



Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food through a comprehensive, high resolution mass spectrometric (HRMS) based approach

Kaatje Van Vlierberghe^a, Maxime Gavage^{b,c}, Marc Dieu^c, Patsy Renard^c, Thierry Arnould^c, Nathalie Gillard^b, Katleen Coudijzer^a, Marc De Loose^a, Kris Gevaert^{d,e}, Christof Van Poucke^{a,*}

^a ILVO Flanders Research Institute for Agriculture, Fisheries and Food, Technology and Food Science Unit, Brusselsesteenweg 370, BE-9090 Melle, Belgium

^b CER Groupe, Rue du Point du Jour 8, 6900, Marloie, Belgium

^c Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, rue de Bruxelles, 5000 Namur, Belgium

^d VIB-UGent Center for Medical Biotechnology, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium

^e Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium

ARTICLE INFO

Keywords:

High-resolution mass spectrometry
Peptide biomarker selection
Food processing
Allergen
Hazelnut

ABSTRACT

The interest of using LC-MS/MS as a method for detection of allergens in food is growing. In such methods, peptides are used as biomarkers for the detection and quantification of the allergens. The selection of good biomarker peptides is of high importance to develop a specific, universal and sensitive method. Biomarkers should, for example, be robust to food processing. To evaluate robustness, test material incurred with hazelnut having undergone different food processing techniques was produced. Proteins of these materials were extracted, digested and further analyzed using HRMS. After peptide identification, selection was carried out using several criteria such as hazelnut specificity and amino acid composition. Further selection was done by comparing peptide MS intensities in the different food matrices. Only peptides showing processing robustness were retained. Eventually, eight peptides coming from three major hazelnut proteins were selected as the best biomarkers for hazelnut detection in processed foods.

1. Introduction

Food allergies – immunological reactions to a certain component, mostly proteins, from a certain food – form a growing health problem around the world, especially in industrialized countries (Costa, Mafra, Carrapatoso, & Oliveira, 2016). Indeed, an estimated 17 million European individuals suffer from food allergy, which is twice as much as 10 years ago (Manea, Ailenei, & Deleanu, 2016). Allergies to tree nuts, among which hazelnut, continues to be globally one of the most common food allergens with an estimated prevalence of up to 4.9% of the general population (Costa et al., 2016; McWilliam et al., 2015). Because of its organoleptic properties, hazelnuts are not only consumed raw, but are widely used in the food industry in chocolate, confectionary, baking, ice cream and dairy products. Furthermore, given their nutritional and nutraceutical benefits, there is a rise in consumption of

hazelnuts in the western countries, which leads to an increase in hazelnut allergy incidence (Özenç & Özenç, 2015).

Allergic patients sensitive to hazelnuts can, after consumption, suffer from rather mild symptoms (hives, dizziness, and/or shortened breath) to very serious and even life-threatening symptoms such as an anaphylactic shock (Özenç & Özenç, 2015). As until today no cure is available to treat food allergies, patients can only protect themselves by avoiding consumption of culprit foods (Popping & Diaz-Amigo, 2017). Therefore, European legislation (Regulation 1169/2011/CE) dictates the declaration of hazelnut and other foods with known allergenic potential when deliberately used as an ingredient, regardless of the concentration or form (Popping & Diaz-Amigo, 2017). However, this is not a complete solution as unintentional presence of trace amounts of hazelnut often occurs, by cross contamination in factories, the use of contaminated ingredients or other causes (Monaci & Visconti, 2010).

* Corresponding author.

E-mail addresses: kaatje.vanvlierberghe@ilvo.vlaanderen.be (K. Van Vlierberghe), m.gavage@cergrupe.be (M. Gavage), marc.dieu@unamur.be (M. Dieu), patsy.renard@unamur.be (P. Renard), thierry.arnould@unamur.be (T. Arnould), n.gillard@cergrupe.be (N. Gillard), katleen.coudijzer@ilvo.vlaanderen.be (K. Coudijzer), marc.deloose@ilvo.vlaanderen.be (M. De Loose), kris.gevaert@vib-ugent.be (K. Gevaert), christof.vanpoucke@ilvo.vlaanderen.be (C. Van Poucke).

<https://doi.org/10.1016/j.foodchem.2019.125679>

Received 22 March 2019; Received in revised form 19 July 2019; Accepted 7 October 2019

Available online 19 October 2019

0308-8146/ © 2019 Elsevier Ltd. All rights reserved.

Since a very small dose of hazelnut can already elicit a reaction in allergic individuals (a few milligrams are enough to cause a severe reaction), this unintentional presence poses a high risk for patients.

To improve protection of patients, accurate, sensitive and reliable analytical methods for the detection and quantification of allergens in food are thus required (Prado et al., 2016). The detection techniques currently applied can generally be divided into two main categories: DNA-based detection (quantitative Polymerase Chain Reaction (qPCR), digital PCR), by which food allergens are detected through amplification of allergen-specific DNA or DNA fragments, and protein-based detection (e.g., Enzyme-Linked Immuno Sorbent Assay (ELISA)), in which detection is based on the binding of a specific protein of the food allergen to a specific receptor (i.e., an antibody) (Prado et al., 2016). Although these techniques are very sensitive and fast, they display some drawbacks, especially for the detection of allergens in processed foods. Indeed, most of the foods consumed in modern developed societies are processed (Sathe, Teuber, & Roux, 2005) and the applied techniques may alter a protein's structure/conformation through denaturation and/or modifications. They could also influence their extractability by formation of protein aggregates. Altering of a protein's structure can thus result in a loss of conformational epitopes, destroying recognition sites used in ELISA detection and thus generate false negative results (Prado et al., 2016). Moreover, cross reactivity of the antibody in the ELISA test with unknown substances can't be fully excluded, giving rise to false positive results (Wei, Sathe, Teuber, & Roux, 2003). DNA-based detection techniques on the other hand do not detect the elicitor (thus the protein) of the allergic reaction itself, which is thus less informative for patients (Poms, Klein, & Anklam, 2004). Indeed, food containing the allergen DNA can be harmless for patients if no allergen proteins are present, while the absence of allergen DNA cannot exclude the presence of allergen proteins.

Another technique that can be used for the detection of allergens in food and which is gaining in popularity is liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) (Pilolli, De Angelis, & Monaci, 2017). This technique is based on the detection of so-called proteotypic peptides, peptides that uniquely match to the ingredient specific protein of interest. These MS-based methods hold particular advantages including the possibility of multiplexing, high sensitivity and specificity, minimization of cross reaction and are qualitative as well as quantitative (Pilolli et al., 2017). Another advantage is the possibility to develop a detection method that is robust to processing techniques, thus universalizing the use of the method. However, in order for the method to be robust to food processing, a proper selection of proteotypic peptides is required. Such peptides have to fulfill several criteria in addition to being specific for the allergen of interest: they have to be stable during food processing and preferably should not contain any amino acids prone to (post-translational) modifications.

In this study we describe a comprehensive approach for the identification and selection of proteotypic peptides robust to food processing that can later be used as analytes for the development of an MS based method for the detection of hazelnut in food. Although several studies based on proteotypic peptides for the detection of hazelnut in food already have already been published (Ansari, Stoppacher, & Baumgartner, 2012; Bignardi et al., 2013; Costa et al., 2016; Cucu, Meulenaer, & Devreese, 2012; Heick, Fischer, Kerbach, Tamm, & Popping, 2011; Pilolli et al., 2017; Pilolli, De Angelis, & Monaci, 2018; Planque et al., 2017, 2019; Sealey-Voyksner, Zweigenbaum, & Voyksner, 2016), to the best of our knowledge, the peptides identified in these studies were not checked for food processing robustness. In the report by Pilolli and collaborators, peptide profiles generated from different food matrices having undergone several processing techniques would need to be compared. However, these matrices should be incurred (incorporation of raw ingredients before food processing), and not spiked (incorporation of extracted proteins from raw ingredients after processing of the matrix). Preferably, production of these matrices can be fully controlled, with exact knowledge of production

circumstances, recipe and ingredients, and keeping cross-contamination events as low as possible. This approach will render trustworthy results, but requires an agreed science based selection of proteotypic peptides.

We tackled this problem by producing five different hazelnut-incurred food matrices. Four different processing conditions (heating, induction of the Maillard reaction, low pH environment and fatty rich environment) were chosen based on techniques that are often used in modern food industry, and material was produced using semi-industrial equipment to mimic realistic conditions of food processing as much as possible. Raw material was self-selected ensuring traceability, and these raw ingredients were used as blank (no processing) samples. These test materials were used for the generation of tryptic peptide profiles by ultra high-performance liquid chromatography coupled to high resolution mass spectrometry (UHPLC-HRMS). Proteotypic peptides were subsequently selected based on multiple selection criteria such as ingredient specificity, amino acid composition and robustness towards food processing. Eventually, we were able to identify eight peptides, originating from two hazelnut proteins – vicilin cor a 11 and legumin cor a 9, two major hazelnut allergens – suitable as good biomarkers for the detection of hazelnut in food.

2. Experimental

2.1. Material

All reagents and chemicals were of analytical grade unless stated otherwise.

Urea, tris(hydroxymethyl)aminomethane (Tris-HCl), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB) and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), and solid phase extraction cartridge (SPE) Pak C18 1 cc Vac Cartridge from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with Trypsin Gold, Mass Spectrometry Grade from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase, Mass Spectrometry Grade (Lys-C) from Wako Pure Chemical Industries, Ltd (Osaka, Japan). Acetonitrile, methanol absolute, 2-propanol (ULC/MS grade for all solvents), and formic acid were from Biosolve (Valkenswaard, the Netherlands). Ammonium phosphate was kindly provided by Palsgaard A/S (Juelsminde – Denmark).

2.2. Production of test materials

Processed hazelnut matrices were produced in the Food Pilot unit of ILVO. All hazelnut matrices were prepared starting from raw hazelnuts from either Turkish origin (Versnoten, Alphen aan den Rijn, The Netherlands) or Italian origin (Quality Nuts BVBA, Zandhoven, Belgium). Both types of purchased hazelnut were dehulled and vacuum packed. Total nitrogen content was determined by the Kjeldahl assay. These values were multiplied with a value of 5.3 (AOAC factor for nuts) (Mariotti, Tomé, & Mirand, 2008) to calculate the protein content in both nut types. This resulted in 14.8% and 14.5% proteins for hazelnuts originating from Turkey and Italy respectively. Production of all test material was done in duplicate, once using hazelnut originating from Turkey and once using hazelnuts originating from Italy.

2.2.1. Preparation of a raw (=unprocessed) hazelnut matrix

Raw whole hazelnuts were grounded using a Kenwood AT286 Glass mini chopper/mill attachment KW714229 (Kenwood, New Hampshire, UK) for 20 s at pulsing mode using maximum speed, vacuum packed (vacuum packaging machine KN5, vc999, Missouri, USA) and stored at 4 °C in the dark.

2.2.2. Preparation of a heated hazelnut matrix

Raw whole nuts were roasted in batches of 500 g in an oven (baking

oven condo CO 6.0608, Miwe, Germany) at 180 °C for 18 min. Nuts were spread in one layer on a baking plate, and each 5 min, nuts were stirred with a spoon to ensure homogenous roasting. Roasted nuts were left to cool down at room temperature, mixed into a nut powder, vacuum packed and stored at 4 °C.

2.2.3. Preparation of a hazelnut matrix subjected to the Maillard reaction (=caramelized nuts)

250 g raw whole nuts were chopped (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3.500 TPM, Robot Coupe, Vincennes Cedex, France) for 3 times 5 s at maximum speed. An amount corresponding to 200 g lactose and 100 mL reverse osmosis water were mixed into a smooth paste. The chopped nuts were subsequently stirred into the lactose paste until nut parts were all covered with the lactose substance. This nut-lactose mixture was spread over an oven plate and placed in a convectional pre-heated (160 °C) oven for 15 min. Every 5 min, the mixture on the plate was stirred with a spoon to obtain equal caramelization. The plate was left at room temperature to cool down overnight, and subsequently mixed into a fine powder (Kenwood Mini Chopper Mill Attachment AT320B). The caramelized nut powder was divided into portions of 50 g, vacuum packed and stored at 4 °C in the dark.

2.2.4. Preparation of a low pH hazelnut matrix (=fermented hazelnut milk)

This matrix was produced by fermenting nut milk using a “set yoghurt” procedure. Nut milk was produced by soaking the nuts in reverse osmosis water overnight, followed by mixing the substance (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3.500 TPM, Robot Coupe, Vincennes Cedex, France) for 8 min at maximum speed upon which a milky substance was reached. Nut milk was subsequently sieved (Retsch Test Sieve, 250 µm pore size, Retsch, Aartselaar, Belgium). Sieved milk was then transferred to a sterile Erlenmeyer of 1 L, and a bacterial culture (nu-trish® BY-Mild – containing the following bacteria: *Bifidobacterium* species, *Lactobacillus delbrueckii* subsp. *Bulgaricus* and *Streptococcus thermophilus*, CHR Hansen, Hørsholm, Denmark) was added (4 g of frozen culture for 600 mL of nut milk). Erlenmeyer recipients were aseptically sealed and incubated in a warm water bath (42 °C). Every 30 min, a sub-sample of the yoghurt was checked for pH until a pH of 4.2 was reached. The yoghurt was subsequently divided into 20 g portions and stored at 4 °C in the dark.

2.2.5. Preparation of a fatty rich hazelnut matrix (=chocolate incurred with hazelnut)

A nut powder was produced by first mixing raw nuts (Kenwood Mini Chopper Mill Attachment AT320B, pulsing speed, 3 min, Kenwood, New Hampshire, UK), resulting in a slushy substance. Nut mash was then defatted by pressing it between extra thick kitchen towels, resulting in a dry nut powder. The defatting step was done to obtain a very fine grain size to ensure a homogenous chocolate after mixing it in. Nut mash and powder was weighed before and after defatting, and loss in mass was documented to determine how much weight was lost during defatting. This fine, dry nut powder was then sieved through a pore size of 500 µm. Secondly, chocolate was prepared: first, a “conche” was made from chocolate refiner flakes (90%) and cacao butter (10%). 450 g refiner flakes (Callebaut, Lebbeke, Belgium) and 50 g cacao butter (Callebaut) were therefore warmed in a water bath in separate bowls until a temperature of 40–45 °C was reached. The temperature was controlled with a food thermometer. Then, the melted butter was slowly added to the melted refiner flakes, and the mix was then stirred for 3 min before addition of 1.2% ammonium phosphatide. The mixture was then stirred for an extra 3 min. Finally, a chocolate incurred with hazelnut (40,207 ppm total hazelnut proteins in matrix for hazelnut origin Italy and 41,046 ppm for hazelnut origin Turkey) was produced by adding 99.996 g of defatted hazelnut powder (containing 24.63 g proteins for hazelnut origin Turkey and 24.12 g proteins for hazelnut

origin Italy) to the liquid chocolate and stirring again for 3 min while keeping the temperature of the chocolate at 42 °C. Ppm levels are mentioned as mg total hazelnut proteins/kg final chocolate product. Liquid chocolate was subsequently poured into chocolate molds, covered with plastic foil and left to cool down. Chocolate was kept overnight at 4 °C to cool further, set and solidify. The next day, chocolate bars were rasped into chocolate flakes. Chocolate flakes were kept in aliquots of 50 g and stored at 4 °C in the dark. A proper blank chocolate was prepared following the same procedure, without adding the hazelnut powder.

2.3. Sample preparation

Sample preparation was similar to the protocol described in our previous study (Gavage et al., 2019). Briefly, proteins were extracted from food matrices, purified and proteolytically digested. From each matrix of both hazelnut origins, three biological repeats were prepared and analyzed by LC-HRMS.

2.3.1. Protein extraction

An amount of 1 g of matrix material was extracted with 10 mL extraction buffer (2 M urea, 200 mM Tris-HCl, pH 9.2) by shaking at 20 °C for 30 min at 300 rpm, followed by 15 min sonication on ice at 100% ultrasound power. The samples were centrifuged at 4600g for 10 min at 4 °C after which the middle liquid phase was transferred to a new tube. The protein concentration of the extract was subsequently determined with A280 ultra-violet absorbance protein assay (NanoPhotometer® N60, Westburg, Leusden, Nederland) (Desjardins, Hansen, & Allen, 2009).

2.3.2. Protein reduction, alkylation and purification

Samples were diluted with extraction buffer to a protein concentration of 4 mg of proteins per mL. Next, proteins were reduced and alkylated (for 15 min at 37 °C in the dark) with TCEP-HCl (15 mM final concentration) and IAA (30 mM final concentration) under agitation at 500 rpm. The samples were desalted and buffer-exchanged by gel filtration chromatography (NAP-10 columns, GE Healthcare, United Kingdom). Proteins were eluted in 1.5 mL 6 M urea in 50 mM TEAB. The protein concentration was measured with the A280 ultra-violet absorbance protein assay.

2.3.3. Enzymatic digestion

An amount corresponding to 100 µg of proteins was transferred to a new micro centrifuge tube and diluted with 6 M urea in 50 mM TEAB to a final volume of 50 µL before addition of 2 µg of Lys-C, and digestion was performed for 2 h at 37 °C under 300 rpm agitation. The samples were diluted with 50 mM TEAB to a final volume of 300 µL, and 1 µg of trypsin was then added. Tryptic digestion was subsequently performed for 16 h digestion at 37 °C under 300 rpm agitation. Digestion was stopped by adding TFA until a pH of ≤ 3 was reached.

2.3.4. Peptide purification

Peptides were purified with C18 SPE cartridges. The cartridges were first washed with 1 mL acetonitrile/water (50/50, v/v) and equilibrated with 1 mL wash solvent (acetonitrile/water/TFA, 2/97.9/0.1, v/v/v). Samples were diluted 10 times with water, loaded onto the cartridges and washed with 1 mL of wash solvent. Peptides were eluted with 2 mL acetonitrile/water (80/20, v/v) and subsequently evaporated under a nitrogen flow at 40 °C. Peptide pellets were dissolved in 200 µL 0.1% formic acid and filtered (centrifugal filter unit 0.22 µm, Merck-Millipore, Carrigtwohill, Ireland).

2.4. Mass spectrometry

Solubilized peptides from a digest of 100 µg proteins were analyzed by high performance liquid chromatography coupled to high resolution

mass spectrometry using a Waters Acquity UPLC® (Waters, Milford, MA VS) coupled to a SYNAPT G2-S High Definition Mass Spectrometer (TOF) (Waters). Tryptic peptides were chromatographically separated on a ACQUITY UPLC Peptide BEH C18 column (300 Å, 1.7 µm, 2.1 mm X 150 mm) (Waters) protected by an ACQUITY UPLC Peptide BEH C18 VanGuard Precolumn (300 Å, 1.7 µm, 2.1 mm X 5 mm) (Waters). Mobile phase A was 0.1% formic acid (solvent A) and mobile phase B was 0.1% formic acid in acetonitrile (solvent B). The solvent gradient was set as follows: initial: 99% solvent A; 60 min: 60% solvent A; 61 min: 15% solvent A; 65 min: 99% solvent A. The injection volume was set at 5 µL, the flow rate at 0.200 mL/min. Peptides were ionized in positive electrospray (ESI +) with the following instrument settings: capillary voltage 0.80 kV, sample cone voltage: 80 V, source temperature 120 °C and desolvation temperature: 400 °C. Data acquisition was done in continuous full scan mode in MS^E mode (data independent acquisition) and the *m/z* range was 50–2,000. Fragmentation was done with collision induced dissociation, with a low trap collision energy for precursor ions of 4 eV and high collision energy ramping from 20 to 40 eV. Recording was done in resolution mode (20 k FWHM). Prior to analysis, the HRMS was calibrated with sodium formate. Leucine enkephalin (200 pg/µL, mass 556.2771 Da, Waters) was used as lock mass and was acquired (scan time of 0.2 s) with an interval of 30 s, but not yet applied for mass correction. Mass correction using lock mass acquisitions is intrinsically imbedded in the ProteinLynx data analysis algorithm.

2.5. Data analysis

Raw HRMS data were analyzed with ProteinLynx Global Server™ (PLGS, Waters). Raw data were first automatically mass corrected using the lock mass recordings. Identified peaks were searched against a database containing the hazelnut proteome stored in Uniprot.org (*Corylus avellana*, 474 entries), added with the 192 reviewed human keratin entries from Uniprot.org, the cationic bovine trypsin sequence, the Lysyl endopeptidase sequence from *Lysobacter enzymogenes* and the possible protein contaminants from cacao and bacteria from Nu-trish® BY-Mild (*Bifidobacterium animalis* subsp. *lactis*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*). Search parameters were as follows: trypsin was chosen as the primary digestion enzyme and carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues, conversion of *N*-terminal glutamine to pyroglutamate, and deamidation of asparagine and glutamine were set as variable modifications and 3 missed cleavages were allowed.

3. Results and discussion

3.1. Test material and hazelnut origin

Our main goal was to identify peptides originating from hazelnut proteins that are robust towards food processing and could thus serve as universal/excellent biomarkers for the detection of hazelnut in food. Therefore, we extracted peptide profiles from different hazelnut-containing food matrices that underwent various processing techniques. Through comparison of these profiles, we expected to identify peptides robust to processing. To obtain these peptide profiles, food matrices needed first to be selected. However, when purchasing such materials from supermarkets, one can never be sure about the raw material that was used, recipes, food processing conditions, and/or putative cross contamination. Therefore, test material was produced in-house using modern food processing techniques and equipment. The four well-controlled processing conditions were heating of the ingredient, induction of the Maillard reaction, incurring hazelnut in a low pH environment or in a fatty rich environment. This generated the following test materials: roasted hazelnuts, caramelized hazelnuts, fermented hazelnut milk and chocolate incurred with hazelnut, and raw hazelnuts

as a non-processed matrix.

A universal detection method for hazelnut implicates that this method should be able to detect hazelnut irrespective of the processing technique the food has undergone, but also irrespective of the origin of hazelnut. However, numerous hazelnut cultivars exist and different cultivars may have different protein profiles and, subsequently, different peptide profiles. Indeed, Garino and collaborators have recently shown that gene transcript levels of known allergenic proteins Cor a 8 and Cor a 11 in hazelnut differ between cultivars and also between year of harvest (Garino et al., 2013). To counter this issue at least partially, all test matrices were produced in “duplo” using hazelnut from two different origins. Based on worldwide production, 70% of all hazelnuts come from Turkey and the second largest producer, counting for 12% of worldwide production, is Italy (Özenç & Özenç, 2015). Hazelnuts from these two origins were therefore used for the production of test material.

3.2. Protein selection

The first step in selecting suitable biomarker peptides is the *in silico* selection in the UniProt database of the proteins from which peptides can be derived. In this study, we initially focused on all known hazelnut proteins, not only on proteins known to cause allergic reactions as this could lead to excluding potentially good biomarkers from high abundant proteins. The full UniProt database for *Corylus avellana* containing 474 entries was therefore used. This is a relatively low amount of entries, considering the 28,255 transcript contigs – of which 93.3% encode proteins – in the hazelnut genome (Rowley et al., 2012). However, hazelnut kernels consist of about 15% proteins, of which 87% are seed storage proteins (Bonvehí, 1995). Of these seed storage proteins in hazelnut, the globulins Cor a 9 (11S globulin, primary accession number Q8W1C2) and Cor a 11 (7S globulin, Q8S4P9), and the albumin Cor a 14 (2S albumin, D0PWG2) are the major ones. Given their high abundance, these proteins were expected to provide the most abundant peptides. Indeed, the high abundance of these proteins was evident after data analysis as identified peptides mainly came from these three proteins in all conditions tested and in hazelnuts of both origins (Table S1). A fourth protein that was also always identified is the lipid transfer protein Cor a 8 (Q9ATH2). Therefore, while present in a lower abundance in the hazelnut kernel (Bonvehí, 1995), this protein was also kept for further peptide selection. All four proteins are known as allergenic proteins and have been used for developing LC-MS/MS methods for the detection of hazelnut in food (Ansari et al., 2012; Bignardi et al., 2013; Costa et al., 2016; Cucu et al., 2012; Heick et al., 2011; Pilolli et al., 2017; Sealey-Voyksner et al., 2016).

3.3. Peptide selection

For LC-MS/MS analysis, peptides are required. Usually, trypsin is the protease of choice because it cuts peptide bonds C-terminal to lysine and arginine with high specificity (Hustoft et al., 2012). As such, trypsin also yields basic peptides of a suitable length (on average 10–20 amino acids) and mass for LC-MS/MS analyses. For each of the hazelnut matrices that underwent the different processing techniques, tryptic peptide profiles were generated upon analysis of the HRMS data. Peptides were considered identified if they were detected in all three biological repeats and in extracts of both origins with high confidentiality (score 2 in ProteinLynx, as analysed with ProteinLynx GlobalSERVER v3.0.3 (Waters Corporation). Parallel to this, theoretical tryptic peptide profiles for the four identified proteins were generated by *in silico* tryptic digestion using PeptideMass (Wilkins et al., 1997), avoiding any missed cleavages and only considering peptides longer than 6 amino acids to ensure specificity. HRMS results from raw hazelnut extracts were compared with this list of peptides. We were able to detect 21 out of the possible 24 peptides for Cor a 9, 19 out of 24 for Cor a 11, 5 out of 7 for Cor a 14 and 7 out of 9 potential peptides for Cor a 8 (Table S1).

3.3.1. Specificity to hazelnut

As the peptide biomarkers for detection of hazelnut in food should only mark the presence of hazelnut proteins, the specificity for hazelnut proteins of each peptide was first checked by performing a Basic Local Alignment Search Tool (BLAST) analysis with the peptide sequence against the full UniProt database. Peptides were considered to be specific to hazelnut only if they showed full identity to only *Corylus avellana*, and not to any other organism. Implementing this criterion resulted in the exclusion of 2 out of the 21 identified peptides for Cor a 9, among which the peptide ³⁶⁴Trp-Arg³⁷¹, used as a biomarker for detection of hazelnut in previous studies (Ansari et al., 2012; Costa, Ansari, Mafra, Oliveira, & Baumgartner, 2014; Gu et al., 2018), 2 out of 19 identified peptides from Cor a 11, and 2 peptides out of the 7 identified peptides from Cor a 8. All identified peptides from Cor a 14 were retained (Table S1).

3.3.2. Amino acid composition

In a second selection step, peptides containing certain amino acids such as methionine, cysteine and *N*-terminal glutamine were excluded. Indeed, we hypothesized that quantification of hazelnut in food products should be reflected by the intensity of the peptide biomarkers used. Thus, these peptides should be ideally present in one form with thus one particular mass. As amino acid modifications alter the mass of peptides, these may result in inaccurate quantification as such modifications can alter both the retention time of the affected peptide as well as precursor and fragment ion masses (Fig. 1B). Thus, we attempted to avoid peptides that were prone to modifications. For example, methionines in the peptide ¹⁰⁵Gln-Arg¹¹⁵ from Cor a 14 can have multiple oxidation states (each with a mass shift of + 15.995 Da). Indeed, we detected multiple forms of this peptide in all matrices (Fig. 1A). Moreover, these results show that the intensities of the peptides with such different oxidation states vary depending on the processing techniques applied, implying that quantification cannot be only based on the intensity of the non-oxidized peptide. Thus, peptides containing methionine were therefore excluded.

Along this line, we also decided to exclude cysteine-containing peptides. Although cysteines are modified by carbamidomethylation, there is a possibility that, depending on the food matrix, some cysteines are not fully reduced and modified. In addition, peptides starting with glutamine were also excluded due to the possibility of this *N*-terminal glutamine to convert to pyroglutamic acid (mass shift of − 17.027 Da) (Neta, Pu, Kilpatrick, Yang, & Stein, 2007). In fact, pyroglutamic acid was also shown to hamper peptide fragmentation, which might negatively influence the overall performance of the LC-MS/MS detection method (Godugu, Neta, Simón-Manso, & Stein, 2010). The peptide ⁴¹⁸Gln-Lys⁴³² for example, cited in several studies as a biomarker (Ansari et al., 2012; Costa et al., 2014; Heick et al., 2011; Pilolli et al., 2018; Planque et al., 2017, 2019), was therefore excluded. Furthermore, we selected against KP or RP motifs in peptides as trypsin might cleave such motifs, albeit with a lower efficiency. Peptides carrying other types of missed cleavages (KK, KR, RR or RK at the beginning or end of their sequence) were also excluded. The sequence motif NG was also avoided, based on the findings that the asparagine (N) in such a motif is susceptible to deamination (Jia & Sun, 2017). Finally, peptides containing the *N*-terminus of a protein were also excluded as those peptides often comprise signal peptides, which are cleaved off once the protein has reached its proper cellular location. This results in both the presence and the absence of these sequences in the extracted proteins, again leading to biased quantification. Also, *N*-terminal acetylation (Ree, Varland, & Arnesen, 2018) and excision of the *N*-terminal initiator methionine (iMet) residue from nascent peptide chains (Jonckheere, Fijałkowska, & Van Damme, 2018) can influence the detection of these peptides, thus making them unsuitable as biomarkers.

Applying these exclusion criteria withheld 10 peptides for Cor a 11, 7 for Cor a 9, 0 for Cor a 14 and 2 for Cor a 8 (Table S1). Hence, these criteria exclude a large number of identified peptides, resulting in the

complete loss of Cor a 14, as well as some peptides with high intensities that could, if withheld, strengthen the power of the detection method. Indeed, an LC-MS/MS diagnostic method is preferably based on multiple peptides from multiple proteins of the ingredient. A solution to keep more proteins and peptides for final selection is to chemically induce protein modifications during sample preparation. For example, one could use hydrogen peroxide to oxidize methionines. However, such chemical modifications are only useful if they completely and specifically convert the targeted amino acids.

3.3.3. Selection based on protein specific properties and observations

Additional selection was done based on protein-specific properties and experimental observations. Each protein was surveyed for sequence-specific knowledge, focusing on sequence conflicts in the database, known isoforms and known natural variants. To ensure the robustness of the method, there should be no variations present in the selected peptides. For example, as Cor a 14 has two known isoforms (⁴⁸Arg ↔ ⁴⁸Ser) (Pfeifer et al., 2015), the tryptic peptide ⁴⁴Gly-Arg⁵³ was excluded. As another example, Cor a 11 has two known potential glycosylation sites (Asn⁸⁵ and Asn³⁰¹). Due to the uncertainty of these asparagine residues being glycosylated, peptides containing them were excluded (Lauer et al., 2004). Information directly extracted from the experimental data was also used for peptide selection. For example, peptides present in a sequence stretch prone to missed cleavages were excluded. The sequence ³⁹⁵Gly-Arg⁴¹⁹ in Cor a 11 is expected to yield only two fully tryptic peptides larger than 6 amino acids: ³⁹⁵Gly-Arg⁴⁰³ and ⁴⁰⁷Glu-Arg⁴¹⁵. However, after HRMS analysis, various peptides were observed containing missed cleavages (Fig. 2). Again, quantification of hazelnut in food cannot be directly based on the intensity of the two fully tryptic peptides, as their true intensities are the sum of their intensities and those of all peptides that embed them. Such peptides are therefore also excluded. For example, the peptides ⁴⁶⁴Ala-Arg⁴⁷⁸ (Ansari et al., 2012; Costa et al., 2014; Heick et al., 2011; Pilolli et al., 2017, 2018; Planque et al., 2017, 2019) and ⁴⁰⁷Glu-Arg⁴¹⁵ (Ansari et al., 2012; Costa et al., 2014) although often used as biomarkers in several studies, lie within a sequence stretch prone to missed cleavages. Since there is a known potential isoform (although partially sequenced, and not yet registered) for Cor a 9 (Nitride et al., 2013), the peptides ³⁵¹Ile-Arg³⁶³, ⁴⁴⁶Thr-Arg⁴⁵⁸ and ⁴⁶⁴Ala-Arg⁴⁷⁸ should also be excluded from selection since they lie in a non-conserved sequence region. These three peptides are also often cited as biomarker peptides for the detection of hazelnut (Ansari et al., 2012; Costa et al., 2016; Heick et al., 2011; Pilolli et al., 2017, 2018; Planque et al., 2017, 2019; Sealey-Voyksner et al., 2016). Applying these criteria led to the exclusion of another 2 peptides from Cor a 11 and 3 peptides from Cor a 9 (Table S1). Applying these selection steps eventually resulted in 2 potential good biomarkers for Cor a 8, 4 for Cor a 9 and 8 for Cor a 11.

3.3.4. Selection based on food process robustness

Further selection of peptides was done by comparing peptide intensities obtained in the five food matrices, and thus, controlling the peptide robustness towards the different food processing techniques. As already emphasized, peptide biomarkers that could be used for detection of hazelnut in food have to be robust to food processing, and thus, the amino acid sequence of this peptide should remain stable during the various processing steps. Also, the intensity of the peptides should be comparable under different circumstances in order to consider the peptides as good biomarkers. Sensitivity to modifications introduced by food processing can vary depending on the protein and on the peptides contained in the protein.

3.3.4.1. Robustness towards roasting and caramelization. In case of roasted and caramelized nuts, peptide intensities can be directly compared as no proteins other than hazelnut proteins are present in the matrix. From each extract, 100 µg of proteins were digested and peptide intensities were compared to the intensities observed for

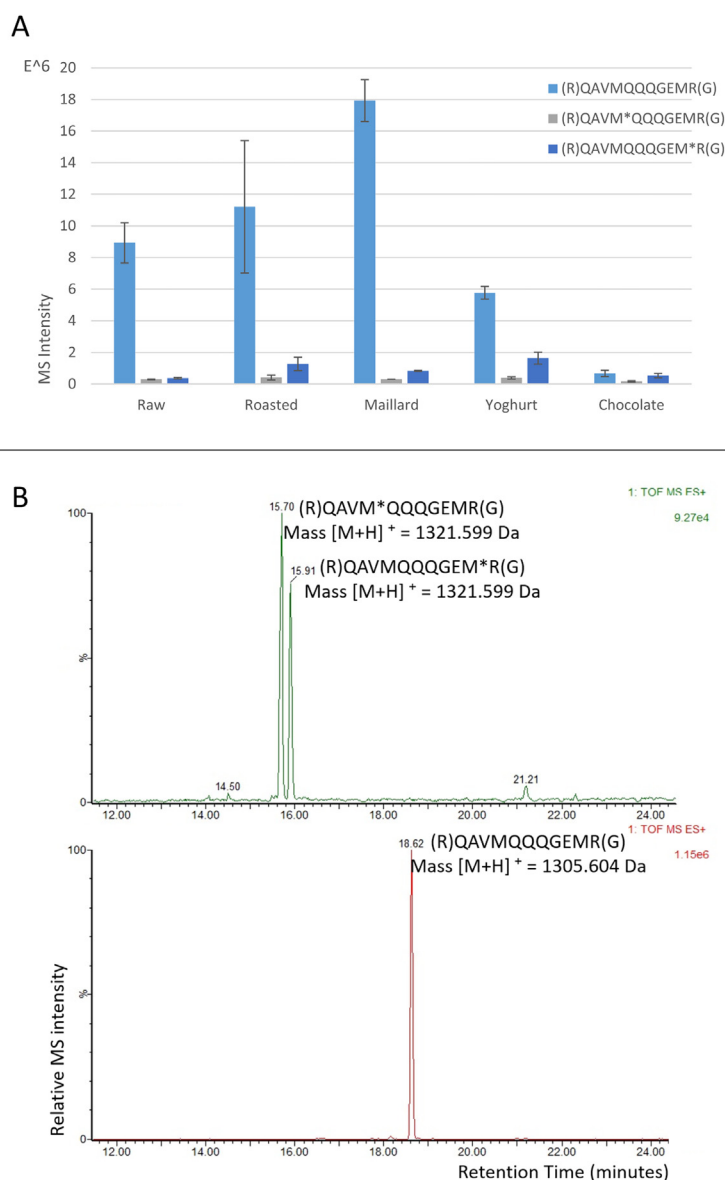


Fig. 1. Methionine oxidation variation. A) MS intensities of the peptide (R)QAVMQQQGEMR(G) from Cor a 14 (D0PWG2) showing different methionine oxidation states after different processing techniques. MS signal intensities are depicted for non-processed hazelnut, roasted hazelnuts, hazelnuts that have undergone induction of Maillard reaction, hazelnuts incurred in a low pH environment (yoghurt) and hazelnuts incurred in a fatty rich environment (chocolate). In each matrix, the intensities of three forms of the peptide are given. In light blue, the native (non-oxidized) peptide, in grey the peptide where the first methionine oxidized (M*) and in dark blue the peptide where the second methionine is oxidized (M*). Results are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. B) Selected ion chromatograms of the three variants (oxidized and reduced) of the peptide (R)QAVMQQQGEMR(G) identified in roasted hazelnuts, which illustrate the difference in retention time of the different forms. Masses are shown for the single charged peptide ion $[M + H]^+$, peak intensities are shown as relative intensity with 100% = peak area of highest peak for each given mass (9.27e⁴ for $MH^+ = 1321.599$ Da, 1.15e⁶ for $MH^+ = 1305.604$ Da). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

peptides generated from the raw sample. These peptide intensities were thus used to select biomarkers robust to processing. Peptide intensities from chocolate and yogurt carry some uncertainty as will be explained later, and will therefore only serve as an indication to the processing

robustness of the selected biomarkers.

At the end of the selection steps, only 2 (⁸¹Gly-Lys⁹⁶ and ⁵²Ala-Arg⁶³) out of the 8 peptides initially identified for Cor a 8 were retained. The peptide ⁸¹Gly-Lys⁹⁶ proved to be sensitive towards

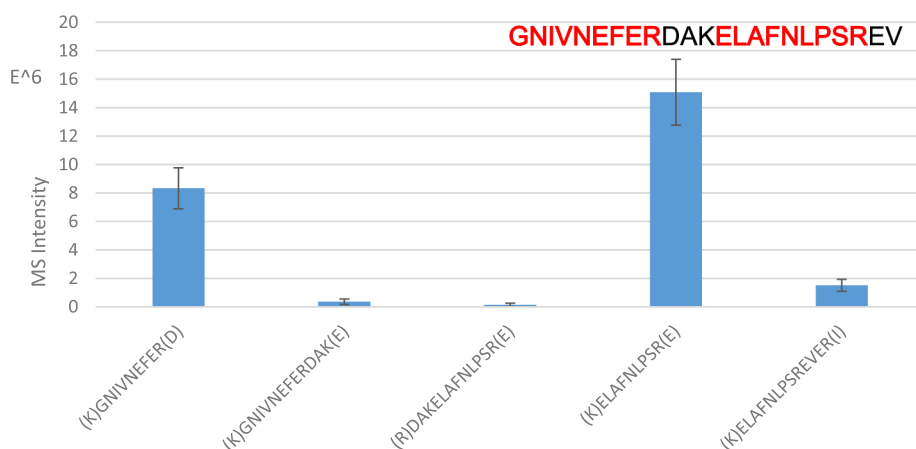


Fig. 2. Comparison of MS intensities of different peptides in a missed cleavage sensitive sequence stretch GNIVNEFERDAKELAFNLPSREVER from Cor a 11 (Q8S4P9). Tryptic peptides without missed cleavage are depicted in red (GNIVNEFER and ELAFNLPSR). MS signal intensities are shown for raw hazelnut and results are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

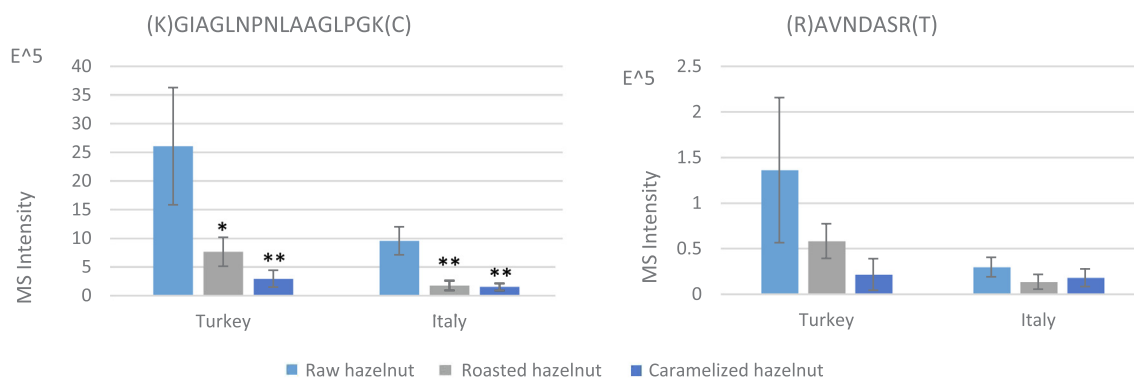


Fig. 3. Comparison of MS intensities of tryptic peptides that were retained after several selection criteria from Cor a 8 (Q9ATH2) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reaction (Caramelized hazelnut). MS signal intensities are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut. (* = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$).

processing, showing significantly lower peptide intensity in both processed matrices when compared to the intensity in raw hazelnut, and this in nuts of both origins (Fig. 3). The peptide $^{52}\text{Ala-Arg}^{63}$, although showing an apparent decrease in intensity in processed matrices when compared to the raw matrix in hazelnut from Turkey, this decrease was not significant. No significant difference in intensity between processed and raw hazelnut matrices in hazelnuts from Italy was observed.

Cor a 11 holds both food processing sensitive as well as robust peptides. Three peptides were robust towards roasting and the Maillard reaction: $^{92}\text{Leu-Arg}^{100}$, $^{175}\text{Ile-Arg}^{201}$ and $^{241}\text{Ala-Arg}^{251}$. For these peptides, no significant difference in intensity was observed between the raw, roasted and caramelized hazelnut matrices (Fig. 4). Of these, $^{92}\text{Leu-Arg}^{100}$ and $^{175}\text{Ile-Arg}^{201}$ are the most intense peptides, and considering the intensities in the chocolate and fermented nut milk matrix, form the most promising biomarker candidates for this protein. The same trend was observed in samples from both nut origins. Interestingly, the peptide $^{202}\text{Ala-Lys}^{213}$, often cited as a biomarker (Ansari et al., 2012; Costa et al., 2014), was excluded from selection because of its sensitivity towards roasting and caramelizing hazelnut.

Selected peptides for Cor a 9 generally showed good robustness towards both roasting and caramelization. No negative significant differences in intensity were observed in peptide intensity between the three matrices (Fig. 5). Some peptides display a higher peak intensity in processed matrices when compared to the raw hazelnut, but this is not a problem for the peptide to serve as a biomarker. Cor a 9 shows overall the highest robustness across all processing techniques when compared to the other proteins. This is in line with the intrinsic properties of this protein being a legumine, and therefore very stable against thermal processes. The secondary structures of these proteins largely remain unchanged and have only minor modifications, even after heating at 94 °C (Costa et al., 2016). This while Cor a 11, a vicilin is less thermostable, showing already structural changes at 70 °C, and is known to form gellish aggregates at high temperatures, making it harder to extract the proteins and thus more difficult to detect the peptides (Sathe et al., 2005).

3.3.4.2. Robustness towards mixing into a fatty rich environment (chocolate matrix) and in a low pH background (fermented nut milk). When yoghurt and chocolate samples are considered, intrinsic to the recipe, non-hazelnut proteins are also present in the matrix. These proteins are co-extracted during sample preparation, digested, and their peptides are also analyzed by HRMS. Thus, in these matrices, only a fraction of the injected peptides comprises hazelnut peptides. This fraction can be theoretically calculated in the case of chocolate. The produced hazelnut incurred chocolate contains 491,099 ppm total chocolate proteins per kg chocolate, 40,708 ppm or 39,860 ppm total hazelnut proteins for Turkey or Italy hazelnut origin, respectively.

Purified proteins extracted from this matrix thus comprises 55.6% cacao proteins and 44.4% hazelnut proteins, assuming cacao proteins and hazelnut proteins are extracted with the same efficiency. Peptide intensities of hazelnut peptides are therefore expected to be around 55.6% lower when compared to the peptide intensities from the same amount of extracted proteins from raw hazelnut. However, when comparing intensities from raw extracts and chocolate extracts, none of the identified peptides showed intensities in this range. On the contrary, intensities were only a fourth of the intensities in the raw sample, or even lower (Fig. S1a). An explanation could be that cacao proteins are more efficiently extracted from the chocolate matrix, enlarging the ratio cacao/hazelnut proteins in the extract. Excluding the possibility that the lower intensities are due to wrong estimation of protein content in the chocolate (indeed, polyphenols residues can influence the nanodrop results), or that protein loss during the desalting step (using NAP-10 column) is the reason for this lower abundance, a different extraction protocol, adapted from Planque and collaborators (Planque et al., 2016) was used, where no correction for protein abundance or a desalting step is done. This resulted again in lower intensities than expected, indicating that it might be hard to detect these peptides in fatty rich environments (Fig. S1b). For the yoghurt, the theoretical ratio of hazelnut proteins to bacterial proteins is not so straightforward to calculate, due to the uncertainty on the bacterial growth rate in the fermented nut milk, and thus the uncertainty on the bacterial protein content. As could be expected, lower peptide intensities compared to the intensities from raw hazelnut were identified, but a comparison of intensities from these two matrices can thus indicate robustness to these food processing techniques, but these results should be taken with caution and are only indicative.

3.3.5. Selection based on robustness to origin variation

Finally, all peptides were checked on similarity of performance in the two hazelnut origins. Robustness towards genetic variations and cultivars strengthens the power of the detection method. Although in general, the same peptides showing the most promising results to serve as good biomarkers were identified in both types of hazelnut, there were some differences in peptide intensity detection. Interestingly, peptides from Cor a 9 usually display a higher intensity in Turkish hazelnuts when compared to Italian hazelnuts (Fig. S2). This trend was visible in all matrices but was more pronounced in roasted hazelnuts. This implies that the protein Cor a 9 was relatively more abundant in the Turkish hazelnuts. Also striking was that peptide intensities from roasted Turkish hazelnut were for the three proteins overall significantly higher compared to peptide intensities from roasted Italian hazelnuts, even when there was no significant difference in intensities from raw nuts (Fig. S2). Peptides from Turkish hazelnuts thus seemed to be more robust towards roasting of the nuts. This was not apparent for

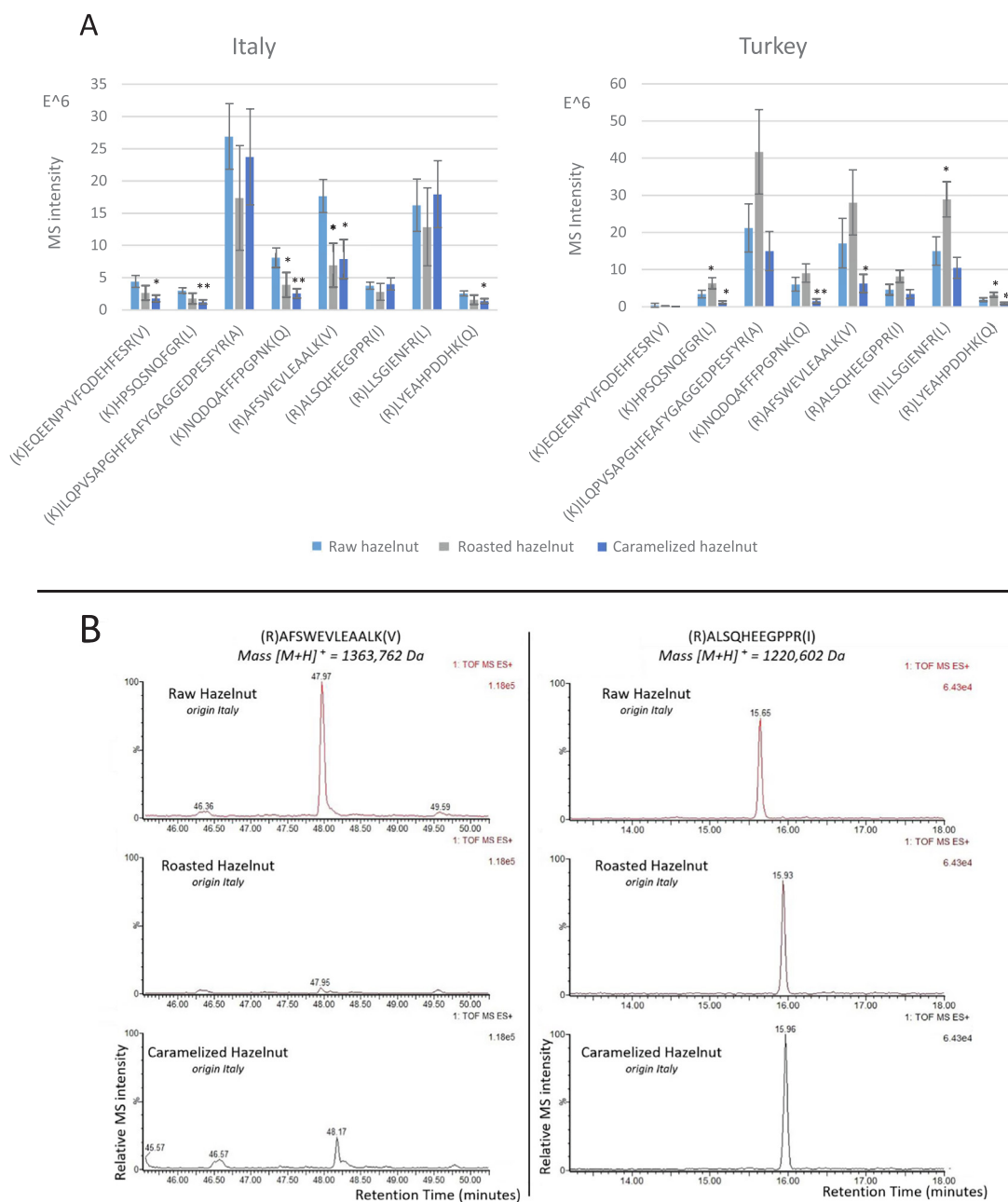


Fig. 4. Comparison of tryptic peptides that were retained after several selection criteria from Cor a 11 (Q8S4P9) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reaction (Caramelized hazelnut)). A) Comparison of MS signal intensities. Intensities are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut (* = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$). B) Exemplary selected ion chromatograms of a peptide showing significant difference in intensity in the different test materials ((R)AFSWEVLEAALK(V)), and a peptide showing no significant difference in intensity in the different test materials ((R)ALSQHEEGPPR(I)). Masses are shown for the single charged peptide ion $[M + H]^+$, peak intensities are shown as relative intensity with 100% = peak area of highest peak for the given mass in the three test materials (1.18e⁵ for (R)AFSWEVLEAALK(V), 6.43e⁴ for (R)ALSQHEEGPPR(I)).

caramelization of nuts.

4. Conclusions

Based on the above selection steps we were able to select 8 peptides from 3 major hazelnut proteins that are suitable to function as biomarkers for detection of hazelnut in food, among which, 5 have not been published yet (Table 1).

These peptides were selected to perform good as universally

applicable biomarkers, robust against multiple processing techniques and matrix effects, and insensitive to post-translational modification of amino acids occurring during production of food or during sample preparation for analysis. Of these 8 peptides, some seem to be more preferred than others. Indeed, in general, the three selected peptides from Cor a 11 are the most suitable for the development of an LC-MS/MS method for the detection of hazelnut in food in terms of similarity in peptide intensity between different matrices and different hazelnut origins. However, peptides from Cor a 9 generally show a higher

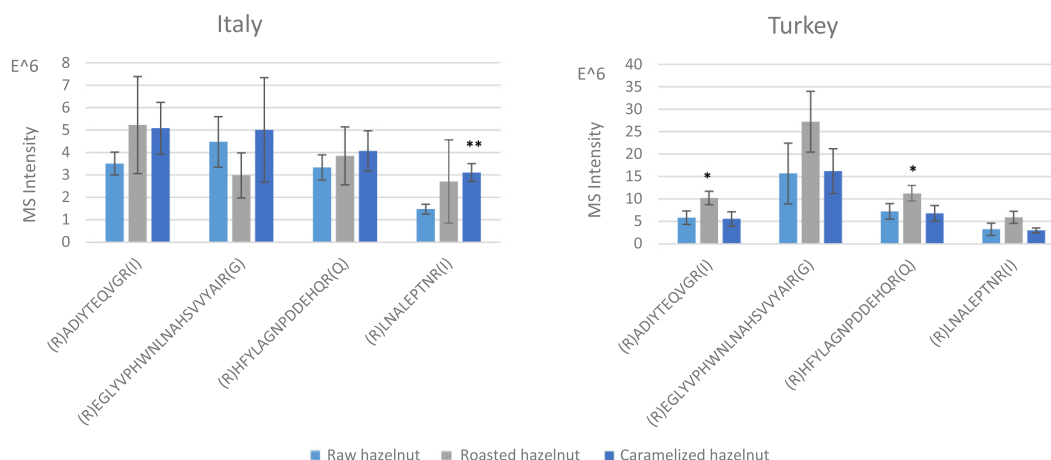


Fig. 5. Comparison of absolute MS intensities of tryptic peptides that were retained after several selection criteria from Cor a 9 (Q8W1C2) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reactions (Caramelized hazelnut). MS signal intensities are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut (* = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$).

Table 1

Final selection of potential good biomarkers for the detection of hazelnut in food. Bold written peptides have not been published yet.

Protein	Peptide	Position	Publications
Cor a 8 (Q9ATH2)	AVNDASR	57–63	
Cor a 9 (Q8W1C2)	LNALPTNR	43–51	Planque et al. (2017); Sealey-Voyksner et al. (2016)
	HFYLAGNPDEHQ(R)	195–208	
	ADIYTEQVGR	341–350	Bignardi, Elviri, Penna, Careri, and Mangia (2010); Bignardi et al. (2013); Gu et al. (2018); Heick et al. (2011); Pilolli et al. (2017, 2018); Planque et al. (2019, 2017)
	EGLYVPHWNLNAHSVVYAIR	377–396	
Cor a 11 (Q8S4P9)	LLSGIENFR	92–100	Ansari et al. (2012); Costa et al. (2014)
	ILQPVSAPGHFEAFYAGGE DPESFYR	175–201	
	ALSQHEEGPPR	241–251	

absolute intensity, and thus, this protein is expected to be detectable in lower concentration of hazelnut in food compared to Cor a 11. Indeed, when developing highly sensitive tandem mass spectrometry detection methods, clearly those peptides giving the highest signals are preferred, this to reach the lowest limit of detection, which will be determined during validation of the method. Secondly, an LC-MS/MS method preferably contains multiple peptides per protein, and multiple proteins of the allergenic ingredient to strengthen reliability of the method. This implicates that the peptide from Cor a 8 on its own would not be ideal as biomarker being the only suitable peptide from this protein, and showing a low absolute intensity. Combining this peptide with peptides from Cor a 9 and Cor a 11 for developing an LC-MS/MS method will higher the confidence of the method. Because of the lack of research into hazelnut protein isoforms, and the small database of known hazelnut proteins, the list of selected peptides should be updated once more information is accessible in literature. Finally, the need of checking the performance of these peptides in different food matrices remains important.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the Belgian Federal Public Service of Health, Food Chain Safety and Environment [RT 15/10 ALLERSENS].

Appendix A. Supplementary data

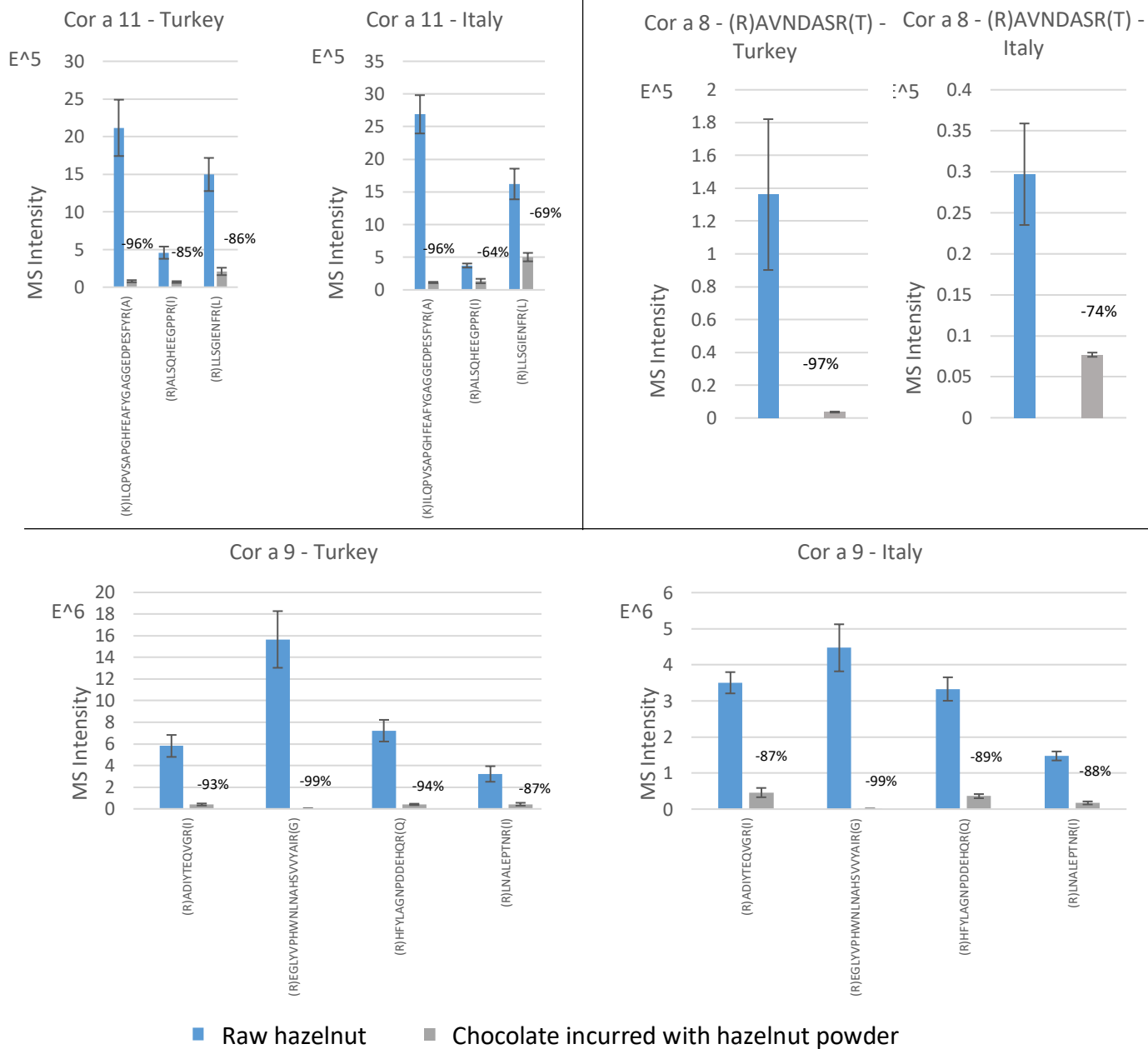
Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2019.125679>.

References

- Ansari, P., Stoppacher, N., & Baumgartner, S. (2012). Marker peptide selection for the determination of hazelnut by LC-MS/MS and occurrence in other nuts. *Analytical and Bioanalytical Chemistry*, 402(8), 2607–2615. <https://doi.org/10.1007/s00216-011-5218-6>.
- Bignardi, C., Elviri, L., Penna, A., Careri, M., & Mangia, A. (2010). Particle-packed column versus silica-based monolithic column for liquid chromatography–electrospray-linear ion trap-tandem mass spectrometry multiallergen trace analysis in foods. *Journal of Chromatography A*, 1217(48), 7579–7585. <https://doi.org/10.1016/j.chroma.2010.10.037>.
- Bignardi, C., Mattarozzi, M., Penna, A., Sidoli, S., Elviri, L., Careri, M., & Mangia, A. (2013). A rapid size-exclusion solid-phase extraction step for enhanced sensitivity in multi-allergen determination in dark chocolate and biscuits by liquid chromatography-tandem mass spectrometry. *Food Analytical Methods*, 6(4), 1144–1152. <https://doi.org/10.1007/s12161-012-9521-4>.
- Bonvehí, J. S. (1995). A chemical study of the protein fractions of Tarragona hazelnuts (*Corylus avellana*). *Zeitschrift Für Lebensmittel-Untersuchung Und -Forschung*, 201(4), 371–374.
- Costa, J., Ansari, P., Mafra, I., Oliveira, M. B. P. P., & Baumgartner, S. (2014). Assessing hazelnut allergens by protein- and DNA-based approaches: LC-MS/MS, ELISA and real-time PCR. *Analytical and Bioanalytical Chemistry*, 406(11), 2581–2590. <https://doi.org/10.1007/s00216-014-7679-x>.
- Costa, J., Mafra, I., Carrapatoso, I., & Oliveira, M. B. P. P. (2016). Hazelnut allergens: Molecular characterization, detection, and clinical relevance. *Critical Reviews in Food Science and Nutrition*, 56(15), 2579–2605. <https://doi.org/10.1080/10408398.2013.826173>.
- Cucu, T., Meulenaer, B. D., & Devreese, B. (2012). MALDI-based identification of stable hazelnut protein derived tryptic marker peptides. *Food Additives & Contaminants: Part A*, 29(12), 1821–1831. <https://doi.org/10.1080/19440049.2012.719639>.
- Desjardins, P., Hansen, J. B., & Allen, M. (2009). Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *Journal of Visualized*

- Experiments: JoVE, (33). doi:10.3791/1610.
- Garino, C., Locatelli, M., Coisson, J. D., D'Andrea, M., Cereti, E., Travaglia, F., & Arlorio, M. (2013). Gene transcription analysis of hazelnut (*Corylus avellana* L.) allergens Cor a 1, Cor a 8 and Cor a 11: A comparative study. *International Journal of Food Science & Technology*, 48(6), 1208–1217. <https://doi.org/10.1111/ijfs.12079>.
- Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., & Gillard, N. (2019). Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *Journal of Chromatography A*, 1584, 115–125. <https://doi.org/10.1016/j.chroma.2018.11.036>.
- Godugu, B., Neta, P., Simón-Manso, Y., & Stein, S. E. (2010). Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions. *Journal of the American Society for Mass Spectrometry*, 21(7), 1169–1176. <https://doi.org/10.1016/j.jasms.2010.03.027>.
- Gu, S., Chen, N., Zhou, Y., Zhao, C., Zhan, L., Qu, L., & Ding, Y. (2018). A rapid solid-phase extraction combined with liquid chromatography-tandem mass spectrometry for simultaneous screening of multiple allergens in chocolates. *Food Control*, 84, 89–96. <https://doi.org/10.1016/j.foodcont.2017.07.033>.
- Heick, J., Fischer, M., Kerbach, S., Tamm, U., & Popping, B. (2011). Application of a liquid chromatography tandem mass spectrometry method for the simultaneous detection of seven allergenic foods in flour and bread and comparison of the method with commercially available ELISA test kits. *Journal of AOAC International*, 94(4), 1060–1068.
- Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsæet, L., Lundanes, E., & Greibrokk, T. (2012). A critical review of trypsin digestion for LC-MS based proteomics. *Integrative Proteomics*. <https://doi.org/10.5772/29326>.
- Jia, L., & Sun, Y. (2017). Protein asparagine deamidation prediction based on structures with machine learning methods. *PLOS ONE*, 12(7), e0181347. <https://doi.org/10.1371/journal.pone.0181347>.
- Jonckheere, V., Fijałkowska, D., & Van Damme, P. (2018). Omics assisted N-terminal proteoform and protein expression profiling on methionine aminopeptidase 1 (MetAP1) deletion. *Molecular & Cellular Proteomics*, 17(4), 694–708. <https://doi.org/10.1074/mcp.RA117.000360>.
- Lauer, I., Foetisch, K., Kolarich, D., Ballmer-Weber, B. K., Conti, A., Altmann, F., & Scheurer, S. (2004). Hazelnut (*Corylus avellana*) vicilin Cor a 11: Molecular characterization of a glycoprotein and its allergenic activity. *Biochemical Journal*, 383(2), 327–334. <https://doi.org/10.1042/BJ20041062>.
- Manea, I., Ailenei, E., & Deleanu, D. (2016). Overview of food allergy diagnosis. *Clujul Medical*, 89(1), 5–10. <https://doi.org/10.15386/cjmed-513>.
- Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Critical Reviews in Food Science and Nutrition*, 48(2), 177–184. <https://doi.org/10.1080/10408390701279749>.
- McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., & Allen, K. (2015). The prevalence of tree nut allergy: A systematic review. *Current Allergy and Asthma Reports*, 15(9), 54. <https://doi.org/10.1007/s11882-015-0555-8>.
- Monaci, L., & Visconti, A. (2010). Immunochemical and DNA-based methods in food allergen analysis and quality assurance perspectives. *Trends in Food Science & Technology*, 21(6), 272–283. <https://doi.org/10.1016/j.tifs.2010.02.003>.
- Neta, P., Pu, Q.-L., Kilpatrick, L., Yang, X., & Stein, S. E. (2007). Dehydration versus deamidation of N-terminal glutamine in collision-induced dissociation of protonated peptides. *Journal of the American Society for Mass Spectrometry*, 18(1), 27–36. <https://doi.org/10.1016/j.jasms.2006.08.016>.
- Nitride, C., Mamone, G., Picariello, G., Mills, C., Nocerino, R., Berni Canani, R., & Ferranti, P. (2013). Proteomic and immunological characterization of a new food allergen from hazelnut (*Corylus avellana*). *Journal of Proteomics*, 86, 16–26. <https://doi.org/10.1016/j.jprot.2013.05.001>.
- Özenç, N., & Özenç, D. B. (2015). Nut traits and nutritional composition of hazelnut (*Corylus avellana* L.) as influenced by zinc fertilization. *Journal of the Science of Food and Agriculture*, 95(9), 1956–1962. <https://doi.org/10.1002/jsfa.6911>.
- Pfeifer, S., Bublin, M., Dubiela, P., Hummel, K., Wortmann, J., Hofer, G., & Hoffmann-Sommergruber, K. (2015). Cor a 14, the allergenic 2S albumin from hazelnut, is highly thermostable and resistant to gastrointestinal digestion. *Molecular Nutrition & Food Research*, 59(10), 2077–2086. <https://doi.org/10.1002/mnfr.201500071>.
- Pilolli, R., De Angelis, E., & Monaci, L. (2017). Streamlining the analytical workflow for multiplex MS/MS allergen detection in processed foods. *Food Chemistry*, 221, 1747–1753. <https://doi.org/10.1016/j.foodchem.2016.10.110>.
- Pilolli, R., De Angelis, E., & Monaci, L. (2018). In house validation of a high resolution mass spectrometry Orbitrap-based method for multiple allergen detection in a processed model food. *Analytical and Bioanalytical Chemistry*, 410(22), 5653–5662. <https://doi.org/10.1007/s00216-018-0927-8>.
- Planque, M., Arnould, T., Delahaut, P., Renard, P., Dieu, M., & Gillard, N. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chemistry*, 274, 35–45. <https://doi.org/10.1016/j.foodchem.2018.08.095>.
- Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2016). Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs. *Journal of Chromatography. A*, 1464, 115–123. <https://doi.org/10.1016/j.chroma.2016.08.033>.
- Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2017). Liquid chromatography coupled to tandem mass spectrometry for detecting ten allergens in complex and incurred foodstuffs. *Journal of Chromatography. A*, 1530, 138–151. <https://doi.org/10.1016/j.chroma.2017.11.039>.
- Poms, R. E., Klein, C. L., & Anklam, E. (2004). Methods for allergen analysis in food: A review. *Food Additives and Contaminants*, 21(1), 1–31. <https://doi.org/10.1080/02652030310001620423>.
- Popping, B., & Diaz-Amigo, C. (2017). European regulations for labeling requirements for food allergens and substances causing intolerances: History and future. *Journal of AOAC International*. <https://doi.org/10.5740/jaoacint.17-0381>.
- Prado, M., Ortea, I., Vial, S., Rivas, J., Calo-Mata, P., & Barros-Velázquez, J. (2016). Advanced DNA- and protein-based methods for the detection and investigation of food allergens. *Critical Reviews in Food Science and Nutrition*, 56(15), 2511–2542. <https://doi.org/10.1080/10408398.2013.873767>.
- Ree, R., Varland, S., & Arnesen, T. (2018). Spotlight on protein N-terminal acetylation. *Experimental & Molecular Medicine*, 50(7), <https://doi.org/10.1038/s12276-018-0116-z>.
- Rowley, E. R., Fox, S. E., Bryant, D. W., Sullivan, C. M., Priest, H. D., Givan, S. A., & Mockler, T. C. (2012). Assembly and characterization of the European hazelnut 'jefferson' transcriptome. *Crop Science*, 52(6), 2679–2686. <https://doi.org/10.2135/cropsci2012.02.0065>.
- Sathe, S. K., Teuber, S. S., & Roux, K. H. (2005). Effects of food processing on the stability of food allergens. *Biotechnology Advances*, 23(6), 423–429. <https://doi.org/10.1016/j.biotechadv.2005.05.008>.
- Sealey-Voyksner, J., Zweigenbaum, J., & Voyksner, R. (2016). Discovery of highly conserved unique peanut and tree nut peptides by LC-MS/MS for multi-allergen detection. *Food Chemistry*, 194, 201–211. <https://doi.org/10.1016/j.foodchem.2015.07.043>.
- Wei, Y., Sathe, S. K., Teuber, S. S., & Roux, K. H. (2003). A Sensitive Sandwich ELISA for the Detection of Trace Amounts of Cashew (*Anacardium occidentale* L.) Nut in Foods. *Journal of Agricultural and Food Chemistry*, 51(11), 3215–3221. doi:10.1021/jf025977x.
- Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F., & Appel, R. D. (1997). Detailed peptide characterization using PEPTIDEMASS—a World-Wide-Web-accessible tool. *Electrophoresis*, 18(3–4), 403–408. <https://doi.org/10.1002/elps.1150180314>.

A



B

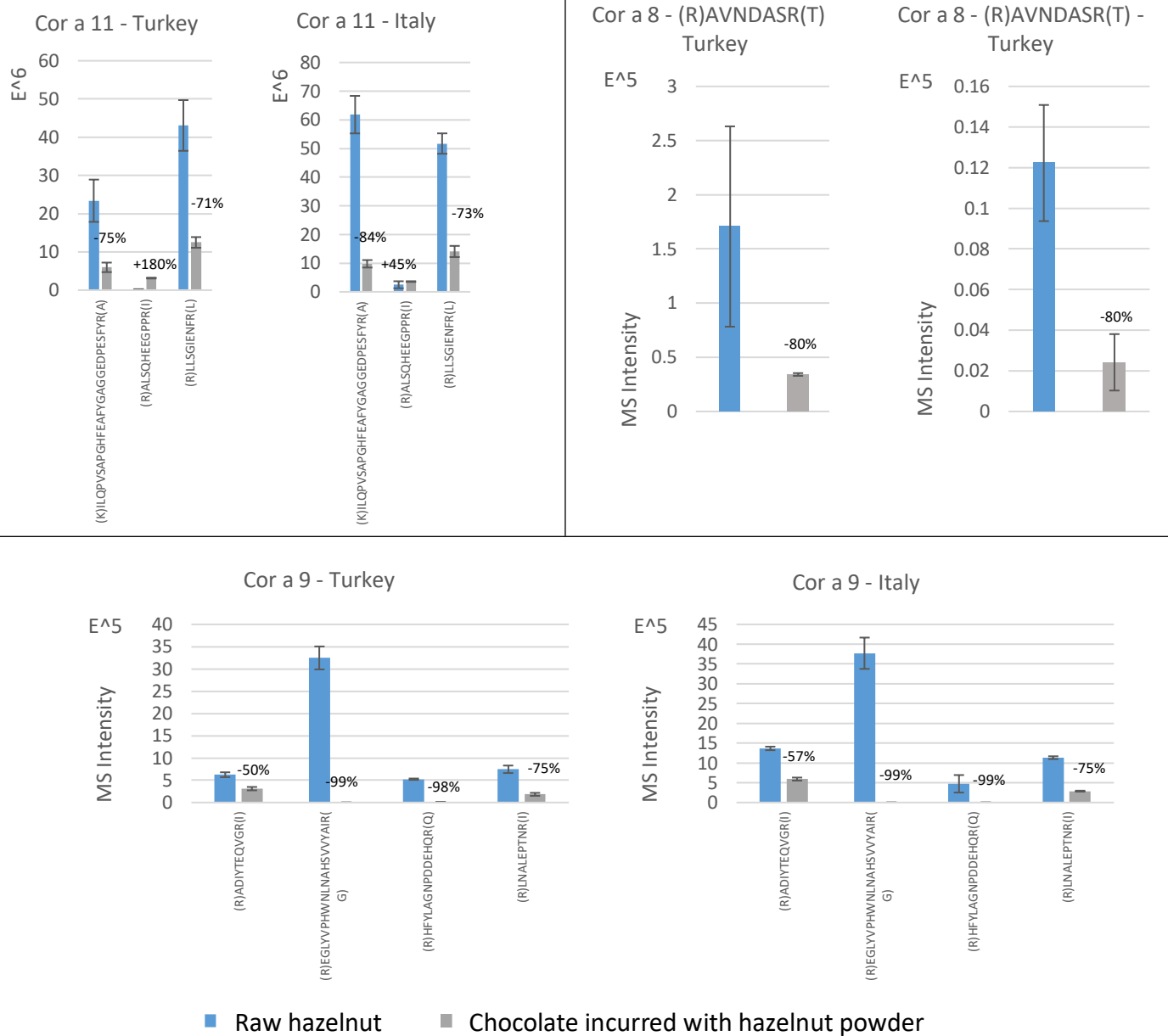


Figure S2: Comparison of absolute peptide MS intensities in raw hazelnuts and chocolate incurred with hazelnut powder from Turkish and Italian origin of tryptic peptides that were retained after several selection criteria from Cor a 11 (Q8S4P9), Cor a 8 (Q9ATH2) and Cor a 9 (Q8W1C2). A) intensities when the Allersens protocol (Gavage et al., 2019) was applied, B) intensities when the Allermass protocol (Planque et al., 2016) was applied. MS signal intensities are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. Intensity reduction between raw hazelnuts and chocolate incurred with hazelnut powder are depicted as percentage.

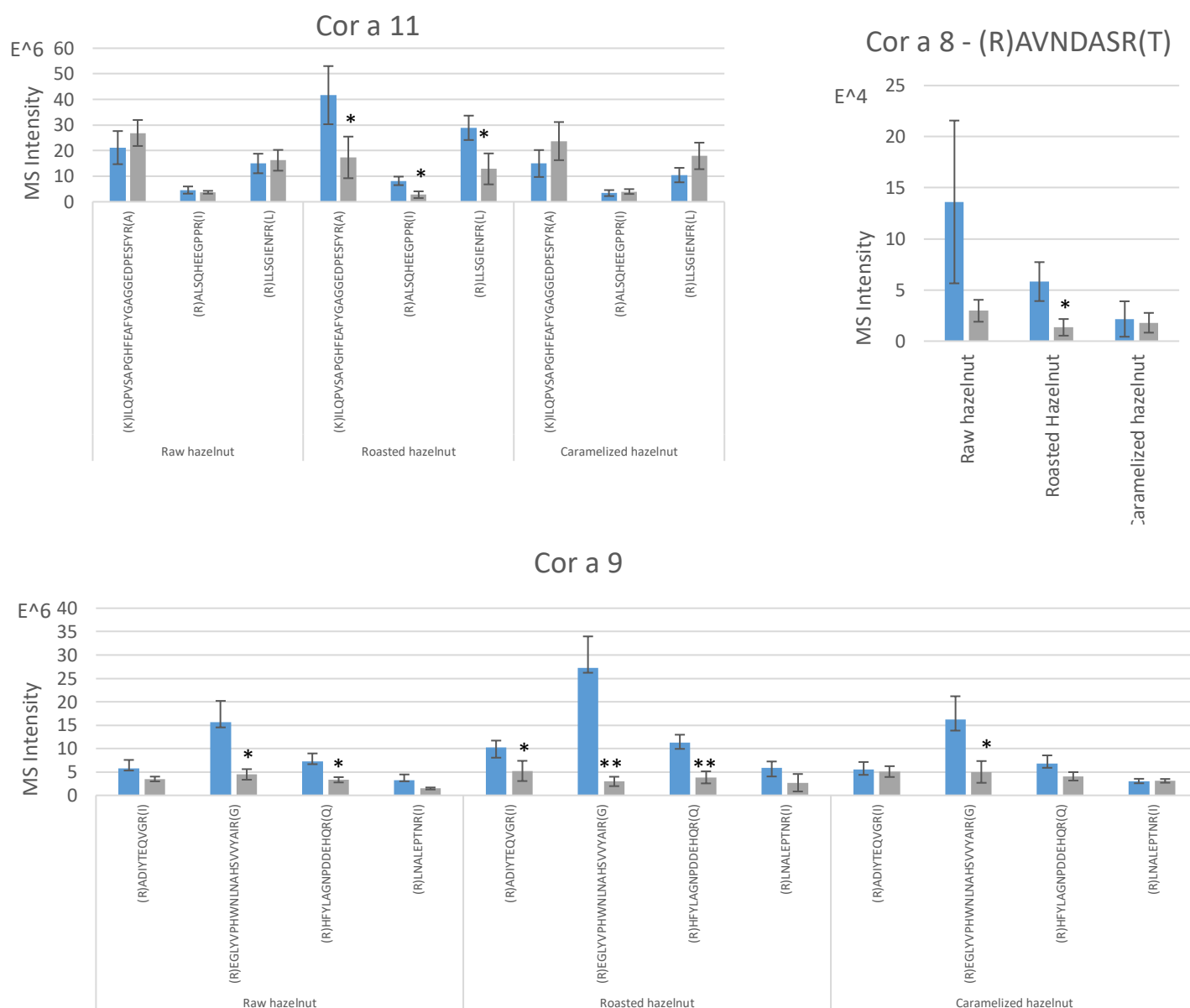


Figure S2: Comparison of absolute MS intensities in hazelnuts from Turkish and Italian origin of tryptic peptides that were retained after several selection criteria from Cor a 11 (Q8S4P9), Cor a 8 (Q9ATH2) and Cor a 9 (Q8W1C2) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reactions (Caramelized hazelnut). MS signal intensities are expressed as the means of the peptide MS signal intensity $\pm$ 1 SD for 3 biological replicates. Significance levels were calculated by Student's t-test, comparing values from Italian and Turkish hazelnut (*= $p < 0.05$; **= $p < 0.01$, ***= $p < 0.001$).

■ Hazelnut Turkish origin ■ Chocolate incurred with hazelnut powder

Protein	Peptides (<6AA)	Position	Criterium 1	Criterium 2	Criterium 3	Criterium 4	Criterium 5A	Criterium 5B	Criterium 5C	Criterium 5D
Cor a 11 (uniprot accession Q8S4P9)	QFDEQQR	21-27	Yes	Yes	RR motif at C-terminus of peptide, N-terminal Q	-	Yes	Yes	Yes	Yes
	QQEEGNSSEESYGK	42-55	No	Yes	N-terminal Q	-	-	-	-	-
	EQEENPYVFQDEHFESR	56-72	Yes	Yes	OK	-	Yes	Yes	Yes	No
	VQVLENFTK	80-88	Yes	Yes	KR motif at C-terminus of peptide	N can be glycosylated	Yes	No	Yes	No
	LLSGIENFR	92-100	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	LAILEANPHTFISPAHFDAE LVLFVAK	101-127	No	Yes	OK	-	-	-	-	-
	ATITMVR	130-136	Yes	Yes	Methionine	-	Yes	Yes	Yes	Yes
	ESFNVEHGDIIR	141-152	Yes	Yes	KR motif at N-terminus of peptide	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	Yes
	IPAGTPVYMINR	153-164	Yes	Yes	Methionine	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	Yes
	ILQPVSAPGHFEAFYGAGG EDPESFYR	175-201	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	AFSWEVLEAALK	202-213	Yes	Yes	OK	-	Yes	Yes	No	No
	VFGEQSK	222-228	No	No	OK	-	-	-	-	-
	ALSQHEEGPPR	241-251	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	IWPFGGESSGPINLLHK	252-268	No	Yes	OK	-	-	-	-	-
	HPSQSNQFGR	269-278	Yes	Yes	OK	-	Yes	No	Yes	No
	LYEAHPDDHK	279-288	Yes	Yes	OK	-	Yes	No	Yes	No
	QLQDLDMVSVFANITK	289-304	No	Yes	N-terminal Q, Methionines	N can be glycosylated	-	-	-	-
	GSMAGPYNSR	305-315	Yes	Yes	Methionine	-	Yes	Yes	Yes	Yes
	ISVVVEGEGFFEMACPHLS SSGSYQK	319-345	Yes	Yes	Methionine, cystein	-	Yes	No	No	Yes
	GVVVFAPAGHPVAVIASQ NNNLQVLCFEVNAHGNS R	353-388	Yes	Yes	RR motif at N-terminus of peptide, cystein	-	Yes	Yes	No	Yes
	GNIVNEFER	395-403	Yes	Yes	OK	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	No

	ELAFNLPSR	407-415	Yes	Yes	OK	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	Yes
	NQDQAFFFPQPNK	423-435	Yes	Yes	OK	-	Yes	No	No	No
	QQEEGGR	436-442	Yes	No	N-terminal Q	-	No	No	Yes	No
Cor a 14 (Uniprot accession D0PWG2)	TTITTVDVDEDIVNQQR	25-42	No	Yes	OK	Partially in small subunit with multiple known isoforms	-	-	-	-
	QQNLNQQR	54-62	Yes	Yes	N-terminal Q, Cystein	In small subunit with multiple known isoforms	Yes	Yes	Yes	Yes
	QQSQYGSYDGSNQQQQ QELEQCCQQLR	66-92	Yes	Yes	N-terminal Q, cysteins	Partially in small subunit with multiple known isoforms	Yes	Yes	Yes	Yes
	QAVMQQQGEMR	102-115	yes	Yes	N-terminal Q, Methionines	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	No
	EVMETAR	121-127	Yes	Yes	Methionine	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	No
	DLPNQCR	128-134	Yes	Yes	OK	In sequence stretch prone to missed cleavage	Yes	Yes	Yes	Yes
Cor a 8 (Uniprot accession Q9ATH2)	LVCAVLLCMMVAAPVAR	6-22	No	Yes	Cysteins, methionines, part of sequence lies within 20 first amino acids of the protein	-	-	-	-	-
	ASLTCPQIK	24-31	No	Yes	Cysteins	-	-	-	-	-
	GNLTPCVLYLK	32-42	Yes	Yes	Cysteins	-	Yes	Yes	No	No
	NGGVLPPSCCK	43-53	Yes	Yes	Cysteins, NG motif	-	Yes	Yes	Yes	Yes
	AVNDASR	57-63	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	QSACNCLK	69-76	Yes	No	Cysteins	-	Yes	Yes	Yes	Yes
	GIAGLNPNLAAGLPGK	81-98	Yes	yes	OK	-	No	No	No	No
	CGVNIPYK	97-104	Yes	No	Cysteins	-	Yes	Yes	Yes	Yes
Cor a 9 (Uniprot	ISPSTNCNNVK	105-115	Yes	Yes	Cysteins	-	Yes	No	Yes	Yes
	LILVSFSLCLLVLFNGCLGIN VGLR	4-28	No	Yes	Cystein, RR motif at C-terminus of peptide, NG motif, part of sequence	-	-	-	-	-

accession Q8W1C2)					lies within 20 first amino acids of the protein					
	YFGECNLDR	34-42	No	Yes	Cystein	-	-	-	-	-
	LNALEPTNR	43-51	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	IEAEACQIESWDHNDQQF QCAGVAVIR	52-78	Yes	Yes	Cystein, RR motif at C-terminus of peptide	-	Yes	Yes	Yes	Yes
	TIEPNGLLLPQYSNAPELIYI ER	80-102	Yes	Yes	NG motif, RR motif at N-terminus of peptide	-	Yes	Yes	Yes	Yes
	GITGVLFPGCPETFEDPQQ Q QQGQR	105-130	Yes	Yes	Cystein	-	Yes	Yes	Yes	Yes
	QGQGQSQR	131-138	Yes	Yes	N-terminal Q	-	Yes	Yes	Yes	No
	EGDIIALPAGVAHWCYND GDSPVVTVSLHTNNYAN QLDENPR	152-194	Yes	Yes	Cysteins	-	Yes	Yes	No	Yes
	HFYLAGNPDDQHQR	195-208	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	QGQQQFGQR	209-217	Yes	Yes	N-terminal Q, RR motif at C-terminus of peptide	-	Yes	Yes	Yes	Yes
	QQQHSHEQGEQEQQG EGNNVFSGFDAEFLADAF NVDVDTAR	220-261	Yes	Yes	Flanked on both N- and C terminus by RR motif, N-terminal Q	-	Yes	Yes	Yes	Yes
	LQSNQDK	263-269	Yes	Yes	RR motif at N-terminus of peptide, KR motif at C-terminus of peptide	-	Yes	Yes	Yes	Yes
	LQVVRPER	280-287	Yes	Yes	RP motif	-	Yes	Yes	Yes	Yes
	DVNGFEETICSLR	318-330	No	Yes	NG motif, cystein	-	-	-	-	-
	ADIYTEQVGR	341-350	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	INTVNSNTLPVLR	351-363	Yes	Yes	OK	In region with potential isoform	Yes	Yes	Yes	Yes
	WLQLSAER	364-371	Yes	No	OK	-	Yes	Yes	Yes	Yes
	EGLYVPHWNLNAHSVVY AIR	377-396	Yes	Yes	OK	-	Yes	Yes	Yes	Yes
	VQVVDDNGNTVFDDELRL	401-417	Yes	Yes	NG motif	-	Yes	Yes	Yes	Yes
	QGQVLTIPQNFVAVAK	418-432	Yes	Yes	KR motif at N-terminus of peptide, N-terminal Q	-	Yes	Yes	Yes	Yes
	AESEGFVWVAFK	434-445	Yes	Yes	KR motif at N-terminus of peptide	-	Yes	Yes	Yes	No
	TNDNAQISPLAGR	446-458	Yes	Yes	OK	In region with potential isoform	Yes	Yes	Yes	Yes

	ALPDDVLANAFQISR	464-478	Yes	Yes	OK	In region with potential isoform, in sequence stretch prone to missed cleavage	Yes	Yes	Yes	Yes
	QETTLVR	489-495	Yes	No	N-terminal Q	-	Yes	Yes	Yes	Yes

Tabel S1: selection of biomarker peptides for the detection of hazelnut in food. Criterium1: yes = detected in all 3 biological replicates in raw, roasted and caramelized hazelnuts of both origins, no = not detected in all 3 biological replicates in raw, roasted and caramelized hazelnuts of both origins; Criterium 2: specific toward hazelnut, Yes = specific; No = not specific; Criterium 3: Amino acid composition, presences of certain motifs and position in the protein sequence, OK = no motifs, amino acids or position in the protein that would exclude this peptide from selection; Criterium 4: protein specific properties and experimental observations, - = no specialties observed/no available data present to exclude this peptide from selection; Criterium 5A: robust toward heating hazelnuts origin Turkey; Criterium 5B: robust toward induction of Maillard reaction hazelnuts origin Turkey; Criterium 5C: robust toward heating hazelnuts origin Italy; Criterium 5D: robust toward induction of Maillard reaction hazelnuts origin Italy. Peptides in green are finally selected as potential biomarkers for the development of an LC-MS/MS method for the detection of hazelnut in food.