 6082 | organic acid metabolic process | 1.27E-53 | 1.11E-50 | 243/1325 18.3% | 710/10365 6.8% |
| 19752 | carboxylic acid metabolic process | 6.71E-51 | 3.92E-48 | 232/1325 17.5% | 679/10365 6.5% |
| 43436 | oxoacid metabolic process | 4.08E-49 | 1.79E-46 | 232/1325 17.5% | 693/10365 6.6% |
| 55114 | oxidation-reduction process | 2.36E-40 | 8.26E-38 | 353/1325 26.6% | 1438/10365 13.8% |
| 6520 | cellular amino acid metabolic process | 2.08E-35 | 6.08E-33 | 164/1325 12.3% | 481/10365 4.6% |
| 1901605 | alpha-amino acid metabolic process | 8.26E-32 | 2.07E-29 | 114/1325 8.6% | 285/10365 2.7% |
| 8152 | metabolic process | 4.30E-30 | 9.42E-28 | 1100/1325 83.0% | 7285/10365 70.2% |
| 6091 | generation of precursor metabolites and energy | 2.46E-29 | 4.80E-27 | 97/1325 7.3% | 230/10365 2.2% |
| 1901564 | organonitrogen compound metabolic process | 8.41E-29 | 1.48E-26 | 604/1325 45.5% | 3314/10365 31.9% |

Strongest enrichment was seen for various metabolic processes and oxidation-reduction processes. After gene ontology enrichment analysis of Anova-significant proteins giving a general overview of underlying biological processes, a more in-depth approach was done. In contrast to Anova analysis, species-to-species comparisons resulted in lists of differentially regulated proteins. For gaining more detailed insight into functional categorization of these proteins, gene ontology enrichment analysis was performed again. Here, only up-regulated proteins for one species-to-species comparison were grouped by their annotated biological processes. In average, 63 % of MS identified proteins from the differentially regulated protein lists were annotated with biological processes. Most often occurring in the comparisons were enriched metabolic processes (small molecule metabolic process, organic acid metabolic process, carboxylic metabolic process, oxoacid metabolic process) and oxidation-reduction processes. Further organonitrogen compound metabolic processes showed enrichment in many comparisons. Exemplary shown in Table 22 are the 15 most enriched processes for proteins upregulated in *T. magnatum* in comparison to *T. indicum*.

Table 22: Top 15 hits of enriched biological processes in upregulated proteins from *T. magnatum* in comparison with *T. indicum*.

| GO ID | GO Description | p-value | corrected p-value | Cluster frequency | Total frequency |
|---------|--|----------|-------------------|-------------------|---------------------|
| 44281 | small molecule metabolic process | 4.71E-17 | 4.02E-14 | 93/331 28.0% | 1195/10365 11.5% |
| 55114 | oxidation-reduction process | 9.55E-14 | 4.08E-11 | 97/331 29.3% | 1438/10365 13.8% |
| 8152 | metabolic process | 2.72E-13 | 7.73E-11 | 288/331 87.0% | 7285/10365 70.2% |
| 6082 | organic acid metabolic process | 5.80E-13 | 1.24E-10 | 61/331 18.4% | 710/10365 6.8% |
| 19752 | carboxylic acid metabolic process | 3.06E-12 | 5.22E-10 | 58/331 17.5% | 679/10365 6.5% |
| 96 | sulfur amino acid metabolic process | 3.72E-12 | 5.30E-10 | 17/331 5.1% | 61/10365 0.5% |
| 1901605 | alpha-amino acid metabolic process | 5.37E-12 | 6.55E-10 | 35/331 10.5% | 285/10365 2.7% |
| 43436 | oxoacid metabolic process | 7.12E-12 | 7.60E-10 | 58/331 17.5% | 693/10365 6.6% |
| 9069 | serine family amino acid metabolic process | 5.88E-11 | 5.58E-09 | 16/331 4.8% | 62/10365 0.5% |
| 6091 | generation of precursor metabolites and energy | 2.09E-10 | 1.78E-08 | 29/331 8.7% | 230/10365 2.2% |
| 6534 | cysteine metabolic process | 2.11E-09 | 1.63E-07 | 11/331 3.3% | 32/10365 0.3% |
| 6520 | cellular amino acid metabolic process | 2.29E-09 | 1.63E-07 | 42/331 12.6% | 481/10365 4.6% |
| 6790 | sulfur compound metabolic process | 1.16E-08 | 7.60E-07 | 25/331 7.5% | 209/10365 2.0% |
| 1901564 | organonitrogen compound metabolic process | 1.42E-08 | 8.66E-07 | 154/331 46.5% | 3314/10365 31.9% |
| 70813 | hydrogen sulfide metabolic process | 2.20E-08 | 1.17E-06 | 7/331 2.1% | 12/10365 0.1% |

Additional to previously mentioned metabolic processes, interestingly processes regarding sulfur-compounds (sulfur amino acid and cysteine metabolic processes, sulfur compound metabolic process and hydrogen sulfide metabolic process) were found to be enriched for *T. magnatum* compared to *T. indicum*. Compared with other truffles, one or more of this sulfur metabolic processes were enriched top hits for *T. magnatum*. Similar, processes regarding sulfur metabolism were found to be enriched top hits for *T. melanosporum* in comparison with the other truffles. Comparing *T. magnatum* with *T. melanosporum*, these sulfur-related processes didn't make it into the top hits on the list of enriched processes for both species.

6.3.9 Profiling truffles regarding their geographical origin

Truffle species could be distinguished well by their species. Next, differences regarding the geographical origin were investigated. This was done by comparing samples of different geographical origin within the same truffle species. Samples of *T. aestivum*, *T. magnatum* and *T. melanosporum*, were used. For *T. aestivum* 23 samples (2x Bulgaria, 1x France, 1x Iran, 6x Italy, 1x Moldova, 11x Romania and 1x Slovenia), for *T. magnatum* 19 samples (2x Bulgaria, 13x Italy, 2x Kroatia and 2x Romania) and for *T. melanosporum* 11 samples (2x Australia, 2x France, 3x Italy and 4x Spain) were included in the analysis. Principal component analysis was performed and the sample projections of the first two principal components are plotted in Figure 34.

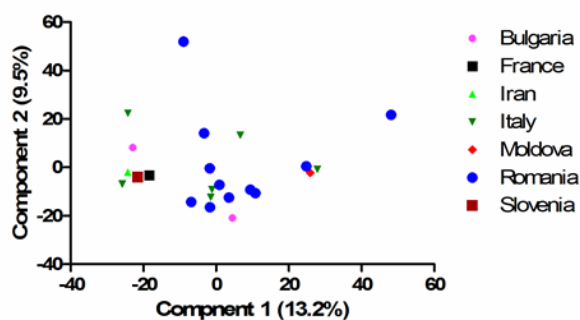
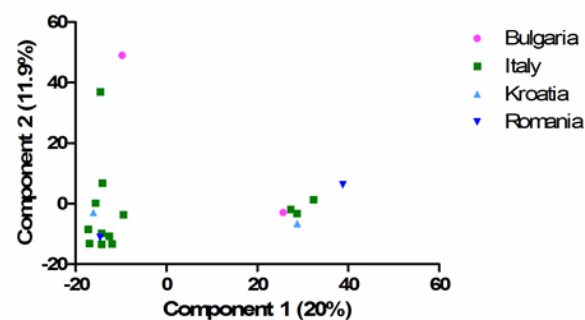
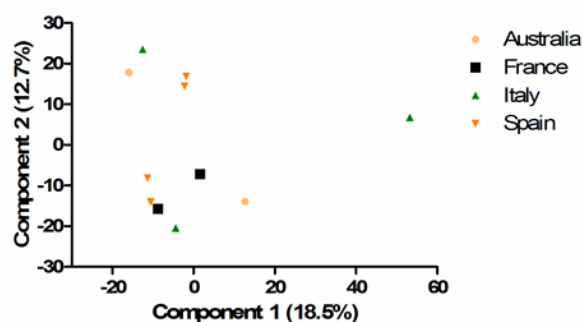
(A) *T. aestivum*(B) *T. magnatum*(C) *T. melanosporum*

Figure 34: Sample projection of the first two components based on principal component analysis for truffle samples of different geographical origin. Species used were (A) *T. aestivum* (B) *T. magnatum* and (C) *T. melanosporum*.

Sample projection of the first two principal components for different truffle species in Partition A to C showed no clustering of samples according to their geographical origin. For each species, samples of the same geographical origin were spread over the whole sample projection. Even for *T. melanosporum* with samples originating in Europe and Australia no

distinction by origin was possible. A distinction of truffle species directly using generated DIA data and protein abundance without utilization of more elaborate statistical analysis was therefore not enabled.

6.4 Summary of results

A final overview of generated results is given in Figure 35.

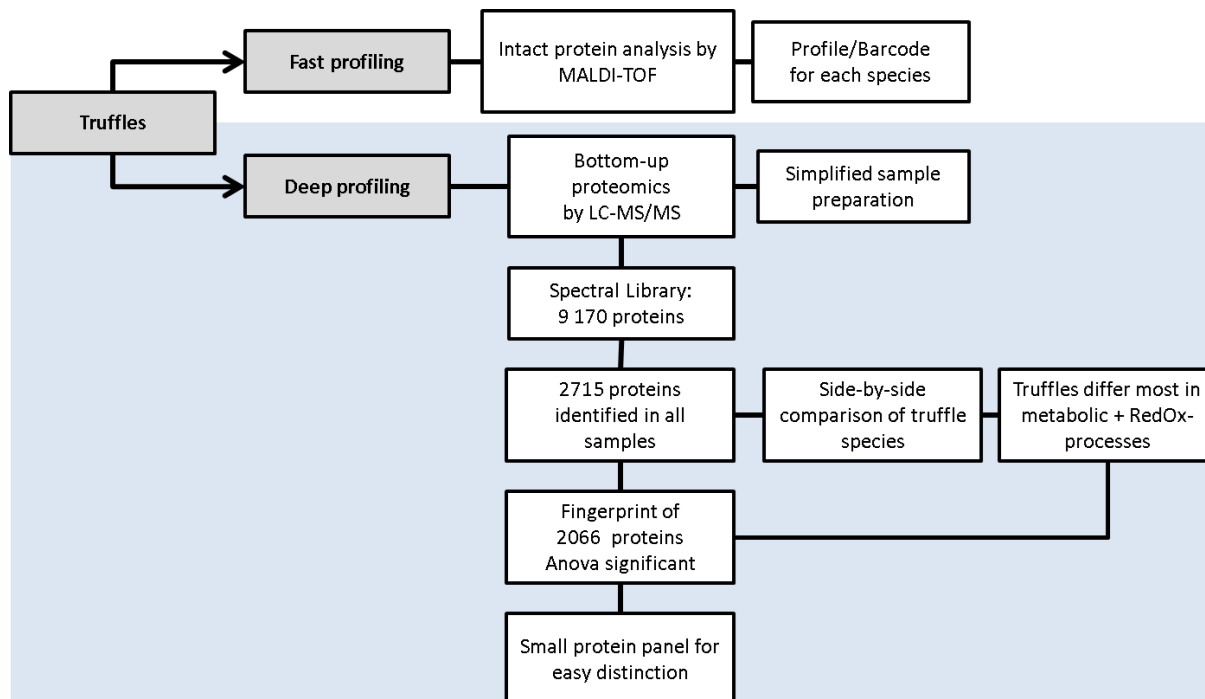


Figure 35: Schematic overview of generated results.

7. Discussion

With increasing globalization, food fraud becomes a topic of uprising interest. This is especially a case for expensive foods like truffles. Most established methods for food authentication rely on utilizing genomic markers. However, more recent proteomic approaches have become of more interest in authentication of biological identity, origin, quality and processing of foods.

First a fast way to profile truffle samples was developed. Here MALDI-TOF based analysis and identification of different truffles was successfully performed. For MALDI-TOF analysis, suitable preparation of truffle samples was needed. By application of formic acid-acetonitrile solution (70 % formic acid with equal volume acetonitrile) proteins were extracted from ground truffle powder and could be co-crystallized with alpha-Cyano-4-hydroxycinnamic acid in a thin layer procedure. Measurements are done on a MALDI-TOF system, recording spectra in the scan range of 2,000-20,000 m/z. This protocol for sample preparation is widely applied in microbiology. It's easy to perform, inexpensive and highly reproducible as it relies on the analysis of highly abundant proteins of bacteria, such as ribosomal proteins^{74,75}.

Using the formic acid-acetonitrile extraction protocol, distinct signals of sufficient intensity for a truffle sample were generated in MALDI-TOF (Figure 10). Established scan range of 2,000-20,000 m/z for measuring high abundant proteins in bacteria could be adopted for truffles as well. With an even and spacious crystallization over the whole target spot an automated scanning with the laser for ablation was enabled. Another preparation approach was tested, a combined extraction and crystallization solution containing 60 % acetonitrile, 3.5 % trifluoroacetic acid and nearly saturated alpha-Cyano-4-hydroxycinnamic acid. By this combination sample processing should be simpler and time efficient. But samples prepared by this method didn't yield any MALDI-TOF spectra with sufficient signal intensity and good signal-noise ratio. By using lower volume of extraction solution, resulting in a much higher sample amount-to-extraction solution ratio, no spectra with satisfactory signals could be recorded. Even with only a quarter of the volume used for the formic acid-acetonitrile approach for the same amount of truffle powder, the second remained superior. Both tested sample preparation methods used acetonitrile as the organic component for extraction. Therefore, the different acidic compositions could account for the difference in extraction efficiency of proteins from ground truffle powder. While the final formic acid concentration was 35 % (v/v) in the first protocol the combined extraction and matrix solution only

contained 2.5 % (v/v) trifluoroacetic acid. Higher acid content clearly had a positive effect on extraction efficiency. Therefore only the formic acid-acetonitrile protocol for sample preparation was further tested in terms of repeatability and capability to differentiate truffle samples from different species. It fulfilled the requirements in both respects. One truffle sample was extracted and measured in triplicates on three different days. Acquired spectra verify the high repeatability observed for bacterial samples in truffle samples as well, shown in Figure 11.

Next, the distinction of truffle species was possible because of species specific signal profiles in acquired mass spectra. These signal patterns were repeatable apparent in all biological samples within the three used species and thereby attributed to their biological identity (Figure 12). Principal component analysis and hierarchical clustering analysis of processed spectra showed biological samples within a species are most similar and separated from the other truffles. Further, even technical replicates within one biological sample were most similar (Figure 13).

With capability to differentiate truffles, all samples were measured and acquired spectra processes. In hierarchical clustering, all samples within a species were most similar and different truffle species separated from each other (Figure 14) with exception of one outlier. This outlier was not as visible in the sample projection of principal components. Nevertheless, it was removed from further analysis to not distort following results. By the acquisition of MALDI-TOF spectra signal pattern profiles for different truffle species were generated. These profiles can be compared for *T. magnatum*, *T. aestivum*, *T. melanosporum*, *T. indicum* and *T. albidum* Pico. Each species shows distinct signal patterns in acquired spectra (Figure 16). Karkouri et al.⁷⁶ previously used MALDI-TOF biotyping derived from microbiology to study commercial truffle samples. Formic acid-acetonitrile extraction, as previously described, was used and crystallization done with alpha-Cyano-4-hydroxycinnamic acid. Measurements were performed in the same scan range as described for bacteria. In their work they could identify seven different *Tuber* species. Four of the investigated species were shared between the studies of Karkouri and this thesis, enabling a comparison of generated profiles for *T. melanosporum*, *T. magnatum*, *T. indicum* and *T. aestivum*. Profiles for *T. aestivum* looked very similar with the strongest and most distinct signal at 2,200 m/z. For *T. indicum* the signal occurrences within the profiles looked similar, but intensities for these signals were different. In the profile of Karkouri a signal at 9,700 m/z was strongest, in here presented results the

signal at 4,000 m/z. Their *T. magnatum* profile does not show the signal triplet between 14,000 and 16,000 m/z as prominent as in here presented results, while the rest of signals look comparable. Last, in the here presented profile of *T. melanosporum* a signal triplet similar to *T. magnatum* is visible and an additional strong signal at 10,900 m/z. These signals are not present in the profiles of Karkouris study. Next generated MALDI-TOF profiles were transferred into signal intensity barcodes. This barcode form of display includes the same information as MALDI-TOF profiles (Figure 17). Nevertheless, it is much easier to extract visualized information in this form of display. Not all signals in profiles are as easy to observe. Examples are the signal at 2,200 m/z for *T. aestivum*, the signal at 10,900 m/z for *T. melanosporum* or the signal at 11,800 m/z for *T. albidum Pico*. Especially for more subtle differences like the shift in maxima of signal intensity over a curve shaped peak assembly, observed in *T. magnatum* and *T. melanosporum*, the barcode form of display is helpful to easily spot differences.

Last step in the MALDI-TOF approach was a classification analysis (Table 18). An according function is implemented in the MassUp software. With a correct classification for 95 % of samples using NaiveBayes classifier the model performed well. Only one sample for *T. albidum Pico* was classified incorrectly. For MALDI-TOF analysis *T. albidum Pico* had the smallest sample set present with only 4 samples. From these 4 samples, two were used as training instances for the classification model. Such a low number of training instances will not lead to a good performance in subsequent classification and matching. A higher number of samples for training would be needed to achieve reliable and correct matching. Using a second classificatory, the random forest classification, and accuracy of the model dropped. Only 21 out of 24 instances were classified correctly. *T. albidum Pico* was incorrectly classified again, most likely due to the still small number of instances. Further, one sample of *T. melanosporum* was not classified correctly. While *T. aestivum* and *T. magnatum* had a higher number of instances, for *T. melanosporum* and *T. indicum* much less samples were available, but still more than for *T. albidum Pico*. Therefore, it is more likely that a mismatch occurs for one of them than for the two species with higher instance numbers. This result marks one downside of classification based on machine learning models. A prediction model of which model will perform best on a given set of data is hard to determine in advance⁷⁷. Often a trial-and-error approach for finding the best classification method is needed⁷⁸⁻⁸⁰.

After the fast identification approach analyzing intact proteins, bottom-up proteomics was performed. Typical and well established bottom-up proteomic workflows for protein extraction and tryptic proteolysis normally used for cells or mammalian tissue were tested on a ground powder truffle sample. First protein extractions with buffers containing sodium deoxycholate (SDC), sodium dodecyl sulfate (SDS) and urea was tested. Resulting protein concentration was estimated by a BCA assay (Figure 18). SDS showed the highest extraction efficiency (6.2 $\mu\text{g}/\mu\text{l}$), indicating most effective membrane disruption and protein solubilization for truffle samples. Extraction efficiency of urea (4.5 $\mu\text{g}/\mu\text{l}$) was superior to SDC (3.6 $\mu\text{g}/\mu\text{l}$). Nevertheless, usage of all extraction agents resulted in sufficient protein concentrations for further processing of samples without limitation in material. Enough material could be extracted from only milligram amounts of ground truffle powder. No further steps for enzymatically breaking down the chitin containing cell wall of truffles were needed⁸¹.

For a more comprehensive comparison protein identification rates by further processing samples with different approaches were compared (Figure 19). Like all detergents SDS needs to be removed prior to mass spectrometric measurement. Left in samples SDS can lead to strong signal suppression⁸². Therefore removal by the filter aided sample preparation (FASP) approach, introduced by Wiśniewski⁶⁷, was performed using two different centrifugal filter devices. Filters utilized either 10 kDa or 30 kDa molecular weight cutoff membranes. Using both filters, protein identification rates for the FASP approach displayed 1,129 (± 45) identified proteins with the 10 kDa filter and 1247 (± 9) with the 30 kDa filter. Usage of the SDC protocol achieved an equally high number of identified proteins (1,169 (± 6)) as the FASP approach with SDS extracted samples. Low concentrations of SDC in solution do not affect the proteolytic activity of trypsin. As the extraction buffer contained only 1 % SDC no prior removal before tryptic proteolysis was required. It was shown that even with 10 % SDC more than 70 % of trypsin activity is retained⁸³. With the lowest obtained extraction efficiency SDC concentration could be increased to test again for a better extraction yield. But with the yield in extracted protein still high enough for subsequent analysis without shortage of material, this test was not needed.

The FASP approach for processing samples extracted by urea led to lowest protein identification rates. A common risk regarding the application of urea is carbamylation of proteins by isocyanic acid. In aqueous solutions urea spontaneously dissociates into cyanate

and ammonium. This modification can firstly block proteolytic cleavage and, when not considered for database search, can lead to low identification rates. Temperature, incubation time and pH value can affect the degree of urea dissociation. Therefore ammonium hydrogen buffer with incubation and sonication on ice was used⁸⁴. Checking for carbamylation was done by including it as possible modification in the database search. With only 1 % of identified proteins modified it was excluded as possible cause for the low identification rate. As urea had a higher efficiency to extract proteins than SDC, seen in the higher obtained protein concentration, it was concluded that not all proteins from the complex truffle proteome were extracted to the same extend. There is a bias towards extraction of certain proteins using urea. A better performance with urea extracted samples was obtained with the Capturem™ kit, which utilizes immobilized trypsin for a fast proteolytic cleavage of proteins. Nevertheless, the protein identification rate was still much lower than for the SDC protocol and SDS extracted samples prepared with the FASP protocol. A problem regarding the usage of immobilized trypsin in the Capturem™ kit can be the short time for proteolysis of a complex protein mixture. In comparison, other tryptic proteolysis protocols emphasizing fast digestion are done in solution and utilizing ultrasound⁸⁵ or microwaves⁸⁶ to enhance trypsin efficiency. Because of the low identification rates urea extracted samples processed by FASP and Capturem™ kit were not included in further experiments. Both sample processing methods did not fit the goal of performing in depth proteomic profiling of truffle samples. With its rapid sample processing time of only a few minutes the Capturem™ kit still identified several hundred proteins and could be greatly utilized in other areas than in depth proteomic profiling.

Identified proteins had a higher quantitative reproducibility when prepared by the SDC protocol. Reproducibility was lower for SDS extracted samples prepared by FASP with both molecular weight cutoff filters. Mean coefficient of variation (CV) for SDC prepared samples was 15.9 %. SDC can be sufficiently and nearly totally removed by acidification⁸³. Therefore, chromatographic separation, ionization process and mass spectrometric measurement are only minimally affected by remaining SDC. FASP settled close to SDCs' reproducibility with application of the 30 kDa cutoff filter (mean CV of 16.3 %). FASP with 10 kDa filters led to a low reproducibility with a mean CV of 27.9 %. Poor repeatability obtained by the 10 kDa cutoff filters is due to the filter units' conical shaped membranes. While the 30 kDa cutoff filters are flat bottom filters, conical bottom membrane filters retain a small reservoir of liquid and SDS is not efficiently removed. Remaining SDS is leading to a loss of sensitivity⁸⁷. This

effect on signal intensity results in misleading assumptions of protein concentration and therefore wrong quantitative values in a relative comparison. Based on the simple, still quick and accessible sample processing with high identification rate and superior quantitative reproducibility, SDC protocol was the protocol of choice.

Handling large sample sets, quick and simple sample preparation methods are preferred. Performing a BCA assay for protein concentration estimation proves to be a time consuming step in this process. To overcome this step a simplified extraction with SDC buffer volume based on the amount of truffle powder weighed in was successfully tried. Using 14 μ l SDC buffer per milligram truffle sample, highest observed deviation in extracted protein amount for a set of different truffle samples was 39 % (Figure 20). Differences in used protein amount for subsequent tryptic proteolysis didn't show strongly in quantitative protein values. Further small differences are removed by a routinely used normalization step in data processing. This is seen by the hierarchical clustering of the same samples prepared according to the BCA assay estimation and prepared by weight-volume ratio in Figure 21. Within a sample set there is no visible distinction between both preparation approaches by quantitative protein values. Same behavior of samples is observed in the multi scatter plot of Figure 22. Additionally shown is the Pearson correlation value. This is a value for measuring the degree of linear correlation between two data sets. Comparing quantitative protein values acquired after both sample preparation approaches, all Pearson correlation values were over 0.9. A value of 0 indicates no correlation and a value of 1 perfect linear correlation. Therefore, the difference in generated quantitative values for proteins using either method is negligibly small. Both preparation approaches could be used in the future to prepare truffle samples.

A second time consuming step in bottom-up proteomics is incubation time for tryptic proteolysis, routinely carried out overnight. Established trypsin-protein ratio for proteolysis is 1:100. To possibly shorten the incubation time to one hour, higher trypsin-protein ratio of 1:25 were tested. Similar approaches were previously applied with success⁸⁸. Protein identification rates stayed constant for overnight and shortened incubation time (Figure 23). Reproducibility of the 1 hour digests was lower on peptide level than overnight, indicated by the higher standard deviation. Therefore 1 hour digestion is not enough time for complete cleavage of proteins into peptides. This lower peptide identification rate is not transferred to protein level as most proteins are still identified with 2 or more unique peptides.

Further, missed cleavage rates of trypsin were compared. Missed cleavages are sites where trypsin didn't cleave the suitable peptide bond. They occurred as expected: With more incubation time and higher trypsin to protein ratio the number of missed cleavages decreased (Figure 24). Here, more trypsin and more time will enable more comprehensive cleavage of suitable sites. There was no strong difference in missed cleavage rate within results for both sample sets with incubation time overnight. Therefore, the overnight incubation with a 1:100 trypsin-protein ratio was further carried out. However, with further testing and careful selection of peptides used for quantitative assays shortened incubation time with higher trypsin amount can be a feasible approach.

After sample preparation and possible simplification of it was done, deep proteomic profiling of truffles was performed. Following analysis was based on LC-MS/MS data with relative quantification between samples. There are different approaches for quantitative bottom-up proteomics, each of them with their own advantages and disadvantages. Labeled approaches, as metabolic labeling and isobaric chemical labeling, have the advantage of high accuracy, precision and reproducibility in quantification. Further they allow multiplexing of samples. This can reduce measurement time and allow the analysis of samples which would be too low in sample amount if measured individually. Downsides are the high costs of isobaric labeling reagents and lower proteome coverage in comparison to label-free approaches⁸⁹. This is due to the increased complexity of samples introduced by isobaric components. Possible precursor ions for fragmentation are now present in twice the number, or in case of isobaric labeling in multiplexed approaches, require more elaborate fragmentation of precursor ions. Therefore a label-free quantification approach was chosen for a most comprehensive characterization.

For reliability in statistical analysis, only species with at least three biological samples were integrated in the process of data analysis. A first set of available samples was prepared and measured. By data-dependent acquisition mode (DDA) 222 proteins were identified with quantitative values in all 40 samples. These quantitative values of proteins were already sufficient for differentiation of the truffle species. In the sample projection after principal component analysis *T. magnatum* and *T. melanosporum* were separated from *T. aestivum* and *T. uncinatum* (Figure 25). *T. aestivum* and *T. uncinatum* could not be separated from each other. There was a long and still ongoing debate whether *T. aestivum* and *T. uncinatum* are different species. Morphologic and phylogenetic analysis resulted in the conclusion that they are co-specific^{72,73}. This co-specificity is shown here on proteomic level as well. Combining

both species under one name, all three species could be separated by quantitative values of identified proteins. This already indicates strong and species specific differences in protein abundance.

With Anova (analysis of variance) testing quantitative protein profiles are generated, seen in Figure 26. While students T-tests are used to compare only two sample sets, Anova testing is needed for the comparison of three or more groups of samples. Proteins responsible for differences in the species are obtained by filtering for proteins which are Anova test significant. Previously gained impression of strong species specific differences was further verified by Anova testing. Of 222 proteins identified in all samples, 161 proteins or 73 % of them were Anova test significant. Already conclusive, the results lacked the desired depth and species of commercial interest were missing at this stage. With only 161 proteins in the quantitative protein profiles the data-dependent acquisition (DDA) approach was not comprehensive enough. DDA lacks in reproducibility of precursor ion selection⁹⁰ and has the problem of undersampling⁹¹, seen in the low protein identification rate of the performed experiment.

To overcome limitations of DDA, the more recently established measurement of samples in data-independent acquisition mode (DIA) was done for truffles. Data extraction of samples measured by DIA requires a previously generated spectral library⁴⁸. This library is commonly created by DDA measurements of the same samples which are later on used for DIA. Only peptides and corresponding proteins which are deposited in the spectral library can be identified and subsequently quantified. Therefore the depth and identification rates by DDA measurements for generation of a library have a high impact on the following DIA analysis. With higher proteomic coverage of a library, more comprehensive analysis of samples in DIA is enabled. To reach this degree of proteomic coverage sample fractionation is needed⁹². Different fractionation methods can be applied. Mostly used are different chromatographic separation approaches. One example for separation of peptides is separation by strong cation exchange chromatography (SCX) followed by reversed phase chromatography. An online approach of this two-dimensional liquid chromatography was introduced by Yates. It was called multidimensional protein identification technology, short MudPIT⁹³. For tryptic peptides SCX followed by reversed phase separation has limited separation efficiency. Peptides with two or three positive charges tend to elute in clusters⁹⁴. More recently used in many proteomic studies, another 2D-LC fractionation approach is combination of offline

reversed phase fractionation at high pH followed by subsequent low pH reversed phase separation and MS/MS measurement. This approach is convenient because low pH reversed phase chromatography is the most common form of chromatography applied before MS/MS measurement of tryptic peptides. Efficiency of fractionation by high pH reversed phase chromatography in a 2D-LC approach can further be improved with fractionation concatenation^{49,50,95}. Fractionation concatenation is a pooling scheme of generated fractions to best utilize chromatographic separation space with reduction of analysis time.

A 2D-LC-MS/ proteomic approach with high pH reversed phase and concatenated fractionation scheme prior to low pH LC-MS/MS was performed to generate a spectral library. Only truffle species that were sequenced are *T. magnatum*, *T. aestivum*, *T. borchii* and *T. melanosporum*^{96,97}. With these limited protein databases available on UniProt and high differences in sample numbers for available different truffles, only samples of *T. magnatum*, *T. aestivum* and *T. melanosporum* were used to generate the library. Missing out on proteins specific for both remaining species, which were not used to generate the library, is possible. Here DIA gives the possibility to expand spectral libraries at a later point with more samples and new databases available. After extension of libraries, datasets can be re-analyzed without the need for re-measuring all samples. Containing 9,170 proteins and 51,628 peptides from the three species, here presented is the most comprehensive library of identified truffle proteins and highest number of identified truffle proteins at all.

With DIA measurements of truffle samples deeper proteomic coverage was reached. Instead of only a few hundred identified proteins now 2,715 proteins with quantitative values were identified over all samples. Principal component analysis prior to statistical analysis using quantitative protein values already shows an exceptionally well separation of different truffle species (Figure 27). This high degree of difference between truffles is also reflected in the fingerprints generated by Anova testing and displayed in a heatmap (Figure 28). From the 2,715 proteins identified in all samples, 2,066 were Anova significant and therefore contributing to the separation of truffle species. Obtained quantitative protein fingerprints for truffles were further reduced to a protein panel consisting of 15 marker candidates. By introduction of the protein panel a helpful tool for comparing truffle samples and a starting point for future perspectives should be generated. Reference proteins for normalization were included in a similar fashion to housekeeping genes used in transcriptomics⁹⁸. They should be as constant as possible in their abundance within all samples (Figure 29). Marker candidate

proteins for the panel were then carefully chosen by a multi-step assessment (Figure 30). This manual assessment is labour-intensive and time consuming. For future perspectives a software based automation of this assessment would be helpful. Requirements to a software solution would be the collected display of gathered information about proteins. In the best case a ranking of marker candidates can be done automatically. With help of the panel a quick distinction of truffles by assessment of the abundance ratios for 15 selected proteins is possible, as each truffle species shows a specific profile (Figure 31). This distinction is further simplified by bi-directional plotting of protein ratios for two species in a direct comparison (Figure 32 and Figure 33), comparable to MALDI-TOF biotyping.

Comparing truffles species-against-species, highest differences in all t-test based direct comparisons of truffle species were shown between *T. magnatum*, *T. melanosporum* and *T. aestivum* (Table 19). These truffles were used to generate the spectral library for data extraction, databases and samples for *T. albidum* and *T. indicum* were not included. Therefore, overrepresentation of proteins for these three species in the generated dataset is possible. A quantification approach independent from previous identification could overcome this problem. Knowledge of protein identity is not required for finding differences between truffles. Mass spectrometric data with chromatographic separation beforehand contains sufficient information about precursor ions. Retention time, m/z value, isotopic pattern and generated fragment ions on MS2 level could be used to be defined as a feature. By this features a comparison of samples independent of identification could be enabled. Both MS1 and/or MS2 level quantification can be performed and relative abundances be compared. But momentarily all routinely used data processing software enables quantification of identified proteins only.

After working out quantitative differences on global proteome level as well as for all species, bottleneck for biological interpretation was the availability of functional information about truffles. Only 14 out of 7,491 protein entries in the UniProt database from *Tuber melanosporum* are reviewed entries. Reviewed entries are manually annotated with information extracted from literature and curator-evaluated computational analysis. Both *Tuber magnatum* with 9,412 entries and *Tuber aestivum* with 9,311 entries are completely un-reviewed. Out of all deposited proteins from the *Tuber* genus on UniProt only 22 of 39,055 entries in total corresponding to 0.72 % are reviewed (status of 04.12.2020). Although first attempts to functionally annotate the black truffles' proteome in 2010 were done, most

proteins are still not annotated with functions or involvement in biological processes⁹⁹. Improvements were done 2018 by further sequencing of three more truffle species, mentioned previously. But still, many proteins remain uncharacterized and functions are not annotated. For *T. magnatum* only 57.6 % of all protein entries are functionally annotated, for *T. melanosporum* its 64.7 % and for *T. aestivum* 46.8 %.

Comparing all different truffles first on global proteome level by gene ontology enrichment analysis of Anova significant proteins, many different metabolic processes and oxidation-reduction processes were over-represented biological processes (Table 21). These differences in metabolic and oxidation-reduction processes were further worked out in more detail by species-to-species comparisons. They can be possibly linked to different growing conditions for each species. For *T. magnatum* it was shown, that aroma and responsible metabolites are influenced by genetically differentiated populations¹⁰³, different bacterial¹⁰⁴ or fungal^{105,106} communities associated with the fruiting body and most strongly by geographical area of origin¹⁰², connected to the climate. In plant metabolomics exogenous factors play a crucial role in the metabolomes' composition. Climatic conditions as warmth, heat waves, drought or frost as well as soil composition, level of elevation, sun exposure and more have to be taken into account. All these factors can have a strong impact on a plants metabolome. Seasonal differences for plants grown in the same geographical location can show strongly on metabolomic level¹⁰⁷. To include the impact of these variations, multi-seasonal sample acquisition needs to be done. Same observations will most likely apply to truffles. Proteins with their enzymatic activity are responsible for all metabolic reactions. Therefore observed enrichment of metabolic and Redox-processes in proteins for different truffles is in accord with these previous metabolomics observations.

Further, analysis of underlying biological processes based on obtained proteomic data revealed some additional conclusive findings. Comparing the white truffle *T. magnatum* with the black Chinese truffle *T. indicum*, sulfur amino acid and cysteine metabolic processes, sulfur compound metabolic processes and hydrogen-sulfide metabolic processes were over-represented in proteins up-regulated in *T. magnatum* (Table 22). Sulfur-containing volatile organic components (VOCs) play an important role for truffles aroma¹⁰⁰, especially for *T. magnatum*^{101,102}. These differences on VOC level are therefore in accord with our findings on proteomic level. A higher abundance of proteins responsible for processes of sulfur-metabolism can lead to more sulfur-containing VOCs in *T. magnatum*, resulting in the

stronger aroma. In comparison of *T. magnatum* with the other truffles *T. indicum*, *T. aestivum* and *T. albidum*, sulfur-metabolic processes were over-represented as well. Similar results were obtained for the black truffle *T. melanosporum*, also known for a strong aroma and sulfur-containing VOCs⁸. Obtained results can be used as repository for future research. With better characterization and annotation of function for truffle proteins a biological interpretation of differences discovered will be possible.

Last, truffles should be investigated regarding their geographical origin. Other than on metabolomics level, no conclusion could be drawn about the origin of truffles analyzing the proteome. Sample projections of principal components, seen in Figure 34, didn't show any clustering or separation of samples linked to their origin. Vita et al found proteins specifically linked to truffles origin for *T. magnatum*¹⁰⁸. Their approach was based on 2-dimensional gel electrophoresis followed by mass spectrometry. Protein extracts of *T. magnatum* samples from different origin were separated on a 2-dimensional gel electrophoresis system. Resulting gels were compared and only spots differentiating in intensity for samples of different origins were chosen for subsequent proteomic analysis. These origin-linked proteins identified by Vita et al were not observed to be origin specific within here presented global proteome dataset. There are differences in abundance for these proteins within samples from *T. magnatum*, but they are not linked to origin. Factors like ripeness of the fruiting body, time point of harvesting, storage, handling and many more will have an impact on a truffles proteome as well. Other techniques that don't analyze the proteomic composition have shown promising results in tracing back geographical origin of truffles. Chemometrics in combination with inductively coupled plasma mass spectrometry (ICP-MS) for analysis of the elemental composition was successfully applied to classify truffle samples based on their origin¹⁰⁹. By analysis of volatile organic compounds, samples of *T. magnatum* could be even traced back to different geographical origins within Italy¹¹⁰.

9. Outlook

In this thesis truffles were investigated by both fast and deep profiling on proteomic level.

MALDI-TOF analysis can be additionally performed on more truffle species of commercial interest. Further advances regarding miniaturization and automation of the sample processing could be made. A possible goal would be the transfer the whole process into the microtiter plate format and working with pipetting and spotting robots for a quick and large scale sample preparation and analysis of big sample sets by MALDI-TOF.

With introduction of the protein panel with 15 proteins a helpful tool for further LC-MS based methods. Next step for a quicker and more sensitive analysis would be the establishment of a targeted approach. This would be done in the form of selected reaction monitoring (SRM)/ multiple reaction monitoring (MRM) on a triple quadrupole mass spectrometry instrument or alternatively parallel reaction monitoring (PRM) on a high resolution mass spectrometer. To implement targeted quantification, the protein panel is a good starting point for marker candidates. Suitable peptides of these proteins need to be chosen, with different retention times in the chromatography, no introduced modifications, good ionization and desorption properties and good fragmentation in the mass spectrometer. For these candidate peptides heavy isotope labeled versions have to be synthesized. Furthermore optimal transition settings need to be tested for maximum signal intensity. Measurements to determine limit of detection (LOD) and limit of quantification (LOQ) for each peptide need to be performed as well. This targeted approach can be tried to be implemented on larger dimensioned chromatography systems for a quicker and more robust analysis. Optimization of sample preparation showed, that limited sample amount will not be a bottleneck for a possible transfer from UPLC- to HPLC-systems with their higher column diameters, higher flow rates and shorter gradients. Orthogonal to chromatography and mass spectrometry, immunological techniques for detection and quantification could be applied as well. Candidates from the protein panel could further be implemented in antibody-based detection methods. Enzyme-linked immunosorbent assays (ELISA) for detection and quantification of suggested proteins from the panel can be developed as an alternative approach for sample identification and verification of MALDI-TOF results.

Last, obtained LC-MS/MS data can be used as repository for future scientific work on truffles. With still limited functional annotation of truffle proteins, biological analysis and interpretation of proteomic differences between truffles is only at a beginning. Only further characterization of truffle proteins will enable a comprehensive analysis and interpretation of underlying biological processes.

10. References

1. Wysocki, C. J., Dorries, K. M. & Beachamp, G. K. Ability to perceive androstenone can be acquired by ostensibly anosmic people. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7976–7978 (1989).
2. Zambonelli, A. True Truffle (Tuber spp.) in the World: Soil Ecology, Systematics and Biochemistry. *Soil Biol.* **47**, 436 (2016).
3. James M. Trappe, A. W. C. The Hidden Life of Truffles. *Sci. Am.* **302**, 78–85 (2010).
4. Hock, B. *Fungal associations 2nd Edition. Fungal Associations, 2nd Edition* **9**, (Springer Berlin Heidelberg, 2012).
5. Allen, M. F., Swenson, W., Querejeta, J. I., Egerton-Warburton, L. M. & Treseder, K. K. Ecology of Mycorrhizae: A Conceptual Framework for Complex Interactions Among Plants and Fungi. *Annu. Rev. Phytopathol.* **41**, 271–303 (2003).
6. Kohler, A. *et al.* Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, 410–415 (2015).
7. Tedersoo, L., Arnold, A. E. & Hansen, K. Novel aspects in the life cycle and biotrophic interactions in Pezizomycetes (Ascomycota, Fungi). *Mol. Ecol.* **22**, 1488–1493 (2013).
8. Culleré, L. *et al.* Characterisation of aroma active compounds in black truffles (*Tuber melanosporum*) and summer truffles (*Tuber aestivum*) by gas chromatography-olfactometry. *Food Chem.* **122**, 300–306 (2010).
9. <https://www.tuber.it/en/truffle/life-cycle/> entered 26.02.2020.
10. Schmidberger, P. C. & Schieberle, P. Characterization of the Key Aroma Compounds in White Alba Truffle (*Tuber magnatum pico*) and Burgundy Truffle (*Tuber uncinatum*) by Means of the Sensomics Approach. *J. Agric. Food Chem.* **65**, 9287–9296 (2017).
11. Payen, T. *et al.* A survey of genome-wide single nucleotide polymorphisms through genome resequencing in the Périgord black truffle (*Tuber melanosporum* Vittad.). *Mol. Ecol. Resour.* **15**, 1243–1255 (2015).
12. Mello, A., Murat, C. & Bonfante, P. Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiol. Lett.* **260**, 1–8 (2006).
13. Martin, F. *et al.* Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**, 1033–1038 (2010).
14. MASTER CLASS: The Secret Life of Truffles by Marc Eber | The Depanneur. Available at: <https://thedepanneur.ca/event/master-class-the-secret-life-of-truffles-by-marc-eber/>. (Accessed: 6th July 2020)
15. Streiblová, E., Gryndlerová, H. & Gryndler, M. Truffle brûlé: an efficient fungal life strategy. *FEMS Microbiol. Ecol.* **80**, 1–8 (2012).

16. Zampieri, E., Chiapello, M., Daghino, S., Bonfante, P. & Mello, A. Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles. *Sci. Rep.* **6**, 1–11 (2016).
17. <https://www.decanter.com/wine-travel/italy/truffle-hunting-alba-332816/> entered 07.06.2020. Available at: <https://www.decanter.com/wine-travel/italy/truffle-hunting-alba-332816/>. (Accessed: 6th July 2020)
18. Streiblová, E., Gryndlerová, H. & Gryndler, M. Truffle brûlé: An efficient fungal life strategy. *FEMS Microbiology Ecology* **80**, 1–8 (2012).
19. Reyna, S. & Garcia-Barreda, S. Black truffle cultivation: a global reality. *For. Syst.* **23**, 317–328 (2014).
20. Truffle Cultivation – New World Truffieres. Available at: <http://www.truffletree.com/cultivation/>. (Accessed: 13th October 2020)
21. Site Selection for Optimum Truffle Cultivation. Available at: <http://www.truffletree.com/site/>. (Accessed: 13th October 2020)
22. Truffle Plantation Establishment & Management. Available at: <http://www.truffletree.com/plantation/>. (Accessed: 13th October 2020)
23. Spink, J. & Moyer, D. C. Defining the Public Health Threat of Food Fraud. *J. Food Sci.* **76**, 1–7 (2011).
24. Creydt, M. & Fischer, M. Omics approaches for food authentication. *Electrophoresis* **39**, 1569–1581 (2018).
25. Von Bargaen, C., Brockmeyer, J. & Humpf, H. U. Meat authentication: A new HPLC-MS/MS based method for the fast and sensitive detection of horse and pork in highly processed food. *J. Agric. Food Chem.* **62**, 9428–9435 (2014).
26. Kim, G. D., Seo, J. K., Yum, H. W., Jeong, J. Y. & Yang, H. S. Protein markers for discrimination of meat species in raw beef, pork and poultry and their mixtures. *Food Chem.* **217**, 163–170 (2017).
27. Stahl, A. & Schröder, U. Development of a MALDI-TOF MS-Based Protein Fingerprint Database of Common Food Fish Allowing Fast and Reliable Identification of Fraud and Substitution. *J. Agric. Food Chem.* **65**, 7519–7527 (2017).
28. Mazzeo, M. F. & Siciliano, R. A. Proteomics for the authentication of fish species. *J. Proteomics* **147**, 119–124 (2016).
29. Gunning, Y. *et al.* Species determination and quantitation in mixtures using MRM mass spectrometry of peptides applied to meat authentication. *J. Vis. Exp.* **2016**, (2016).
30. Wang, J. *et al.* Rapid determination of the geographical origin of honey based on protein fingerprinting and barcoding using MALDI TOF MS. *J. Agric. Food Chem.* **57**, 10081–10088 (2009).

31. Nawrocki, A., Thorup-Kristensen, K. & Jensen, O. N. Quantitative proteomics by 2DE and MALDI MS/MS uncover the effects of organic and conventional cropping methods on vegetable products. *J. Proteomics* **74**, 2810–2825 (2011).
32. Abd El-Salam, M. H. Application of proteomics to the areas of milk production, processing and quality control - A review. *Int. J. Dairy Technol.* **67**, 153–166 (2014).
33. Holland, J. W., Gupta, R., Deeth, H. C. & Alewood, P. F. Proteomic analysis of temperature-dependent changes in stored UHT milk. *J. Agric. Food Chem.* **59**, 1837–1846 (2011).
34. Meltretter, J., Wüst, J. & Pischetsrieder, M. Modified peptides as indicators for thermal and nonthermal reactions in processed milk. *J. Agric. Food Chem.* **62**, 10903–10915 (2014).
35. Wilkins, M. R. *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology. (N. Y.)* **14**, 61–5 (1996).
36. Wasinger, V. C. *et al.* Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**, 1090–4 (1995).
37. Kelleher, N. L. A cell-based approach to the human proteome project. *J. Am. Soc. Mass Spectrom.* **23**, 1617–1624 (2012).
38. Jungblut, P. R., Holzhütter, H. G., Apweiler, R. & Schlüter, H. The speciation of the proteome. *Chemistry Central Journal* **2**, (2008).
39. Kelleher, N. L. Top-Down Proteomics. *Anal. Chem.* **76**, 196 A-203 A (2004).
40. Nilsson, T. *et al.* Mass spectrometry in high-throughput proteomics: Ready for the big time. *Nature Methods* **7**, 681–685 (2010).
41. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C. & Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews* **113**, 2343–2394 (2013).
42. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **85**, 2–9 (2013).
43. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews Molecular Cell Biology* **5**, 699–711 (2004).
44. Nesvizhskii, A. I. Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. in *Mass Spectrometry Data Analysis in Proteomics* 87–120 (Humana Press). doi:10.1385/1-59745-275-0:87
45. Kapp, E. & Schütz, F. Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms. in *Current Protocols in Protein Science* **Chapter 25**, 25.2.1-25.2.19 (John Wiley & Sons, Inc., 2007).
46. Shalit, T., Elinger, D., Savidor, A., Gabashvili, A. & Levin, Y. MS1-based label-free proteomics using a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* **14**, 1979–1986 (2015).

47. Zanivan, S., Krueger, M. & Mann, M. In vivo quantitative proteomics: The SILAC mouse. *Methods Mol. Biol.* **757**, 435–450 (2011).
48. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, (2012).
49. Batth, T. S., Francavilla, C. & Olsen, J. V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **13**, 6176–6186 (2014).
50. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–2026 (2011).
51. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat. Protoc.* **10**, 887–903 (2015).
52. Hillenkamp, F., Karas, M., Beavis, R. C. & Chait, B. T. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Anal. Chem.* **63**, 1193A–1203A (1991).
53. Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Analytical Chemistry* **60**, 2299–2301 (1988).
54. Tanaka, K. *et al.* Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).
55. Lavigne, J. P. *et al.* Mass spectrometry: A revolution in clinical microbiology? *Clinical Chemistry and Laboratory Medicine* **51**, 257–270 (2013).
56. Pranada, A. B., Schwarz, G. & Kostrzewa, M. MALDI biotyping for microorganism identification in clinical microbiology. in *Advances in MALDI and Laser-Induced Soft Ionization Mass Spectrometry* 197–225 (Springer International Publishing, 2015). doi:10.1007/978-3-319-04819-2_11
57. Pavlovic, M., Huber, I., Konrad, R. & Busch, U. Application of MALDI-TOF MS for the Identification of Food Borne Bacteria. *Open Microbiol. J.* **7**, 135–141 (2013).
58. Valentine, N., Wunschel, S., Wunschel, D., Petersen, C. & Wahl, K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Appl. Environ. Microbiol.* **71**, 58–64 (2005).
59. Wunschel, D. S. *et al.* Effects of varied pH, growth rate and temperature using controlled fermentation and batch culture on Matrix Assisted Laser Desorption/Ionization whole cell protein fingerprints. in *Journal of Microbiological Methods* **62**, 259–271 (Elsevier, 2005).
60. Wunschel, S. C. *et al.* Bacterial analysis by MALDI-TOF mass spectrometry: An inter-laboratory comparison. *J. Am. Soc. Mass Spectrom.* **16**, 456–462 (2005).
61. Luzzatto-Knaan, T., Melnik, A. V. & Dorrestein, P. C. Mass spectrometry tools and workflows for revealing microbial chemistry. *Analyst* **140**, 4949–4966 (2015).

62. Freiwald, A. & Sauer, S. Phylogenetic classification and identification of bacteria by mass spectrometry. *Nat. Protoc.* **4**, 732–742 (2009).
63. Sauer, S. *et al.* Classification and Identification of Bacteria by Mass Spectrometry and Computational Analysis. *PLoS One* **3**, e2843 (2008).
64. Reeve, M. A. *et al.* A highly-simplified and inexpensive MALDI-TOF mass spectrometry sample-preparation method with broad applicability to microorganisms, plants, and insects. *J. Biol. Methods* **5**, 103 (2018).
65. López-Fernández, H. *et al.* Mass-Up: An all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. *BMC Bioinformatics* **16**, 318 (2015).
66. Strohalm, M., Hassman, M., Košata, B. & Kodíček, M. mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry* **22**, 905–908 (2008).
67. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
68. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
69. Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrometry Reviews* **39**, 229–244 (2020).
70. Maere, S., Heymans, K. & Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**, 3448–3449 (2005).
71. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
72. Wedén, C., Danell, E. & Tibell, L. Species recognition in the truffle genus *Tuber* - The synonyms *Tuber aestivum* and *Tuber uncinatum*. *Environ. Microbiol.* **7**, 1535–1546 (2005).
73. Molinier, V. *et al.* A multigene phylogeny demonstrates that *Tuber aestivum* and *Tuber uncinatum* are conspecific. *Org Divers Evol* **13**, 503–512 (2013).
74. Maier, T., Klepel, S., Renner, U. & Kostrzewa, M. Fast and reliable MALDI-TOF MS-based microorganism identification. *Nat. Methods* **3**, i–ii (2006).
75. Ryzhov, V. & Fenselau, C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Anal. Chem.* **73**, 746–750 (2001).
76. El Karkouri, K., Couderc, C., Decloquement, P., Abeille, A. & Raoult, D. Rapid MALDI-TOF MS identification of commercial truffles. *Sci. Rep.* **9**, (2019).
77. White, B. J., Amrine, D. E. & Larson, R. L. Big data analytics and precision animal agriculture symposium: Data to decisions. *J. Anim. Sci.* **96**, 1531–1539 (2018).

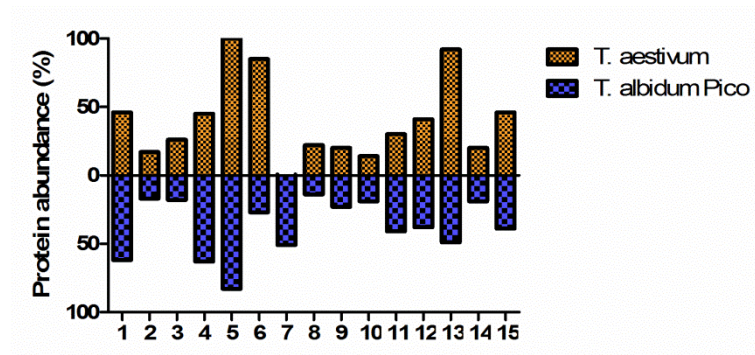
78. Amrine, D. E., White, B. J. & Larson, R. L. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Comput. Electron. Agric.* **105**, 9–19 (2014).
79. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, 321–332 (2015).
80. van der Heide, E. M. M. *et al.* Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *J. Dairy Sci.* **102**, 9409–9421 (2019).
81. Latgé, J. P. The cell wall: A carbohydrate armour for the fungal cell. *Molecular Microbiology* **66**, 279–290 (2007).
82. Rundlett, K. L. & Armstrong, D. W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **68**, 3493–3497 (1996).
83. Lin, Y. *et al.* Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal. Biochem.* **377**, 259–266 (2008).
84. Sun, S., Zhou, J. Y., Yang, W. & Zhang, H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal. Biochem.* **446**, 76–81 (2014).
85. López-Ferrer, D., Capelo, J. L. & Vázquez, J. Ultra fast trypsin digestion of proteins by high intensity focused ultrasound. *J. Proteome Res.* **4**, 1569–1574 (2005).
86. Sun, W. *et al.* Microwave-assisted protein preparation and enzymatic digestion in proteomics. *Mol. Cell. Proteomics* **5**, 769–776 (2006).
87. Lipecka, J. *et al.* Sensitivity of mass spectrometry analysis depends on the shape of the filtration unit used for filter aided sample preparation (FASP). *Proteomics* **16**, 1852–1857 (2016).
88. Egeland, S. V., Reubsæet, L. & Halvorsen, T. G. The pros and cons of increased trypsin-to-protein ratio in targeted protein analysis. *J. Pharm. Biomed. Anal.* **123**, 155–161 (2016).
89. Li, Z. *et al.* Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ orbitrap velos. in *Journal of Proteome Research* **11**, 1582–1590 (American Chemical Society, 2012).
90. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
91. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10**, 1785–1793 (2011).
92. Wang, H., Chang-Wong, T., Tang, H. Y. & Speicher, D. W. Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. *J. Proteome Res.* **9**, 1032–1040 (2010).

93. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247 (2001).
94. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Orthogonality of separation in two-dimensional liquid chromatography. *Anal. Chem.* **77**, 6426–6434 (2005).
95. Dwivedi, R. C. *et al.* Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal. Chem.* **80**, 7036–7042 (2008).
96. Martin, F. *et al.* Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**, 1033–1038 (2010).
97. Murat, C. *et al.* Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nat. Ecol. Evol.* **2**, 1956–1965 (2018).
98. Mane, V. P., Heuer, M. A., Hillyer, P., Navarro, M. B. & Rabin, R. L. Systematic method for determining an ideal housekeeping gene for real-time PCR analysis. *J. Biomol. Tech.* **19**, 342–347 (2008).
99. Islam, M. T. *et al.* Unlocking the Puzzling Biology of the Black Périgord Truffle *Tuber melanosporum*. *J. Proteome Res.* **12**, 5349–5356 (2013).
100. Splivallo, R., Ottonello, S., Mello, A. & Karlovsky, P. Truffle volatiles: From chemical ecology to aroma biosynthesis. *New Phytologist* **189**, 688–699 (2011).
101. Fiecchi, A., Kienle, M. G., Scala, A. & Cabella, P. Bis-methylthiomethane, an odorous substance from white truffle, tuber magnatum pico. *Tetrahedron Lett.* **8**, 1681–1682 (1967).
102. Vita, F. *et al.* Volatile organic compounds in truffle (*Tuber magnatum* Pico): comparison of samples from different regions of Italy and from different seasons. *Sci. Reports 2015 51* **5**, 1–15 (2015).
103. Rubini, A., Paolocci, F., Riccioni, C., Vendramin, G. G. & Arcioni, S. Genetic and phylogeographic structures of the symbiotic fungus *Tuber magnatum*. *Appl. Environ. Microbiol.* **71**, 6584–6589 (2005).
104. Benucci, G. M. N. & Bonito, G. M. The Truffle Microbiome: Species and Geography Effects on Bacteria Associated with Fruiting Bodies of Hypogeous Pezizales. *Microb. Ecol.* **72**, 4–8 (2016).
105. Buzzini, P. *et al.* Production of volatile organic compounds (VOCs) by yeasts isolated from the ascocarps of black (*Tuber melanosporum* Vitt.) and white (*Tuber magnatum* Pico) truffles. *Arch. Microbiol.* **184**, 187–193 (2005).
106. Pacioni, G. *et al.* Isolation and characterization of some mycelia inhabiting *Tuber* ascomata. *Mycol. Res.* **111**, 1450–1460 (2007).
107. Kfoury, N. *et al.* Plant-Climate Interaction Effects: Changes in the Relative Distribution and Concentration of the Volatile Tea Leaf Metabolome in 2014–2016. *Front. Plant Sci.* **10**, 1518 (2019).

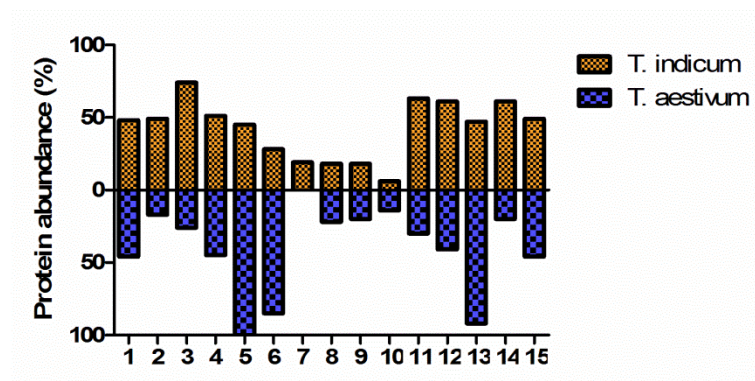
108. Vita, F. *et al.* Tuberomics: A molecular profiling for the adaption of edible fungi (tuber magnatum pico) to different natural environments. *BMC Genomics* **21**, 90 (2020).
109. Segelke, T., von Wuthenau, K., Neitzke, G., Mueller, M.-S. & Fischer, M. Food Authentication: Species and origin determination of truffles (*Tuber* spp.) by inductively coupled plasma mass spectrometry (ICP-MS) and chemometrics. *J. Agric. Food Chem.* (2020). doi:10.1021/acs.jafc.0c02334
110. Gioacchini, A. M. *et al.* Geographical traceability of Italian white truffle (*Tuber magnatum Pico*) by the analysis of volatile organic compounds. *Rapid Commun. Mass Spectrom.* **22**, 3147–3153 (2008).

11. Appendix

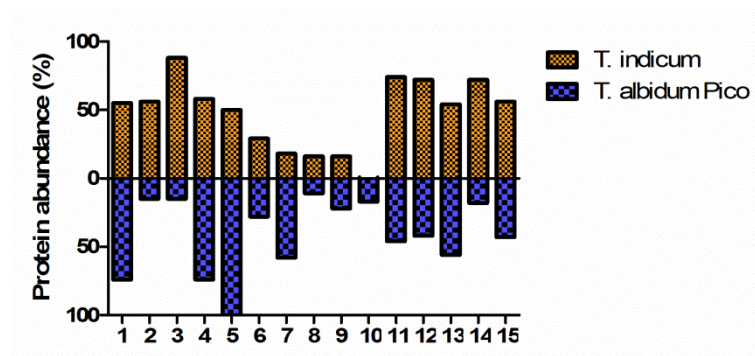
11.1 Supplemental figures



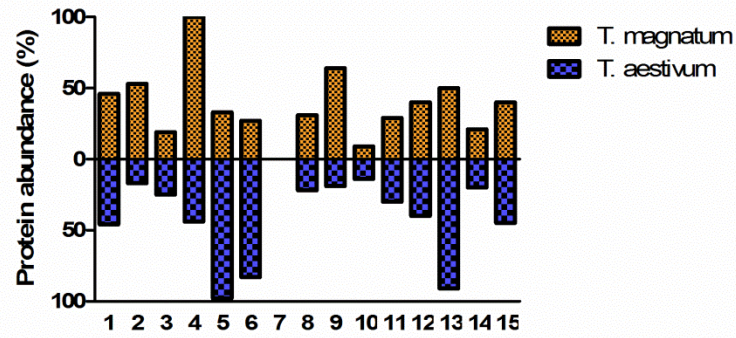
Supplemental Figure 1: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. aestivum* with *T. albidum Pico*.



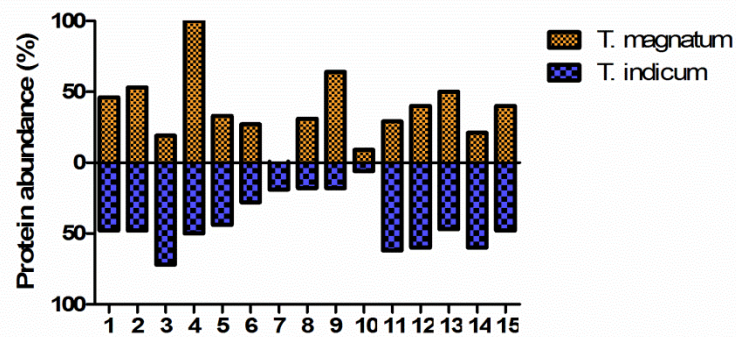
Supplemental Figure 2: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. indicum* with *T. aestivum*.



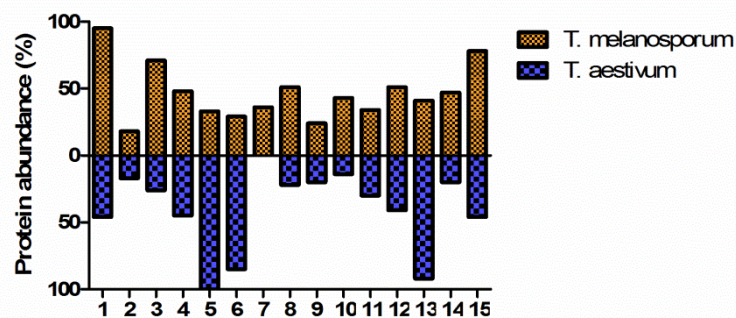
Supplemental Figure 3: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. indicum* with *T. aestivum*.



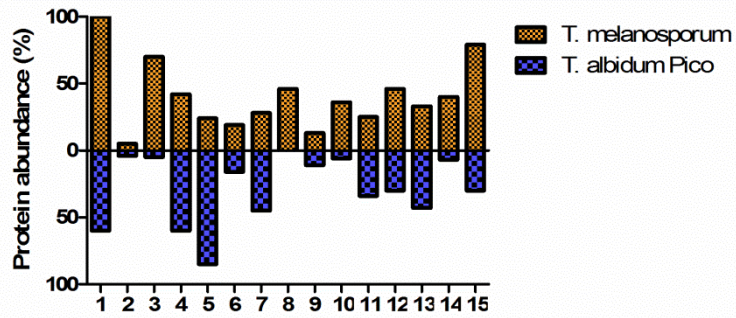
Supplemental Figure 4: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. magnatum* with *T. aestivum*.



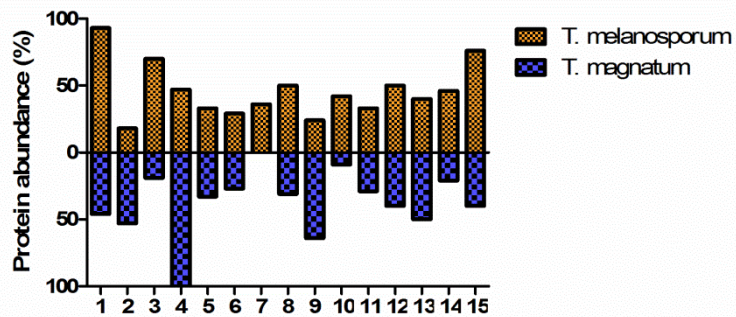
Supplemental Figure 5: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. magnatum* with *T. indicum*.



Supplemental Figure 6: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. melanosporum* with *T. aestivum*.



Supplemental Figure 7: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. melanosporum* with *T. albidum* Pico.



Supplemental Figure 8: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for *T. melanosporum* with *T. magnatum*.

11.2 List of figures

| | |
|---|----|
| Figure 1: The truffles life cycle (Modified after Centro nazionale studi tartufo ⁹)..... | 12 |
| Figure 2: Picture of different economical relevant truffle species ¹⁴ | 13 |
| Figure 3: Truffle hunt with the help of a dog on the left ¹⁷ and the characteristic brûlé around trees in symbiotic association with truffles ¹⁸ on the right. | 13 |
| Figure 4: Schematic representation of the electrospray ionization (ESI) process ⁴² | 17 |
| Figure 5: Schematic construction of the Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific)..... | 18 |
| Figure 6: Graphical comparison of data dependent acquisition and data independent acquisition. ⁵¹ | 22 |
| Figure 7: Principle of Matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) ⁵⁵ | 23 |
| Figure 8: Principle of MALDI-TOF biotyping for bacteria. | 24 |
| Figure 9: Characterization of truffles by fast and deep profiling. | 43 |
| Figure 10: MALDI-TOF spectra for proteins extracted from a truffle sample by formic acid-acetonitrile extraction and combined solution preparation. | 44 |
| Figure 11: MALDI-TOF spectra for proteins extracted from a truffle sample by formic-acid/acetonitrile extraction on three different days in three individual preparations. | 45 |
| Figure 12: MALDI-TOF spectra for protein extracts from different truffle species..... | 46 |
| Figure 13: (A) Sample projection of the first two components based on principal component analysis (PCA) and (B) dendrogram showing the degree of similarity for different truffle species after hierarchical clustering analysis. | 47 |
| Figure 14: Dendrogram showing the degree of similarity for truffle samples of different species after hierarchical clustering analysis. | 48 |
| Figure 15: MALDI-TOF sample projection of the first two components based on principal component analysis (PCA)..... | 49 |
| Figure 16: MALDI-TOF profiles of different truffle species. | 50 |
| Figure 17: Signal intensity barcodes derived from MALDI-TOF spectra of different truffle species. | 51 |
| Figure 18: Bar plot comparing the extracted protein amount using different chaotropes/surfactants/detergents according to a performed BCA assay. | 54 |
| Figure 19: Bar plot comparing the number of identified protein using different sample preparation protocols for truffle powder. | 55 |

| | |
|---|----|
| Figure 20: Bar plot showing the protein amount extracted with SDC buffer from different truffle species and origins. | 56 |
| Figure 21: Heat map displaying protein abundance in different truffle species after hierarchical clustering. | 58 |
| Figure 22: Multi scatter plot comparing the protein intensities obtained by sample preparation according to weight-volume ratio with sample preparation according to BCA test. | 59 |
| Figure 23: Box plots comparing identification rates using different digestion time and trypsin-to-protein ratios. | 60 |
| Figure 24: Grouped box plots comparing the rate of missed cleavages using different digestion times and enzyme-to-protein ratios. | 61 |
| Figure 25: Sample projection of the first two components based on principal component analysis (PCA) for different truffle species. | 62 |
| Figure 26: Heat map using Anova test with 5 % FDR displaying significant protein abundance in different truffle species after hierarchical clustering. | 63 |
| Figure 27: Sample projection of the first two components based on principal component analysis (PCA) for different truffle species. | 65 |
| Figure 28: Heat map using Anova test with 5 % FDR displaying significant protein abundance in different truffle species after hierarchical clustering. | 66 |
| Figure 29: Profile plot showing the intensity of 5 proteins with small changes in abundance over the different truffle species. | 68 |
| Figure 30: Overview of the process to choose proteins for the protein panel on the example of A0A317SIJ8_9PEZI. | 69 |
| Figure 31: Bar plots displaying the protein abundance based profiles for the different truffle species. | 72 |
| Figure 32: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for <i>T. melanosporum</i> with <i>T. indicum</i> | 73 |
| Figure 33: Birectional grouped bar plot comparing the signal intensities of the 15 proteins from the protein panel for <i>T. magnatum</i> with <i>T. albidum</i> Pico. | 73 |
| Figure 34: Sample projection of the first two components based on principal component analysis for truffle samples of different geographical origin. | 77 |
| Figure 35: Schematic overview of generated results. | 78 |

11.3 List of tables

| | |
|---|----|
| Table 1: List of chemicals and enzymes used. | 26 |
| Table 2: List of disposables used. | 26 |
| Table 3: List of devices used. | 27 |
| Table 4: List of chromatography instruments used. | 27 |
| Table 5: List of chromatography columns used. | 27 |
| Table 6: List of mass spectrometry instruments used. | 28 |
| Table 7: List of software used. | 28 |
| Table 8: Tuber species with corresponding origin used for testing the sample preparation by weight of the amount of truffle powder used. | 33 |
| Table 9: Parameters for MS1 spectra acquisition (DDA measurements). | 34 |
| Table 10: Parameters for MS/MS spectra acquisition (DDA measurements). | 34 |
| Table 11: Gradient for separation of peptides by high pH reversed phase liquid chromatography. | 36 |
| Table 12: Concatenated pooling scheme of fractions collected by offline high pH reversed phase fractionation on a HPLC. | 36 |
| Table 13: Parameters for MS1 spectra acquisition (DDA measurements for spectral library). | 37 |
| Table 14: Parameters for MS/MS spectra acquisition (DDA measurements for spectral library). | 38 |
| Table 15: Parameters for MS1 spectra acquisition (DIA measurements). | 39 |
| Table 16: Parameters for MS/MS spectra acquisition (DIA measurements). | 40 |
| Table 17: DIA windows with corresponding start and end m/z values. | 41 |
| Table 18: Classification analysis of truffle samples. | 52 |
| Table 19: Results of the species-to-species comparison listing the number of T-test significant proteins upregulated with an at least two-fold change. | 67 |
| Table 20: Protein Panel for the differentiation of truffle species. | 71 |
| Table 21: Top 10 hits of enriched biological processes in Anova significant proteins from the comparison of all truffles. | 74 |
| Table 22: Top 15 hits of enriched biological processes in upregulated proteins from <i>T. magnatum</i> in comparison with <i>T. indicum</i> | 76 |

11.4 List of publications and conference contributions

Publications:

1. “Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability for the Analysis of Complex Proteomes”

Marcel Kwiatkowski, **Dennis Krösser**, Marcus Wurlitzer, Pascal Steffen, Andrei Barcaru, Christoph Krisp, Péter Horvatovich, Rainer Bischoff, and Hartmut Schlüter.

Anal. Chem. 2018, 90, 16, 9951–9958

Conference contributions:

1. “Differential Proteomic Profiling of Truffles”

Poster, 22nd International Mass Spectrometry Conference, Florenz, 2018










2. “Differential Proteomic Analysis for Validation of Truffle Origin”


Poster, 52. Jahrestagung der Deutschen Gesellschaft für Massenspektrometrie, Rostock, 2019

3. „Quantitative Proteomic Profiles generated by non-targeted Liquid Chromatography Tandem Mass Spectrometry distinguishes different Truffle Species”

Poster, 53. Jahrestagung der Deutschen Gesellschaft für Massenspektrometrie, Münster, 2020

11.5 Safety and disposal

| Chemical name and formula | GHS symbol | Hazard statement | Precautionary statement | Disposal code |
|---|---|-------------------------------------|---|---------------|
| Acetonitrile C_2H_3N |  | 225-332-302- 312-319 | 210-240-302+352- 305+351+338- 403+233 | 1 |
| alpha-Cyano-4-hydroxycinnamic acid $C_{10}H_7NO_3$ |  | 315-319-335 | 261-305+351+338 | 2 |
| Ammonium hydrogen carbonate NH_4HCO_3 |  | 302 | 301+312-330 | 2 |
| Dithiothreitol $C_4H_{10}O_2S_2$ |  | 302-315-319- 335 | 261-305+351+338 | 2 |
| Formic acid CH_2O_2 |  | 226-302-314- 331 EUH: 071 | 210-280- 303+361+353- 304+340+310- 305+351+338- 403+233 | 2 |
| Iodoacetamide C_2H_4INO |  | 301-317-334- 413 | 261-280-301+310- 342+311 | 2 |
| Methanol CH_4O |  | 225-331-311- 301-370 | 210-233-280- 302+352-304+340- 308+310-403+235 | 1 |
| Sodium deoxycholate $C_{24}H_{40}O_4$ |  | 302 | 301-312-330 | 2 |
| Sodium dodecyl sulfate $C_{12}H_{25}NaO_4S$ |  | 228-302+332- 315-318-335- 412 | 210-261-280- 301+312+330- 305+351+338+310- 370+378 | 2 |

| | | | | |
|---|---|---------------------------------|---|---|
| Trifluoroacetic acid $C_2HF_3O_2$ |  | 290-331-314- 412 EUH: 071 | 260-273-280- 303+361+353- 305+351+338-312 | 2 |
|---|---|---------------------------------|---|---|

*1: Disposal in collecting tank for halogen-free, organic solvents and solutions.

*2: Disposal in collecting tank for salt solutions, pH adjusted to 6-8.

Biological materials were disposed in the collection bin for autoclaving waste.

11.6 CMR list

No cancerogen, mutagen or reprotoxic substances (CMR substances) from the GHS category 1A or 1B were used in the work.

12. Danksagung

Zum Abschluss dieser Arbeit möchte ich gerne meine Dankbarkeit an Alle ausdrücken, die mich bei meiner Dissertation unterstützt und begleitet haben.

Der erste und wichtigste Dank geht dabei an meinen Betreuer, Prof. Hartmut Schlüter. Im Bachelor wurde mein Interesse für Chromatographie und Massenspektrometrie geweckt, im Master hat er mich dann durch seine Vorlesungen vollends für dieses Feld begeistert. Vielen Dank für das ermöglichen dieser Arbeit, für die Unterstützung und den Rückhalt in der Betreuung, insbesondere aber die vielen fachlichen und fachfremden Diskussionen mit dem Anspruch, immer kritisch zu hinterfragen.

Weiterhin danke ich Prof. Christian Betzel von der Universität Hamburg, der sich bereit erklärt hat Zweitgutachter dieser Dissertation zu sein.

Großer Dank gebührt auch PD Dr. Friedrich Buck, dem Betreuer meiner Bachelorarbeit, der mich in die Welt der Chromatographie und Massenspektrometrie eingeführt hat. Ein Danke an Sönke Harder für die vielen Kaffees und die langen Gespräche und Diskussionen, egal ob gleicher oder geteilter Meinung. Ein Dank geht auch an all die (ehemaligen) Kollegen und Freunde aus der Arbeitsgruppe. An die nun alle promovierten Dr. Marcel Kwiatkowski, Dr. Marcus Wurlitzer, Dr. Pascal Steffen und Dr. Laura Heikau. An Hannah Voß, Manuela Moritz, Benjamin Dreyer, Dr. Jahn Hahn und nochmal besonders an Dr. Christoph Krisp. Danke für die vielen Anregungen, die vielen Gespräche, die vielen Anregungen und das stets offene Ohr. Zu euch bin ich gerne in die Arbeitsgruppe gekommen, mit euch habe ich gerne zusammen gearbeitet und außerhalb des Labors als Freunde eine wunderbare Zeit gehabt.

Ein riesiges Dankeschön geht an meine Familie. An meine Eltern, meine Großeltern und meinen Bruder. Ihr habt mich immer dabei unterstützt und ermutigt diesen Weg zu gehen, nur durch eure Hilfe, euren Rückhalt und eure Unterstützung war es mir möglich dorthin zu kommen, wo ich nun stehe.

Zuletzt geht natürlich noch ein ganz besonderer Dank an meine bessere Hälfte, Karen. Du hast mich immer weiter und immer wieder dazu motiviert, diese Arbeit fertig zu stellen. Du hast mich beim Schreiben unterstützt, hast dir all meine Gedanken und Abschweifungen dazu angehört (und einiges an Frust, wenn es mal nicht so lief) und mich immer dazu gebracht, das Beste aus mir heraus zu holen.

13. Eidesstattliche Erklärung

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.

Hamburg, den 23.02.2021