



Polymorphism at donkey β -lactoglobulin II locus: identification and characterization of a new genetic variant with a very low expression

Andrea Criscione¹ · Vincenzo Cunsolo² · Serena Tumino¹ · Antonella Di Francesco² · Salvatore Bordonaro¹ · Vera Muccilli¹ · Rosaria Saletti² · Donata Marletta¹

Received: 10 January 2018 / Accepted: 13 March 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

In the last years, donkey milk had evidenced a renewed interest as a potential functional food and a breast milk substitute. In this light, the study of the protein composition assumes an important role. In particular, β -lactoglobulin (β -LG), which is considered as one of the main allergenic milk protein, in donkey species consists of two molecular forms, namely β -LG I and β -LG II. In the present research, a genetic analysis coupled with a proteomic approach showed the presence of a new allele, here named F, which is apparently associated with a null or a severely reduced expression of β -LG II protein. The new β -LG II F genetic variant shows a theoretical average mass (M_{av}) of 18,310.64 Da, a value practically corresponding with that of the variant D ($\Delta_{mass} < 0.07$ Da), but differs from β -LG II D for two amino acid substitutions: Thr¹⁰⁰ (variant F) $\rightarrow$ Ala¹⁰⁰ (variant D) and Thr¹¹⁸ (variant F) $\rightarrow$ Met¹¹⁸ (variant D). Proteomic investigation of the whey protein fraction of an individual milk sample, homozygous FF at β -LG II locus, allowed to identify, as very minor component, the new β -LG II F genetic variant. By MS/MS analysis of enzymatic digests, the sequence of the β -LG II F was characterized, and the predicted genomic data confirmed.

Keywords Polymorphism · LGB2 gene · Donkey milk · Allergy · Mass spectrometry · SNPs

Introduction

In the last years, the use of donkey milk (DM) as a valuable product with numerous health-promoting properties aroused an increasing scientific interest (Salimei and Fantuz 2012; Brumini et al. 2016). Many studies have been carried out with the aim to improve the knowledge about

DM composition, to detect unintended contaminations, to improve dairy productions, and to accomplish a more accurate comprehension of its potential benefits for human nutrition (Claeys et al. 2014; Cunsolo et al. 2013; Malacarne et al. 2017; Medhammar et al. 2012). On this respect, the high lactose content, the characteristic protein profile characterized by a low casein/whey protein ratio, and the relevant percentage of essential amino acids, make the DM a potential functional product (Marletta et al. 2016). In particular, many clinical and in vitro studies indicated that DM, albeit with an adequate lipid integration, may represent a valid natural substitute of cow milk (CM) for feeding children affected by Cow Milk Protein Allergy (CMPA) and multiple food intolerances (Mansueto et al. 2013; Monti et al. 2012; Swar 2011; Vincenzetti et al. 2014). Although the mechanism of this tolerance has not yet been fully clarified, it is likely related to the low level of total proteins (very similar to that of human milk) and to the protein fraction composition (Bertino et al. 2010; Wasserman and Watson 2011). A recent study conducted on 14 Ragusano jennies, reported, on average, a total protein content of 1.54 g 100 g⁻¹, the

Handling Editor: P. Beltran-Alvarez.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00726-018-2555-1>) contains supplementary material, which is available to authorized users.

✉ Vincenzo Cunsolo
vcunsolo@unict.it

¹ Animal Production Section, Department of Agriculture, Food and Environment, University of Catania, Via Valdisavoia 5, 95123 Catania, Italy

² Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy

casein content of $0.71 \text{ g } 100 \text{ g}^{-1}$, a little higher than the whey protein content ($0.58 \text{ g } 100 \text{ g}^{-1}$) (Tidona et al. 2014). Moreover, many MS-based investigations aimed to the characterization of DM proteins clearly showed the remarkable differences in the primary structure of DM components with respect to the main CM allergens (Cunsolo et al. 2009a, b; Saletti et al. 2012). An evidence that may contribute to explain the reduced allergenic properties of DM as well as the rare cases of allergy to this milk (Cunsolo et al. 2017). In CM the β -lactoglobulin (β -LG) is the major whey protein and represents one of the most allergenic proteins (allergen Bos d5) (Swaigood 1995). Allergenicity of β -LG can also be related to its high resistance to enzymatic hydrolysis that occurs in the intestinal lumen (Restani et al. 2009; Tidona et al. 2011). On the contrary, β -LG is not expressed in human milk, whereas in DM it is approximately 40% of the whey proteins equal to the level in mare milk and lower than that in cow milk. In DM, β -LG consists of two molecular forms, the major β -LG type I and the minor β -LG type II. The minor component is expressed by a second gene which seems to be originated by gene duplication (Conti et al. 1984; Godovac-Zimmermann et al. 1985, 1990). Donkey β -LG I and the minor β -LG II show differences in their electrophoretic mobility, chain length and primary structure. Up to date, only two variants of β -LG I (UniProt Accession No. P13613), named A and B (Godovac-Zimmermann et al. 1988) have been detected and characterized. On the contrary, donkey β -LG II locus shows a remarkable genetic polymorphism since five different variants named A (UniProt Accession No. P19647) (Godovac-Zimmermann et al. 1990), B, C (Herrouin et al. 2000; Zuccaro et al. 2010), D (Cunsolo et al. 2007) and E (Chianese et al. 2013) have been identified and characterized, so far. Moreover, the occurrence of a null or severely reduced expression of β -LG II protein has been reported in Italian donkey breeds (Criscione et al. 2009; Chianese et al. 2013).

On the light of this scenario, the aim of this study was to investigate the polymorphism at β -LG II in Ragusano donkey breed. The genomic sequence of the β -LG II gene revealed the presence of B and C alleles as well as a new unreported nucleotide sequence (named F) at homozygous and heterozygous status. To confirm the genomic data, the whey protein fraction of an individual milk sample homozygous FF was investigated by a proteomic approach.

Materials and methods

Chemicals

Dithiothreitol (DTT), acetonitrile (ACN), trifluoroacetic acid (TFA), iodoacetamide (IAA), ammonium bicarbonate, formic acid (FA), endoproteinase GluC (*Staphylococcus aureus*

strain V8) and all other chemicals used all along the experimental work were current pure analytical grade products and purchased from Sigma-Aldrich Corporation, St Louis, MO, USA. Water and acetonitrile (OPTIMA® LC/MS grade) for LC/MS analyses were purchased from Fisher Scientific Italy (Rodano, MI, Italy). Modified porcine trypsin was purchased from Promega (Madison, WI, USA).

Blood samples and DNA extraction

Blood was sampled in K3-EDTA vacutainer from ten Ragusano donkeys producing milk with different isoelectrophoretic (IEF) patterns at whey proteins according to Criscione et al. (2009). Animals were handled by specialized personnel who took care of animal welfare according to the European Union guidelines (EC 2010). In particular, sampling was performed including animals which produced milk with a detectable amount of β -LG II and animals which apparently lacked this protein fraction in milk. Genomic DNA was extracted using “Illustra blood Prep Mini Spin” kit (GE Healthcare, Little Chalfont, UK) according to the manufacturer’s protocol and then checked for quality and concentration by NanoDrop ND 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA).

DNA sequencing

Eight primer pairs (Table 1) were designed using Oligo 7 software (Molecular Biology Insights, Inc., Cascade, CO, USA) according to the genomic sequence of donkey’s β -LG II (GenBank Accession no. HM012799), to amplify the entire β -LG II gene and a part of the region 5′ (227 bp), for a total of 4511 bp. Each PCR reaction was performed in a 25 μ l of final volume containing 50 ng of genomic DNA, 0.4 μ M of each primer, 1X PCR buffer, 1.5 mM MgCl₂, 200 μ M dNTPs (dATP, dTTP, dCTP and dGTP) and 1 units of BIOTAQ™ DNA Polymerase (BIOLINE). Amplification was conducted in a Thermal Cycler 2720 (Applied Biosystems Foster City, CA, USA) for 29 cycles; temperatures of annealing (T_a) are reported in Table 1. The PCR products were checked by electrophoresis on 1.6% agarose gel, purified by Wizard® SV Gel and PCR Cleaning-Up System (Promega Corporation, Madison, WI, USA) and sequenced using the BigDye Terminator Kit, on an ABI PRISM® 3130 Genetic Analyzer equipped with Sequencing Analysis software and SeqScape 2.6 software (Applied Biosystems Foster City, CA, USA).

Sequence data analysis

The nucleotide sequences were checked using Sequencing Analysis Software (Applied Biosystems Foster City, CA, USA) and subsequently analysed with SeqScape 2.6

Table 1 Amplification PCR primer sequences of asinine LGB2 (designed on GenBank Acc. No. HM012799)

Region	Primers e seq. (5' 3')	Ta ^a (°C)	Sizes (bp)
PROMOTER	F: 5'-GAACCTCTCTGCCCCGTGCTC-3' R: 5'-GTGACTCATCCTCGGCCCTC-3'	61	725
Ex I	F: 5'-CCCAGTTGAGGAAGCCACTGAG-3' R: 5'-CCCCCTGGATGTGGAGAGCAAG-3'	63	428
INT 1-Ex II	F: 5'-TGGCCTTGCTCTCCACATCCAG-3' R: 5'-CAGCATGGCCAGGTCCCATCAG-3'	63	524
INT 2 - Ex III	F: 5'-ATCTGGGCGGTGTCTTGCATC-3' R: 5'-ACCAGCGCTAACCCACACCTG-3'	63	536
INT 3	F: 5'-GCCTCATCGCGAGCACCAG-3' R: 5'-CTCCACCGGGACAGCATCCTC-3'	62	1320
Ex IV	F: 5'-CTGCGCCTCCGTTCTCTCAC-3' R: 5'-CGACCAAGGGCCACATCTCAG-3'	64	918
INT 4	F: 5'-CTTGTCTGCCCTGCCGTCGTC-3' R: 5'-CCATGACCTCCTCGTCGACCTTC-3'	63	1591
Ex V-VI-INT 5	F: 5'-TGAGCCCTGGACCACCCTG-3' R: 5'-TGAGACCCCAACCATGTGCCAG-3'	64	612

^aTa annealing temperature

Software (Applied Biosystems Foster City, CA, USA). Genomic sequences of the asinine β -LG II gene and a part of the region 5' (227 bp), accounting for 4511 bp, were aligned with *Equus caballus* β -LG II gene reference (GenBank Accession no. AF107201) and translated to amino acid sequences using MEGA 6 software (Tamura et al. 2013). Protein sequence handling, including molecular mass, isoelectric point and theoretical isotopic mass distribution calculation, were performed using the General Protein Mass Analysis for Windows 9.5 (GPMW) software (Lighthouse data, Odense, Denmark).

Sample milk collection and treatment

A milk sample was collected from a multipara Ragusano jenny with FF genotype at LGB2 locus. The animal, reared in a specialized dairy farm, was clinically healthy and at middle lactation stage. The milk sample, kept cool at 4 °C, was defatted by centrifugation (1700×g, 30 min) at 15 °C; the fat layer was solidified at − 20 °C for 15 min and removed. Separation of the protein fraction into caseins and whey proteins was conducted according to a modified method from Ochirkhuyag et al. (2000). Specifically, whole casein was obtained from skimmed milk by isoelectric precipitation (pH 4.2) at 22 °C, using 1 mol L^{−1} HCl. The supernatant, containing the whey proteins, was recovered by centrifugation (2800×g, 30 min) at room temperature. This procedure was repeated three times. The total supernatants were pooled and subsequently freeze-dried and stored at − 20 °C.

Whey proteins fraction digestion

An aliquot corresponding to 5 μ g of the whey proteins fraction was dissolved in 100 mM NH₄HCO₃, pH 8.3, at a concentration approximately of 1 μ g/ μ L and then, subjected to reduction, alkylation and enzymatic digestion using different enzymes. In detail, DTT (50-fold molar excess over the disulfide bonds) was added, and the reduction was carried out at 57 °C for 30 min in nitrogen atmosphere. At the end of the reduction reaction, IAA was added to the mixture at a molar IAA/DTT ratio of 2:1, and the alkylation was performed at room temperature for 60 min in the dark. The reaction was stopped by cooling in liquid nitrogen. To maximize the sequence coverage, the reduced and alkylated protein fraction was subjected to two different enzymatic digestions. An aliquot was digested using modified porcine trypsin, whereas another one was digested using a mixture (1:1) of porcine trypsin and endoproteinase Glu-C. Each enzyme was dissolved in 100 mM NH₄HCO₃, pH 8.3, added at a molar enzyme/substrate ratio 1:50 and the solutions were incubated at 37 °C for 4 h. The digestion reactions were stopped by cooling in liquid nitrogen.

Mass spectrometry analysis

nLC-nESI MS of whey proteins was carried out performed using a Thermo Scientific Dionex UltiMate 3000 RSLC-nano system (Sunnyvale, CA, USA) coupled on-line with a Thermo Fisher Scientific Orbitrap Fusion Tribrid® (Q-OT-qIT) mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The solution of the whey proteins was filtered

through a 0.45 μm filter (Alltech). Then, 1 μL of the solution was loaded onto an Acclaim[®] Nano Trap C18 Column (2 cm $\times$ 100 μm ID, 5 μm , 100 $\text{\AA}$). After washing the trapping column with solvent A ($\text{H}_2\text{O}/\text{ACN}$, 95/5 + 0.1% FA) at a flow rate of 7 $\mu\text{L}/\text{min}$ for 3 min, the solution was switched from the trapping column onto a PepMap[®] RSLC C4 EASY-Spray column (15 cm $\times$ 75 μm ID, 2.6 μm , 150 $\text{\AA}$). The column was eluted at a flow rate of 0.3 $\mu\text{L}/\text{min}$ and 40 $^\circ\text{C}$ with a linear gradient of solvent B (ACN + 0.1% FA) in A from 5 to 60% in 35 min, followed by 60–95% in 10 min, and then by holding 95% B for 5 min. Finally, the column was eluted with the solvent B from 95 to 5% in 10 min, and finally equilibrated at 5% of B for 15 min. Eluting proteins were converted to gas-phase positive ions by electrospray ionization using a source voltage of 1.7 kV and introduced into the mass spectrometer through a heated ion transfer tube (320 $^\circ\text{C}$). Survey scans of multi-charged protein ions from 600 to 1800 m/z were performed at 120 K resolution (@ 200 m/z) using the following parameters: RF lens, 80%; Auto Gain Control target: 20,000; maximum injection time, 50 ms. MS analysis of whey proteins was performed in technical duplicates. The experimental monoisotopic mass and isotopic mass distribution of proteins were determined by deconvolution of multi-charged mass spectra using the Xcalibur Xtract deconvolution algorithm optimized for isotopically resolved data (Thermo Fisher Scientific, Bremen, Germany) (Senko et al. 1995; Horn et al. 2000; Goraczko 2005). Nano LC–nESI MS/MS analyses of peptide mixtures were performed using the same system above described, but under different chromatographic and mass spectrometric conditions. Briefly, 1 μL of the peptide mixtures was loaded onto an Acclaim[®] Nano Trap C18 Column (100 μm $\times$ 2 cm, 5 μm , 100 $\text{\AA}$). After washing the trapping column with solvent A ($\text{H}_2\text{O}/\text{ACN}$, 95/5 + 0.1% FA) at a flow rate of 7 $\mu\text{L}/\text{min}$ for 3 min, the solution was switched from the trapping column onto a PepMap[®] RSLC C18 EASY-Spray column (50 cm $\times$ 75 μm ID, 2 μm , 100 $\text{\AA}$). Peptides were separated by elution at a flow rate of 0.25 $\mu\text{L}/\text{min}$ and 40 $^\circ\text{C}$ with a linear gradient of solvent B (ACN + 0.1% FA) in A from 5 to 65% in 82 min, followed by 65–95% in 5 min. We finished by holding 95% B for 5 min and re-equilibrating the column at 5% B for 20 min. The eluting peptide cations were converted to gas-phase ions by electrospray ionization using a source voltage of 1.6 kV and introduced into the mass spectrometer through a heated ion transfer tube (275 $^\circ\text{C}$). Survey scans of peptide precursors from 400 to 1800 m/z were performed at 120 K resolution (@ 200 m/z). Tandem MS was performed by isolation at 1.6 Th with the quadrupole, HCD fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. Only those precursors with charge state 1–4 and an intensity above the threshold of 5×10^3 were sampled for MS². The dynamic exclusion duration was set to 60 s with a 10 ppm tolerance

around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 3 s cycles, meaning that the instrument would continuously perform MS² events until the list of non-excluded precursors diminishes to zero or 3 s, whichever is shorter. MS/MS spectral quality was enhanced enabling the parallelizable time option (i.e. using all parallelizable time during full scan detection for MS/MS precursor injection and detection). Peptide mixtures were analysed running a MS technical duplicate per sample. Mass spectrometer calibration was performed using the Pierce[®] LTQ Velos ESI Positive Ion Calibration Solution (Thermo Fisher Scientific). MS data acquisition was carried out by utilizing the Xcalibur v. 3.0.63 software (Thermo Fisher Scientific).

Protein identification

LC/MS/MS data were analyzed and searched by the PEAKS de novo sequencing software (v. 7.0, Bioinformatics Solutions Inc., Waterloo, ON, Canada). In particular, the de novo amino acid sequences generated from each spectrum were searched against a custom protein database, constituted by the five known variants of the donkey β -LG II *plus* the predicted sequence of the new β -LG II variant F. The following parameters were used: enzyme: trypsin (or the combined action of trypsin and endoproteinase Glu-C); maximum missed cleavages for peptide: 3; allowed non-specific cleavage at one end of the peptide; cysteine carbamidomethylation as fixed modification; oxidation of methionine, transformation of N-terminal glutamine and N-terminal glutamic acid residue in the pyroglutamic acid form, as variable modifications. The precursor mass tolerance threshold was 10 ppm and the max fragment mass error was set to 0.6 Da. Peptide spectral matches (PSM) were validated using a Target Decoy PSM Validator node based on q values at a 1% FDR. All Peptides Spectrum Matches were also manually checked.

Results

Sequencing of the LGB2 gene

The entire asinine LGB2 gene was sequenced (Table 2). In comparison with the equine counterpart (GenBank Accession no. AF107201) the aligned exonic regions revealed nine Single Nucleotide Polymorphisms (SNPs). All the SNPs were responsible for amino acid substitutions (Table 3), except for a synonymous mutation, detected in exon V (c. 525 G > A; p. Gly157Gly). In particular exon I, encoding for the signal peptide and the N-terminus of the protein, resulted always monomorphic and identical to the horse's one, whereas exons II and III showed one transition each (c.164 T > C; c.248 C > T), responsible for the substitutions,

Table 2 Nucleotide sequence of *Equus asinus* LGB2 allele F (4511 bp)

1	AGTGGATCTT	CCGGTTTGCT	CTCGGAGGGA	AAAAGGCTTC	CTATTGTCCC	AGTTGAGGAA
61	GCCACTGAGG	GCCGAGGATG	AGTCACGGAG	ATTGCGGGAA	CCTCATGGCG	GGGGGCTGGA
121	GGGCGCTGGC	GGGCGGCCCC	TCTCCAGCTC	CCCTCCTTTA	TAGGGCGCCG	AGCCCGTGCC
181	TCGGTCTCCT	CACCTCCCGG	AGCTCAGAGT	GCACCCGCAG	CGCAGCCATG	AAGTGCCCTCC
241	TGCTTGCCCT	GGGCTTGTC	CTCATGTGTG	GCAACCAGGC	CACCGACATC	CCCCAGACGA
301	TGCAGGATCT	GGACCTCCAG	GAGGTTCCAG	GATGGCTGGG	GTGGGCTGAG	CTGCGGGGCG
361	GGCAGGGGGC	TGGCTCGGAA	ATGGAAGGTT	GCACAAATGG	AGCCATGACG	TGAGGAAACA
421	CTTGGCCTTG	CTCTCCACAT	CCAGGGGGTT	TGCTACTTGG	TTTCCTAAGA	TTCAGATGGT
481	CTCGGGTGAC	TCAGAATTTG	GACTCTAAGT	CCATCGGGGT	TGACACAGCA	GGGTTATTTT
541	GAGGGTGGGA	AGCTTGGCAG	CCTAGCGTCT	GGCAGCCGCG	AGAGGGTCAG	CCAGTGTGTA
601	TGTCCATCCC	GTTAGACTCG	GCCTGGCCGG	CATGGGGTGG	GACGGAGAGT	CCACTGTGGG
661	GCTTGGGGGG	CTCTTCTGCG	CCCTGGAGTG	CAGCTCGAGG	CCCCCCTCA	GGTGGCAGGG
721	AGGTGGCACT	CCGTGGCCAT	GGTGGCCAGC	GACATCTCCC	TCCTGGACCT	CGAGATGTCC
781	CCTCTGAGAG	TGTACGTCGA	GGAGTCCGAG	AGGACCCCGG	AGGGCAACCT	GGAGATCATC
841	CTGCGCGAAG	GGTGGGTGTG	CCCCCGGGCT	GCGGGCCAGT	GAGGGCCTGG	CTCCAGGAGG
901	GGTGGGTGGG	CAGGTGGGGC	TCCTGATGGG	ACCTGGCCAT	GCTGTGAGTG	GGCATGGTGC
961	CCAGGGGGTG	GGGATTGGGG	CTTGGAGGTT	CCTAGTCCCG	TAGAGGGCCG	CGCCAGGGCT
1021	CTGGAATAAC	AGGCTGACCA	CATAGAGTGA	CCCCGGTCAG	TGTGGGGAGA	GGCCGGCGTG
1081	GAGGCCGGCT	GGTCAGCTGC	AGGCTTCAGG	TCCTCAGGGC	CTGGTCCGTT	CCTGCCGTGA
1141	ATCCTGTCTC	CCCCCAGAGT	GGGTGACCAA	GGGTGACAGA	ATTCTGAGGT	TGAGCTGTGA
1201	TGACGTCCTT	GGAAATGCTG	TGATCAAGCC	CCACATAGAC	AGGGTCTGTT	TTGAGCTGCC
1261	ATGAGCCCTC	GGCCACAGGC	CCCCACACCT	GCCTGCCGGG	TTTCCTAGCT	GTGTCCCTTG
1321	GTCCCTTGCT	CAACCTTCCT	GAGTCGCTGG	CTTGGGGCCA	CGAGGGCCCG	TCACCTTCAG
1381	GCCGAGACCT	CACGGCCTTG	CTACCAGCCA	GCCCAAGGCC	ATCTGGGCCG	GTGTCTTGTC
1441	ATCCTGTGGA	CGCTTCTGTC	AGCCACAGCTC	CTGTCCCTGC	ACCGGGTTCC	TGAATGACCC
1501	TCACCACTGC	GGGGCTTTTC	AGGGCAAAAC	ACGTATGTGT	TGAGAGGAAC	ATTGTGCCCC
1561	AGAAGACTGA	GGACCCAGCT	GTGTTACAGG	TCAACTGTAA	GTGTCCCTAC	TCTGTGCCCA
1621	CAGGGCCACC	TGGCACCCAC	CCTCAGGTGG	CCCTGGGCGA	GGCTGTACCC	CTCCTGACGC
1681	TGGGCCCTCT	GAGGCTGCTC	TGAGTCACTC	AACGTCCTGC	AGCAGGAAGA	GCAGGGCTTG
1741	CGTCTGGCGC	CTCATCGCGA	GCACAGGCC	CCGCTTGCCG	CACCCCTGCA	GCTGTGTCAG
1801	CACCTCGGGG	GTGATGAGGC	CTAGGGACTC	CACCTCCCTC	ATGGCTCACC	CTGCTGCCCA
1861	GCTCCTGTTT	TCCTGCTTGG	ACGTAGGAAG	AGAGCGGGCA	GCCTCTTGTC	CACTGCAAGG
1921	TGGGTGCCAT	CCACGCAGGT	GTGGGTAGCG	GCTGGTTCCA	ACAGCAGGCA	TGTGGCAACG
1981	TCCCTGCTCA	GGGGCTGGCA	GGCTGTGGCT	CCTTAAAAAT	ATCCAAATAA	ACCAGACCTG
2041	TAGGCAAAAC	CACATATGGT	TTGCCTGCCA	AACCTGGAAG	CTGCTTCTGG	AACGTTCTGG
2101	GCTTCACCCA	CCTCTCTCCT	CACTCTGGCG	ACTGCCCCCT	GGCCCTGGCA	GCCTCCCATC
2161	CTGCAGTCAG	CCTCATCTGG	GACAGGCCCC	TGCACCCACC	CCCGACTCCA	CCCTCCCCAC
2221	CAGGGGCCCA	CCTGCCCCCT	CTCCCAAGTC	CACAGGCTT	CCTGCCCTCT	CTTCTGTAC
2281	TTCTCTTCTC	TCTGAGAGGC	GGTGAGTGTG	GTGGGTACAA	TCTTGTCCCC	AGAAGTGACC
2341	CTGTACACGTC	TGGTCCCAGC	AACCCAGGGC	AGCCCTTAG	CCCCCGTCCC	TGCGCCTCCC
2401	GTTCTCTACC	TGGGAGCGGG	CTCAGCTGCT	GTTCGCGCCG	TGCAGCTGAG	AGGTTGCGGG
2461	GAGTCTCTGA	GAGCTGGGGC	AGTGCTGGGG	GGGCTGCTTG	TCTGCCCTGC	CAGTGTCTTG
2521	CACCAAAAGT	GGTCAGGGCT	GTGGGTCTCC	ACCAGTGCCC	TGGTTTCTGC	CTCTGAAGAA
2581	GGGAGACGTG	CCTACCCAC	CTCGCAGTAA	TTCCCCATTT	CTTTCTTCCA	CTCTCTGATG
2641	TCCACGTGTA	ACTCTCGACT	CAAGAGGGGC	CCCGATCCTG	GACGTTCTTG	TGATGCTCTT
2701	GAGGACCTGG	CACCTGCCCT	TGGGAGGGGT	CGCCTGGATT	TAGTGCTGTA	GTATGCTGTA
2761	ATGTCTGAAT	GAATTAACGC	CACCTCAGGA	CGTGGGGGCT	GGAGTTTGGC	CAAGCTGTGA
2821	GTAGGTCTGA	GCAGCTTCGT	TGGAGCGCCA	ACACATCGAG	GGGCTGTTTA	CAGCTGGAGC
2881	CAGGGGACAG	AGGGGCTGCC	TGGGAGCGCG	TGCCGCATAG	GGGCCCTCT	GGGGCCCAT
2941	ACACTTGTGG	TGGGGCTGCC	AGTCATGGGC	TTTTGTTGGA	ACCGTGGAAG	TGAACACAG
3001	GGGGAAGGAG	ATGCCCTGAG	GTGCGGGTGT	GTCTGCACT	GTGCGGGAGG	ATGCTGTCCC
3061	GCTGGAGGTC	AGAGGGGCAC	TCAGATGAGA	ATCCAGGCTT	CTGTTTTCGC	CTCTTTTCCA
3121	GATCAAGGGG	AGAGAAAGAT	TTCCGTGCTG	GACACAGACT	ACACCCACTA	CATGTTCTTC
3181	TGTGTGGGGC	CCCCCTTGCC	CAGCGCTGAG	CAGGGCACGG	TGTGCCAGTA	CCTGGGTACG
3241	TGGTGTGTCT	AGCCAGGGGG	GCCTTCGGCG	GGGGCCACTC	CTCATGAGAT	GGGGCCCTTG
3301	GTCGTCTTTA	AACAACGAGC	ACTCATGACG	ACTGACCCCT	GGCTGGAGGA	AGGGAAGGGG
3361	CTGCCCACCC	GCCTCTGTGG	GTCCCCTTGA	CGAAGGTCAT	CGGGTGTGGT	GTCCGACACG
3421	AGTGTGGCTC	TGATGTGGGT	GACTTTGGGA	TGAGGGGACG	GGGCTGGTGC	TGATGTCCGG
3481	GACAGGCTGA	GTAGGGGCGA	GGCAGAGGAT	GAGGGGTCTG	TCTCAAGTGG	AGGGGCCACA
3541	CTCACCCCTG	GGGGAACAGG	CTGTATTGAT	CCAATAACCT	GACGGAAGGA	GATTTTCAAT
3601	GACCAGAAAA	GTTCAGAAAT	GAAAGCAACT	TTAAAGGGTG	AGATTCTCAC	AGAAGTAAAT
3661	ATGTGGTTGA	TGCCACCATT	GTCTCAGGG	GATCCCAGGG	TCTGAGTAGA	GCATGGTGTG
3721	GGGCTGCTGA	GGGCTCCCTG	ATCCATAAAT	CGGCCACCT	GGGTGGAGGA	GACCCACAG
3781	CCTGGGAGCA	GGGGCCCTTT	TTCTTACCAT	GACTGGGCGA	CACCCCCAGC	TGCTGCTCTG
3841	GGAGGAGGAG	GGGGCCTGGG	CCGAGGAGGG	AGGGAGAGAC	CCTGGAGGAC	ACGTGGGCTG
3901	AGCCCTGGAG	ACCACCTGG	GAGGCTGGGC	TGGACCAAGC	TCTACCGACC	TCAGGACGTT
3961	GGCTTGTCTT	TCCATGCTC	CCCAGCCGGG	GCACCCGGCC	CCCTGAGGTC	TTGTCAAAGC
4021	CCCCACAGCC	CAGGGACTGA	CCCCCGACCC	CCACACAGCA	GGACCCAGAA	GGTCCACGAG
4081	GAGGTCATGG	AGAAATTGAG	CAGAGCCCTC	CAGCCTCTGC	CAGGGCACGT	CCAGATCATC
4141	CAGGACCCGA	GCGGGGGACA	GGGTGAGCTC	TGGCCCGCGG	GGGCCCTGCC	TGGGTTTGAG
4201	AAGCAGAGGC	TGTCTCTGCT	CTGAGTCCCT	ATTGCTGGTT	CTGTGGGAAC	CAGGGGCGCA
4261	TCTTTCTCCC	ACGGTGGCCT	CCCAGGAGGC	GGGTGGCGGG	CAGGTGGAGG	GTGCAGAGCC
4321	CCTGAGGTGG	GGTCCCCCGG	CCCCACCCAG	ACCCCCCTGG	GGCCGGGGCT	GCCTGTCTGA
4381	CTCAGAGTCT	TTCCCCCGCG	CTCTCCCTCC	CCCTGCAGAG	CGATGCGATT	TCTAGCGGAG
4441	CTCCGGCCCG	TCCGCCTCAT	CGGGTAACGT	ACCAACCAGG	CCCCACGTCC	TGGGCACATG
4501	GTGGGGTCTC	A				

Coding region (546 bp) is highlighted in grey (ExI 96 bp, ExII 140 bp, ExII 74 bp, ExIV 114 bp, ExV 105 bp, ExVI 17 bp)

Table 3 β -LG II DNA based variants identified in 10 Ragusano donkeys

Genotypes (<i>n</i>)	Amino acid position								
	37	65	96	100	110	118	147	151	162
F/F (1)	Ala	Val	Asp	Thr	Pro	Thr	His	Ile	Asp
B/F (3)	Ala	Val	Asp	Ala/Thr	Pro	Thr	His	Ile	Gly/Asp
C/F (1)	Ala	Val	Glu/Asp	Ala/Thr	Pro	Thr	His	Ile	Gly/Asp
B/C (3)	Ala	Val	Asp/Glu	Ala	Pro	Thr	His	Ile	Gly
B/B (2)	Ala	Val	Asp	Ala	Pro	Thr	His	Ile	Gly

Val37Ala and Ala65Val, respectively. Exons IV and V were the most polymorphic, showing 1 transversion and 5 transitions, (c. 342 C > G; c. 352 G > A; c. 407 T > C; c. 494 G > A; c. 505 G > A; c. 525 G > A). These mutations resulted in the substitutions Asp96Glu and Ala100Thr and in three other substitutions (Met118Thr, Arg147His and Val151Ile). Furthermore, the small exon VI (17 bp) presented, in 5 samples, the transition c. 539 G > A resulting in the substitution Gly162Asp. BLASTN analysis of these results, highlighted two sequences which correspond to the already known B (GenBank Accession No. HM012799) and C (GenBank Accession No. HM012800) alleles. On the contrary, no data were found for a third sequence, which represented a novel allele, here called F, and characterized by two SNPs with respect to the allele B; namely the c.352 G > A SNP (GCC > ACC), located in exon IV, leading to a substitution

p. Ala100Thr, and the c.539 G > A SNP (GGT > GAT), located in exon VI, giving the substitution p. Gly162Asp. In the present study, among the ten donkeys investigated, DNA sequence analysis allowed to detect one homozygote with FF genotype and four heterozygotes carrying the new allele (Table 3). In the light of these findings the genetic variability of the LGB2 gene in donkey can be updated as showed in Table 4, in which the coding sequence variations (where available) and the corresponding amino acid substitutions for asinine β -LG II variants are reported with respect to LGB allele/variant B. It is interesting to note that the new predicted β -LG II F genetic variant practically presents the same molecular mass of the already known variant D ($\Delta_{\text{mass}} < 0.07$ Da), even if carrying two amino acid substitutions: Thr¹⁰⁰ (variant F) $\rightarrow$ Ala¹⁰⁰ (variant D) and Thr¹¹⁸ (variant F) $\rightarrow$ Met¹¹⁸ (variant D). To confirm the expression

Table 4 List of known donkey β -LG II variants including the new variant F

β -Lactoglobulin II variants	Monoisotopic Mass (Da) ^{a)}	Average Mass (Da)	<i>pI</i>	Nucleotide position ^{b)} (Amino acid position)									
				342		352		539					
				(2)	(18)	(25)	(96)	(100)	(110)	(118)	(162)		
A	18246.900	18258.690	4.45							(Cys)	(Met)		
B^{d)}	18210.958	18222.576	4.45	GAC (Asp)	AGG (Arg)	GTG (Val)	GAC (Asp)	<u>G</u> CC (Ala)	CCC (Pro)	ACG (Thr)	GGT (Gly)		
C^{d)}	18224.973	18236.603	4.47				GAG (Glu)						
D	18298.956	18310.706	4.39									(Met)	(Asp)
E	18241.935	18253.654	4.45									(Met)	(Asp)
F^{c,d)}	18298.974	18310.639	4.39	(Asn)	(Lys)	(Ala)							
								<u>A</u> CC (Thr)				GAT (Asp)	

Coding sequence variations (where available) and the resulting amino acid substitutions are reported taking as reference the allele/variant B

^{a)}Monoisotopic and Average Molecular masses were calculated considering the cysteines in oxidized form (-SS); isoelectric point calculated by GPMW considering the cysteines in oxidized form (-SS), and based on the algorithm of Skoog and Wichman (1986)

^{b)}The bold number denotes the position within the coding sequence whereas the italic number is referred to the amino acid position within the protein sequence. Amino acid differences are reported with respect to the variant B

^{c)}Properties of variant F are referred to the predicted amino acid sequence

^{d)}Underlined nucleotides indicate the SNPs that differentiate alleles B, C and F

of the new β -LG II allele, the whey protein fraction of an individual milk sample, produced by a jenny with FF genotype, was characterized by means of nUPLC-nESI high resolution MS, enzymatic digestions and nUPLC/nESI MS/MS.

Mass spectrometry results

Figure 1 shows the total ion current (TIC) chromatogram of the whey protein fraction obtained from the individual milk sample under investigation. The chromatogram shows the presence of four main peaks, here labelled as peak 1, 2, 3 and 4. The experimentally deduced monoisotopic masses (M_{mi}) of components eluting in the chromatographic peaks 1, 3 and 4 agree with those of three already known donkey milk components, such as lysozyme B (peak 1), α -lactalbumin (peak 3) and β -LG I variant B (peak 4), respectively (see Supplementary Material, Figures S1, S5 and S6). In the chromatographic peaks 2 and 2' (r.t. range 26.7–27.2 min) elute some components with an experimentally deduced M_{mi} ranging from 5 to 9 kDa (Supplementary Material, Figures S2 and S3). These polypeptides may be identified as casein fragments mainly originated by the action of the native plasmin milk protease (Supplementary Material, Figure S4). Figure 2a shows the multi-charged spectra of the chromatographic fraction with retention time range of 28.7–29.0 min, and labelled as peak 3'. Monoisotopic deconvoluted mass spectrum (Fig. 2b), reveals that the main component is the α -lactalbumin (experimental M_{mi} : 14,204.792 Da; theoretical M_{mi} : 14,204.805 Da; ΔM_{mi} = 0.91 ppm), which mainly elutes in peak 3. In addition, a very minor protein component with an experimental M_{mi} of 18,298.959 Da is also present. In particular, Fig. 2c reports the isotope resolution

of protein charge states of this component across the mass range. This molecular mass might be related to the already characterized donkey β -LG II variant D (theoretical M_{mi} 18,298.956 Da; ΔM_{mi} = 0.16 ppm) (Cunsolo et al. 2007). Alternatively, it may be due to the new β -LG II F genetic variant (theoretical M_{mi} 18,298.974 Da; ΔM_{mi} = - 0.81 ppm) (Table 4; see Supplementary Material, Figure S7, for the comparison from the theoretical isotopic and the experimentally determined mass distribution of β -LG F). As above described, these two variants even if show the same molecular mass (Table 4), differ for two amino acid point substitutions, located at the positions 100 and 118; particularly, variant D presents an alanine and a methionine residue, respectively; on the contrary, variant F shows a threonine residue in both 100 and 118 positions. Therefore, to confirm the expression of the new β -LG II F genetic variant, the entire whey milk fraction was subjected to reduction and alkylation of cysteine residues and finally digested with porcine trypsin. LC/MS/MS analysis, database search and manual inspection of MS data allowed the characterization of five peptides corresponding to theoretical tryptic fragments common to both β -LGs II variant D and F (Table 5, Fig. 3). Specifically, the Thr¹-Arg¹⁸, Trp¹⁹-Arg⁴⁰, Val⁴¹-Arg⁵⁹, Asn⁷⁰-Arg⁹⁰ and Ala¹⁴⁰-Arg¹⁶⁰ peptides were found and characterized. MS/MS spectra and correlation with the theoretical expected fragment ions of these peptides are reported in the Supplementary Figures S8, S9, S10, S11 and S12, respectively. Unfortunately, these tryptic fragments cannot be used to discriminate the β -LGs II variant D and F because they do not cover the aforementioned two amino acid differences at position 100 and 118. Therefore, with the aim to obtain a more comprehensive characterization of the β -LG II variant present

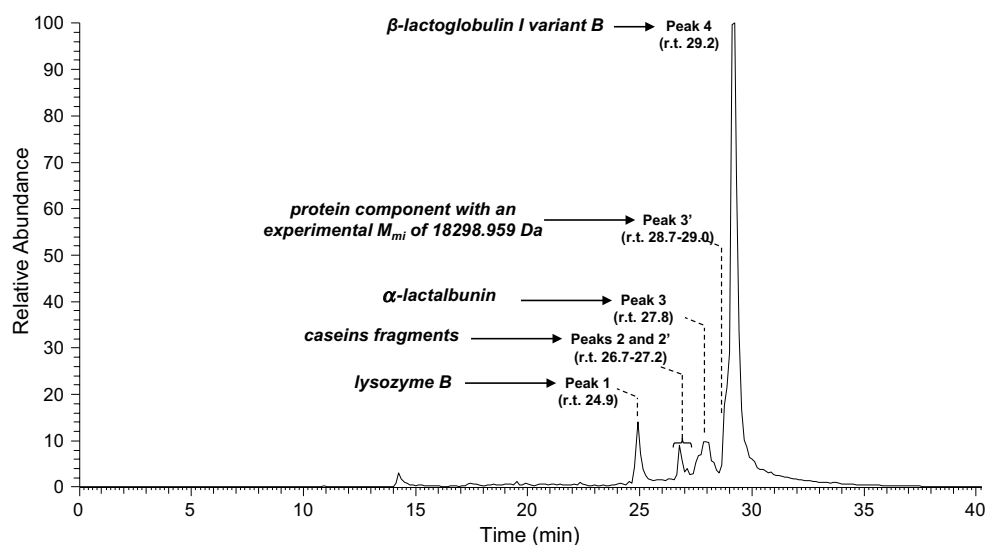


Fig. 1 Total Ion Current (TIC) chromatogram of the whey protein fraction obtained from the milk sample under investigation (genotype FF)

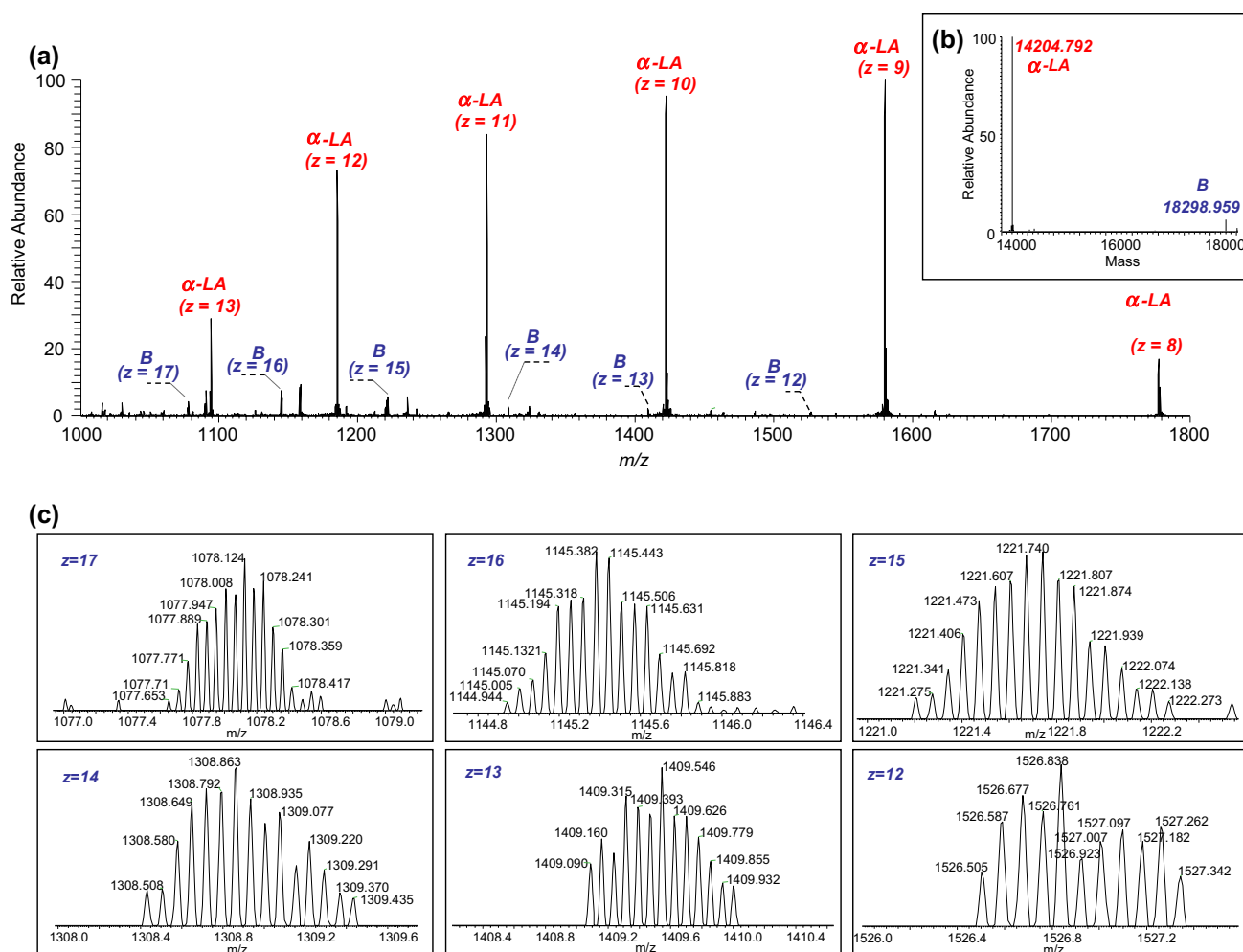


Fig. 2 **a** Multi-charged mass spectra relative to the components eluting in the chromatographic peak 3'; **b** monoisotopic deconvoluted mass spectrum (mass zero-charge); **c** isotope resolution of protein charge states across the mass range

Table 5 Amino acid position, theoretical monoisotopic mass, theoretical monoisotopic doubly or triply charged ions, experimentally measured mass and error of enzymatic fragments from β -LG identified and characterized in peaks 3' are reported (Fig. 1)

Position	Theoretical monoisotopic mass	Theoretical monoisotopic m/z (z)	Experimental m/z (z)	Error (ppm)	Reference sequence (cDNA derived sequence of β -LG variant F)
1–18	2028.9786	1015.4966 (2)	1015.4958 (2)	– 0.8	TDIPQTMQDLDLQEVAGR
19–40	2383.1842	795.4020 (3)	795.4013 (3)	– 0.9	WHSVAMVASDISLLDSESAPLR
41–59	2239.2212	747.4144 (3)	747.4136 (3)	– 1.0	VYVEELRPTPEGNLEILR
70–90	2378.1866	793.7362 (3)	793.7373 (3)	1.4	NIVAQKTEDPAVFTVNYQGER
116–125	1203.5819 ^a	602.7982 (2)	602.8018 (2)	5.9	HGTVCQYLAR ^b
140–160	2239.1709	747.3976 (3)	747.3977 (3)	0.1	ALQPLPGHVQIIQDPSGGQER

The cDNA derived sequence of β -LG variant F (Fig. 3) was taken as reference

The peptide marker of the new variant F, which allows to distinguish it with respect to the variant D is italicized

^aMolecular mass is calculated considering the cysteine residue as carbamidomethylcysteine

^bFound in the peptide mixture obtained by the combined use of trypsin and endoproteinase Glu-C

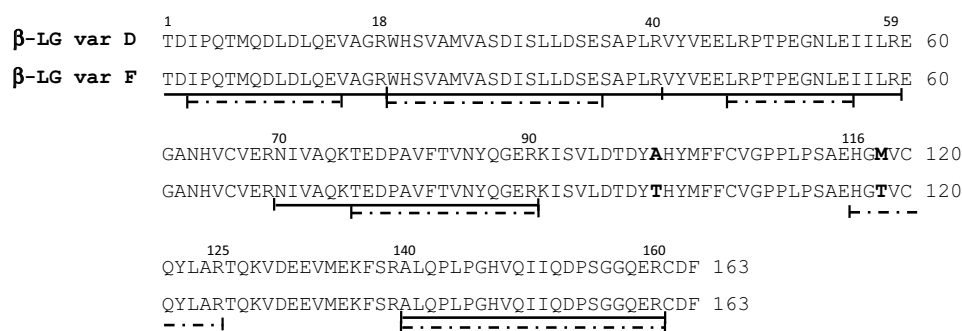


Fig. 3 Alignment of β -LG sequences of variant D and F. The amino acid differences, located at the positions 100 and 118, are reported in bold. Sequence traits characterized by trypsin are underlined by

the continuous line. Sequence traits characterized by the mixture of trypsin and endoproteinase Glu-C are underlined by the dashed line

in the sample investigated, an aliquot of entire whey milk fraction was digested by a mixture (1:1) of porcine trypsin and endoproteinase Glu-C, analyzed by LC/MS/MS, and finally MS data searched against the custom database. MS/MS analysis of the digest obtained by coupling these two enzymes (see Supplementary Table S1 and Figures S13, S14, S15, S16 and S17) confirmed most of the results already obtained by the analysis of the tryptic mixture (Fig. 3). In addition, *de novo interpretation* of the MS/MS spectrum of the doubly charged ion at m/z 602.8018 ($M_{mi} = 1203.5879$ Da) (Fig. 4) permitted to deduce the sequence HGTVCQYLAR. This peptide corresponds to

the amino acidic trait 116–125 of the new β -LG II variant F, which arises by the combined cleavages of endoproteinase Glu-C and trypsin, at level of the glutamic acid 115 and arginine 125, respectively (Fig. 3). It is important to note that this sequence represents a peptide marker of the new variant F, which allows to assert the presence of this novel variant. Indeed, the peptide His¹¹⁶-Arg¹²⁵ of the new variant F carries a threonine residue at position 118, whereas the corresponding peptide of the β -LG II variant D shows a methionine (Fig. 3). In conclusion, even if it was not possible to characterize the entire sequence of the very minor component with an experimental M_{mi} of

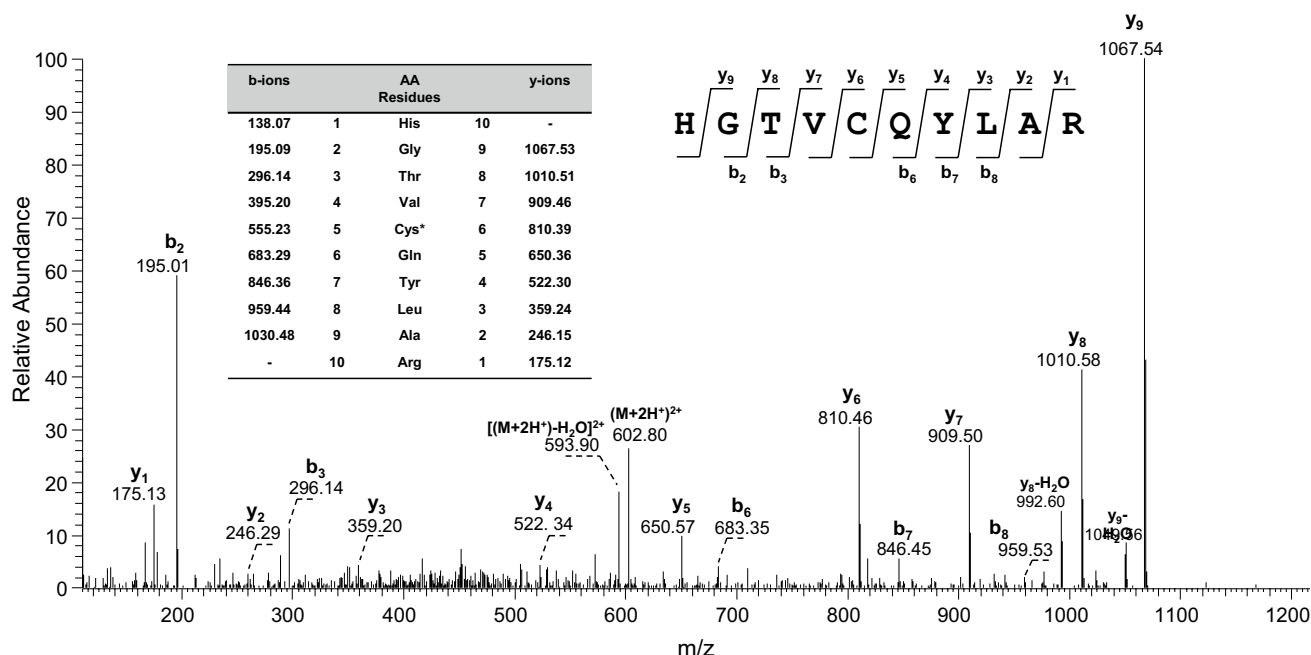


Fig. 4 ESI-MS/MS spectrum of the doubly charged molecular ion at m/z 602.8018 (M_{mi} 1203.5879 Da). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the theoretical m/z values of the b - and y -ion fragments

18,298.959 Da (sequence coverage 68%), the proteomic investigation here used allowed to identify this protein with the predicted new β -LG II F genetic variant, confirming the genomic data.

Discussion and conclusion

In a study carried out by the combined approach of mono-dimensional electrophoresis and mass spectrometric analysis, Criscione et al. (2009) reported a polymorphism at β -LG II fraction with a semi quantitative evidence of differentiation between two main isoelectrophocusing (IEF) patterns, one showing evidence of the protein presence (normal pattern) and one apparently lacking this protein (defective pattern). In the present study, we sampled DNA from jennies accounting for both patterns. Sequence analysis revealed the presence of the already known B and C alleles together with the evidence of a new allele, here named F. In particular, a donkey carrying the F allele at homozygous status produced milk apparently without or with less β -LG II fraction; so that the new allele seems to be associated with a null or a severely reduced expression of this whey protein fraction. With the aim to confirm the genomic data, the milk whey protein fraction obtained from a jenny homozygous FF at β -LG II locus, was characterized by a typical proteomic method (Cunsolo et al. 2014), involving the use of high resolution mass spectrometry at both protein and peptide level. The proteomic approach here applied allowed to confirm the genomic data through the identification, as very minor protein component, of the new predicted β -LG II F genetic variant. It is important to highlight that the confirmation of the genomic data was achieved by joining two methodologies. The first one is related to the unique scan rate, very sensitive high resolution and high mass accuracy of the Orbitrap Fusion Tribrid mass spectrometer used to obtain mass spectral data at both protein and peptide level. The second strategy here adopted which has permitted to achieve definitively the identification of the new predicted β -LG II F was the coupling of two enzymes (i.e. trypsin and endoproteinase Glu-C) for the sequence characterization. Indeed, in this aspect, the peptide marker which allowed the unambiguous identification of the new variant F was obtained only by the contemporary action of porcine trypsin and endoproteinase Glu-C. It is interesting to note that, as previously reported by Chianese et al. (2013), HPLC analysis of whey milk samples homozygous DD at β -LG II locus, showed a low content ($<1\%$) of β -LG II fraction. So that, since only two amino acid substitutions characterized the new variant F in comparison with variant D, [e.g. Thr¹⁰⁰ (variant F) $\rightarrow$ Ala¹⁰⁰ (variant D) and Thr¹¹⁸ (variant F) $\rightarrow$ Met¹¹⁸ (variant D)], a direct phylogenetic origin of F from D variant can be hypothesized. Taking into account these data

and considering that the two observed mutations (c. 352 G $>$ A and c. 539 G $>$ A) in the nucleotide sequence of the new allele F introduce only amino acid substitutions, there is not a direct evidence for the very low expression of the new variant. The low amount of the F variant could be explained by those mechanisms triggered by the specific nucleotide sequence and/or by possible synonym mutations leading to the change of the secondary structure of mRNA and thus causing low protein expression levels (Mitra et al. 2016; McCarthy et al. 2017). Numerous qualitative and quantitative polymorphisms at goat casein *loci* and the related effect of technological and nutritional properties of goat's milk have been widely reported (Marletta et al. 2007; Ballabio et al. 2011). In this light, the identification of jennies producing milk with an extremely low level of β -lactoglobulin II appears promising. Indeed, quantitative variation has shown to have an effect on the digestibility of β -LG II fraction, enhancing the protein degradation particularly during the gastric digestion step (Tidona et al. 2014). An appreciable range of quantitative variation of β -LG ($1.3\text{--}5.5\text{ g L}^{-1}$) has been reported in the literature (Brumini et al., 2016). This aspect could also significantly affect the nutritional and allergenic properties of this milk. In fact, since β -LG can elicit an allergic reaction in sensitive subjects (Restani et al. 2009), the identification of animals producing milk having very low β -LG II protein appears promising to further reduce the allergenic capacity of individual donkey's milk and solve potential residual cases of reactivity (Wal 2002).

Acknowledgements This work was supported by grants from PO FERS 2007/13 4.1.2.A, project "Piattaforma regionale di ricerca translazionale per la salute", CUP B65E12000570008. The authors gratefully acknowledge the Bio-Nanotech Research and Innovation Tower (BRIT; PON project financed by the Italian Ministry for Education, University and Research, MIUR) for the availability of the Orbitrap Fusion mass spectrometer.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

References

- Ballabio C, Chessa S, Rignanese D, Gigliotti C, Pagnacco G, Terracciano L, Fiocchi A, Restani P, Caroli M (2011) Goat milk allergenicity as a function of α S1-casein genetic polymorphism. *J Dairy Sci* 94:998–1004. <https://doi.org/10.3168/jds.2010-3545>
- Bertino E, Gastaldi D, Monti G, Baro C, Fortunato D, Perono Garoffo L, Coscia A, Fabris C, Mussap M, Conti A (2010) Detailed proteomic analysis on DM: insight into its hypoallergenicity. *Front Biosci* 2:526–536

- Brumini D, Criscione A, Bordonaro S, Vegarud GE, Marletta D (2016) Whey proteins and their antimicrobial properties in donkey milk: a brief review. *Dairy Sci Technol* 96:1–14. <https://doi.org/10.1007/s13594-015-0246-1>
- Chianese L, De Simone C, Ferranti P, Mauriello R, Costanzo A, Quarto M, Garro G, Picariello G, Mamone G, Ramunno L (2013) Occurrence of qualitative and quantitative polymorphism at donkey betalactoglobulin II locus. *Food Res Int* 54:1273–1279. <https://doi.org/10.1016/j.foodres.2012.11.005>
- Claeys WL, Verraes C, Cardoen S, De Block J, Huyghebaert A, Raes K, Dewettinck K, Herman L (2014) Consumption of raw or heated milk from different species: an evaluation of the nutritional and potential health benefits. *Food Control* 42:188–201
- Conti A, Godovac-Zimmermann J, Liberatori J, Braunitzer G (1984) The primary structure of monomeric beta-lactoglobulin I from horse colostrum (*Equus caballus*, *Perissodactyla*). *Hoppe Seylers Z Physiol Chem* 365:1393–1401
- Criscione A, Cunsolo V, Bordonaro S, Guastella AM, Saletti R, Zucaro A, D'Urso G, Marletta D (2009) Donkeys' milk protein fraction investigated by electrophoretic methods and mass spectrometric analysis. *Int Dairy J* 19:190–197. <https://doi.org/10.1016/j.idairyj.2008.10.015>
- Cunsolo V, Costa A, Saletti R, Muccilli V, Foti S (2007) Detection and sequence determination of a new variant beta-lactoglobulin II from donkey. *Rapid Commun Mass Spectrom* 21:1438–1446. <https://doi.org/10.1002/rcm.2978>
- Cunsolo V, Cairone E, Fontanini D, Criscione A, Muccilli V, Saletti R, Foti S (2009a) Sequence determination of alpha(s1)-casein isoforms from donkey by mass spectrometric methods. *J Mass Spectrom* 44:1742–1753. <https://doi.org/10.1002/jms.1683>
- Cunsolo V, Cairone E, Saletti R, Muccilli V, Foti S (2009b) Sequence and phosphorylation level determination of two donkey beta-caseins by mass spectrometry. *Rapid Commun Mass Spectrom* 23:1907–1916. <https://doi.org/10.1002/rcm.4087>
- Cunsolo V, Muccilli V, Saletti R, Foti S (2013) MALDI-TOF mass spectrometry for the monitoring of she-donkey's milk contamination or adulteration. *J Mass Spectrom* 48:148–153
- Cunsolo V, Muccilli V, Saletti R, Foti S (2014) Mass spectrometry in food proteomics: a tutorial. *J Mass Spectrom* 49:768–784. <https://doi.org/10.1002/jms.3374>
- Cunsolo V, Saletti R, Muccilli V, Gallina S, Di Francesco A, Foti S (2017) Proteins and bioactive peptides from donkey milk: the molecular basis for its reduced allergenic properties. *Food Res Int*. <https://doi.org/10.1016/j.foodres.2017.07.002>
- European Commission. 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. In: Official Journal of the European Union. L 276/33. 20.10.2010
- Godovac-Zimmermann J, Conti A, Liberatori J, Braunitzer G (1985) The amino-acid sequence of beta-lactoglobulin II from horse colostrum (*Equus caballus*, *Perissodactyla*): beta-lactoglobulins are retinol-binding proteins. *Biol Chem Hoppe Seyler* 366:601–608
- Godovac-Zimmermann J, Conti A, James L, Napolitano L (1988) Microanalysis of the amino-acid sequence of monomeric beta-lactoglobulin I from donkey (*Equus asinus*) milk. *Biol Chem H-S* 369:171–179
- Godovac-Zimmermann J, Conti A, Sheil M, Napolitano L (1990) Covalent structure of the minor monomeric β -lactoglobulin II component from donkey milk. *Biol Chem H-S* 371:871–879
- Goraczko AJ (2005) Molecular mass and location of the most abundant peak of the molecular ion isotopomeric cluster. *J Mol Model* 11:271–277
- Herrouin M, Mollé D, Faunquant J, Ballestra F, Maubois JL, Léonil J (2000) New variants identified in donkey's milk whey proteins. *J Protein Chem* 19:105–115
- Horn DM, Zubarev RA, McLafferty FW (2000) Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom* 11:320–332
- Malacarne M, Criscione A, Franceschi P, Tumino S, Bordonaro S, Di Frangia F, Marletta D, Summer A (2017) Distribution of Ca, P and Mg and casein micelle mineralisation in donkey milk from the second to ninth month of lactation. *Int Dairy J* 66:1–5
- Mansueto P, Iacono G, Taormina G, Seidita A, D'Alcamo A, Adragna F, Randazzo G, Carta Rini G, Carroccio A (2013) Ass's milk in allergy to cow's milk protein: a review. *Acta Med Mediterr* 29:153–160
- Marletta D, Criscione A, Bordonaro S, Guastella AM, D'Urso G (2007) Casein polymorphism in goat's milk. *Le Lait* 87:491–504
- Marletta D, Tidona F, Bordonaro S (2016) Donkey milk proteins: digestibility and nutritional significance. In: "Milk Proteins—From Structure to Biological Properties and Health Aspects" pp 199–209. <https://doi.org/10.5772/62597>
- McCarthy C, Carrea A, Diambra L (2017) Bicondon bias can determine the role of synonymous SNPs in human diseases. *BMC Genom* 18:227. <https://doi.org/10.1186/s12864-017-3609-6>
- Medhammar E, Wijesinha-Bettoni R, Stadlmayr B, Nilsson E, Charroindiere UR, Burlingame B (2012) Composition of milk from minor dairy animals and buffalo breeds: a biodiversity perspective. *J Sci Food Agr* 92:445–474. <https://doi.org/10.1002/jsfa.4690>
- Mitra S, Ray SK, Banerjee R (2016) Synonymous codons influencing gene expression in organisms. *Rese Rep Biochem* 6:57–65
- Monti G, Viola S, Baro C, Cresi F, Tovo PA, Moro G, Ferrero MP, Conti A, Bertino E (2012) Tolerability of donkey's milk in 92 highly-problematic cow's milk allergic children. *J Biol Regul Homeost Agents* 26:75–82
- Ochirkhuyag B, Chobert JM, Dalgalarrodo M, Haertlé T (2000) Characterization of mare caseins. Identification of α s1- and α s2-caseins. *Lait* 80:223–235
- Restani P, Ballabio C, Di Lorenzo C (2009) Molecular aspects of milk allergens and their role in clinical events. *Anal Bioanal Chem* 395:47–56
- Saletti R, Muccilli V, Cunsolo V, Fontanini D, Capocchi A, Foti S (2012) MS-based characterization of as2-casein isoforms in donkey's milk. *J Mass Spectrom* 47:1150–1159. <https://doi.org/10.1002/jms.3031>
- Salimei E, Fantuz F (2012) Equid milk for human consumption. *Int Dairy J* 24:130–142. <https://doi.org/10.1016/j.idairyj.2011.11.008>
- Senko MW, Beu SC, McLafferty FW (1995) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom* 6:229–233
- Skoog B, Wichman A (1986) Calculation of the isoelectric points of polypeptides from the amino acid composition. *Trends Anal Chem* 5:82–83
- Swaigood HE (1995) Protein and amino acid composition of bovine milk. In: Jensen RC (ed) *Handbook of milk composition*. Academic Press Inc., San Diego, pp 464–468
- Swar MO (2011) Donkey milk-based formula: a substitute for patients with cow's milk protein allergy. *Sudan J Paediatr* 11:21–24
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tidona F, Sekse C, Criscione A, Jacobsen M, Bordonaro S, Marletta D, Vegarud GE (2011) Antimicrobial effect of donkeys' milk digested in vitro with human gastrointestinal enzymes. *Int Dairy J* 21:158–165. <https://doi.org/10.1016/j.idairyj.2010.10.008>
- Tidona F, Criscione A, Devold TG, Bordonaro S, Marletta D, Vegarud GE (2014) Protein composition and micelle size of donkey milk with different protein patterns: effects on digestibility. *Int Dairy J* 35:57–62. <https://doi.org/10.1016/j.idairyj.2013.10.018>

- Vincenzetti S, Foghini L, Pucciarelli S, Polzonetti Cammertoni N, Beghelli D, Polidori P (2014) Hypoallergenic properties of donkey's milk: a preliminary study. *Vet Ital* 50:99–107. <https://doi.org/10.12834/VetIt.219.125.5>
- Wal JM (2002) Cow's milk proteins/allergens. *Ann Allergy Asthma Immunol* 89:3–10. [https://doi.org/10.1016/S1081-1206\(10\)62115-1](https://doi.org/10.1016/S1081-1206(10)62115-1)
- Waserman S, Watson W (2011) Food allergy. *Allergy Asthma Clin Immunol* 7(Suppl 1):57. <https://doi.org/10.1186/1710-1492-7-S1-S7>

SUPPLEMENTARY INFORMATION

Polymorphism at donkey -Lactoglobulin II *locus*: identification and characterization of a new variant with a very low expression

Polymorphism at donkey -Lactoglobulin II *locus*: identification and characterization of a new variant with a very low expression

Andrea Criscione¹, Vincenzo Cunsolo^{2*}, Serena Tumino¹, Antonella Di Francesco²,
Salvatore Bordonaro¹, Vera Muccilli¹, Rosaria Saletti², Donata Marletta¹

¹Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy

²Agricultural, Food and Environment, Animal Production and Phytopathology Section, University of Catania, Via Valdisavoia 5, 95123 Catania, Italy

Keywords: polymorphism, BGLII gene, donkey milk, allergy, mass spectrometry, SNPs

* Corresponding Author:

Vincenzo Cunsolo,

Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy.

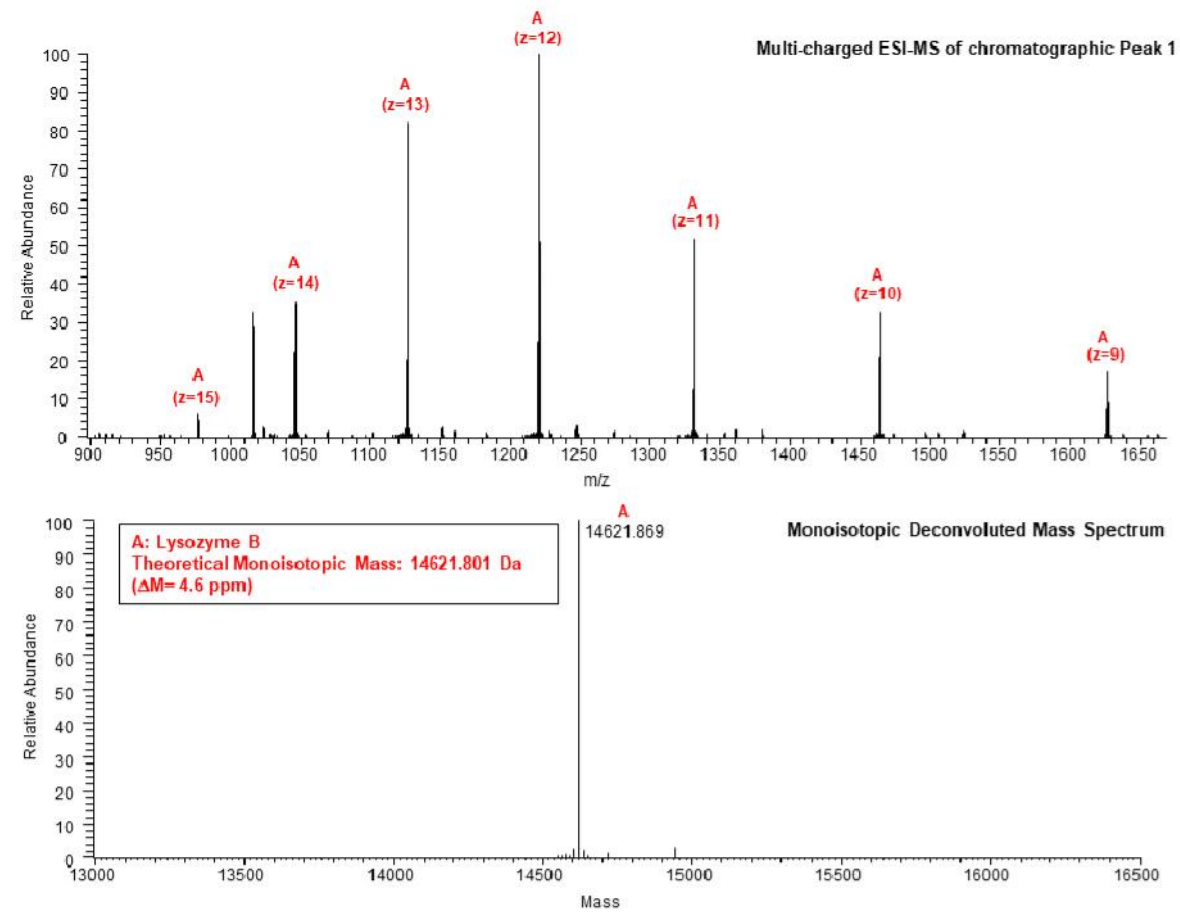
Tel. 0039 095 7385029

Fax 0039 095 580138

E-mail: vcunsolo@unict.it

SUPPLEMENTARY INFORMATIONS

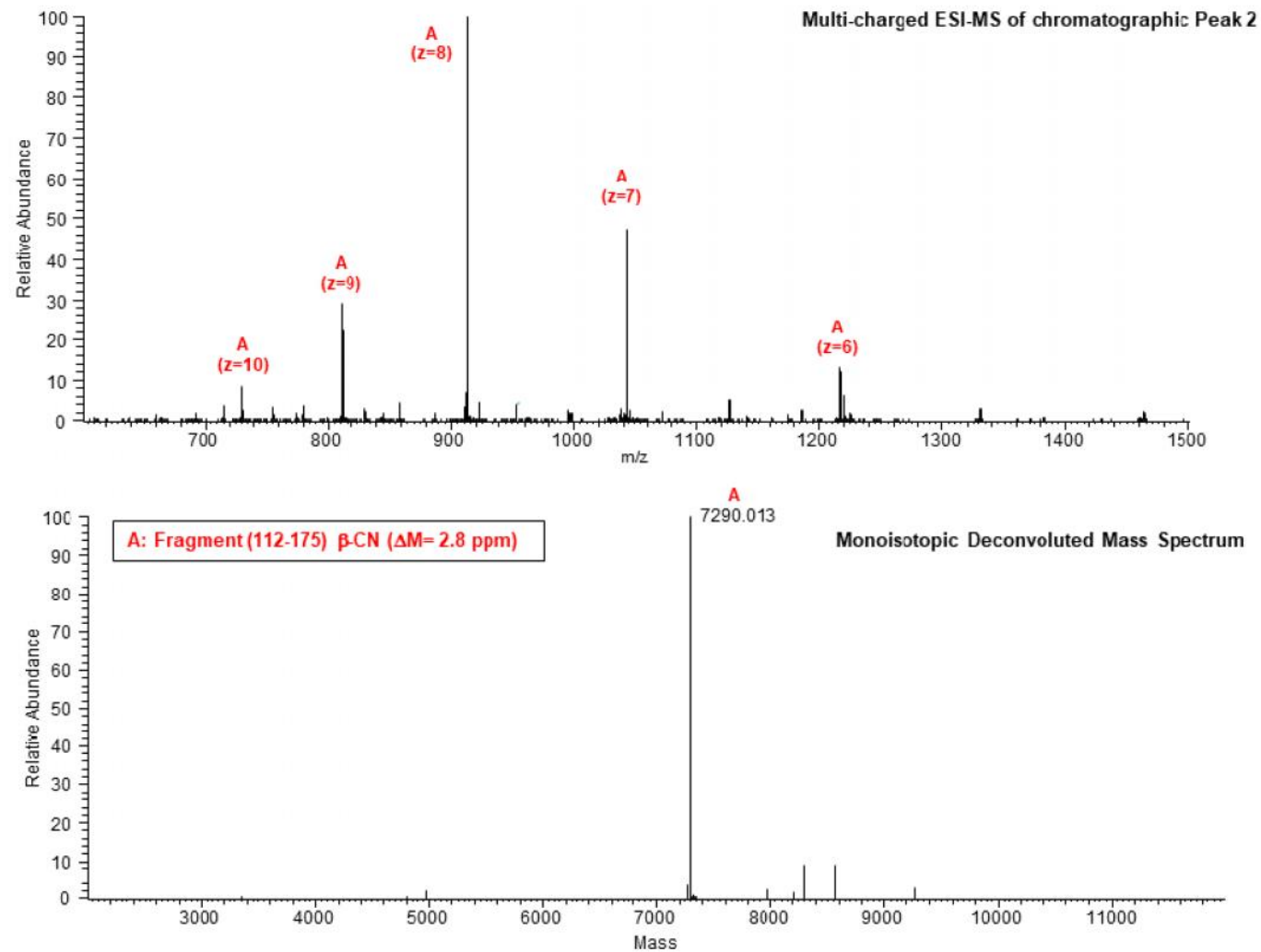
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S1. Multi-charged ESI mass spectrum and monoisotopic deconvoluted mass spectrum of chromatographic Peak 1.

SUPPLEMENTARY INFORMATIONS

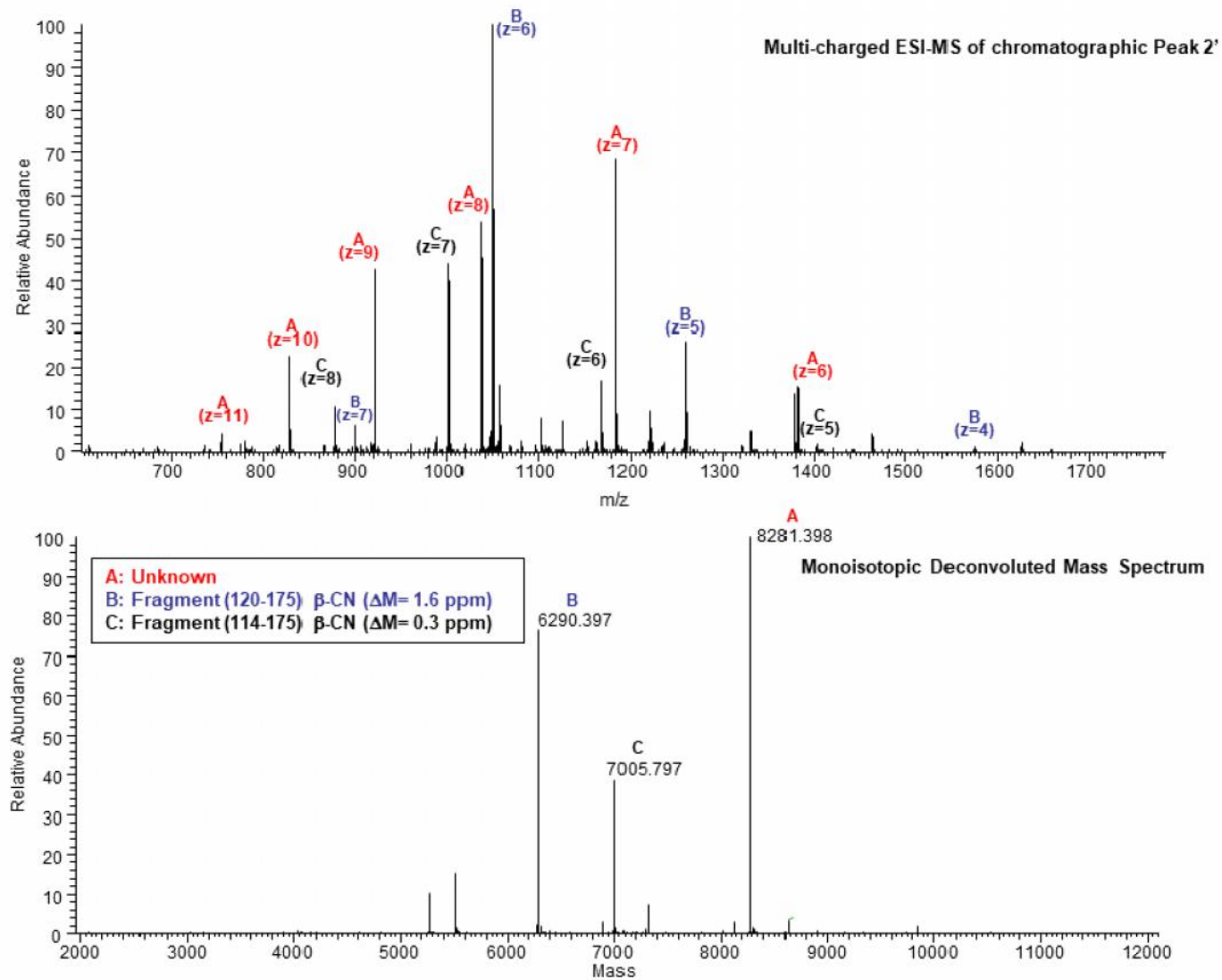
Polymorphism at donkey β -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S2. Multi-charged ESI mass spectrum and monoisotopic deconvoluted mass spectrum of chromatographic Peak 2

SUPPLEMENTARY INFORMATIONS

Polymorphism at donkey β -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S3. Multi-charged ESI mass spectrum and monoisotopic deconvoluted mass spectrum of chromatographic Peak 2'.

SUPPLEMENTARY INFORMATIONS

Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression

```

REKEELNVSSETVESLSSNEPDSSSEESITHINKEKSQKFKHEGQQQREVEHQDKISRfV    60
QPQPVVYPYAEPVPYAVVPQNILVLAQPPIVPFLQPEIMEVSQAKETILPKRKVMPFLKS    120
PIVPFSERQILNPTNGENLRLPVHLIQPFMHQVPQSLLQTLMLPSQPVLSPPQSKVAPFP    180
QPVVPYPQRDTPVQAFLLYQDPQLGLTGEFDPATQPIVPVHNpVIV

```

Arrows indicate Lys-X bonds cleaved by plasmin at positions 111, 113, 119, and 175.

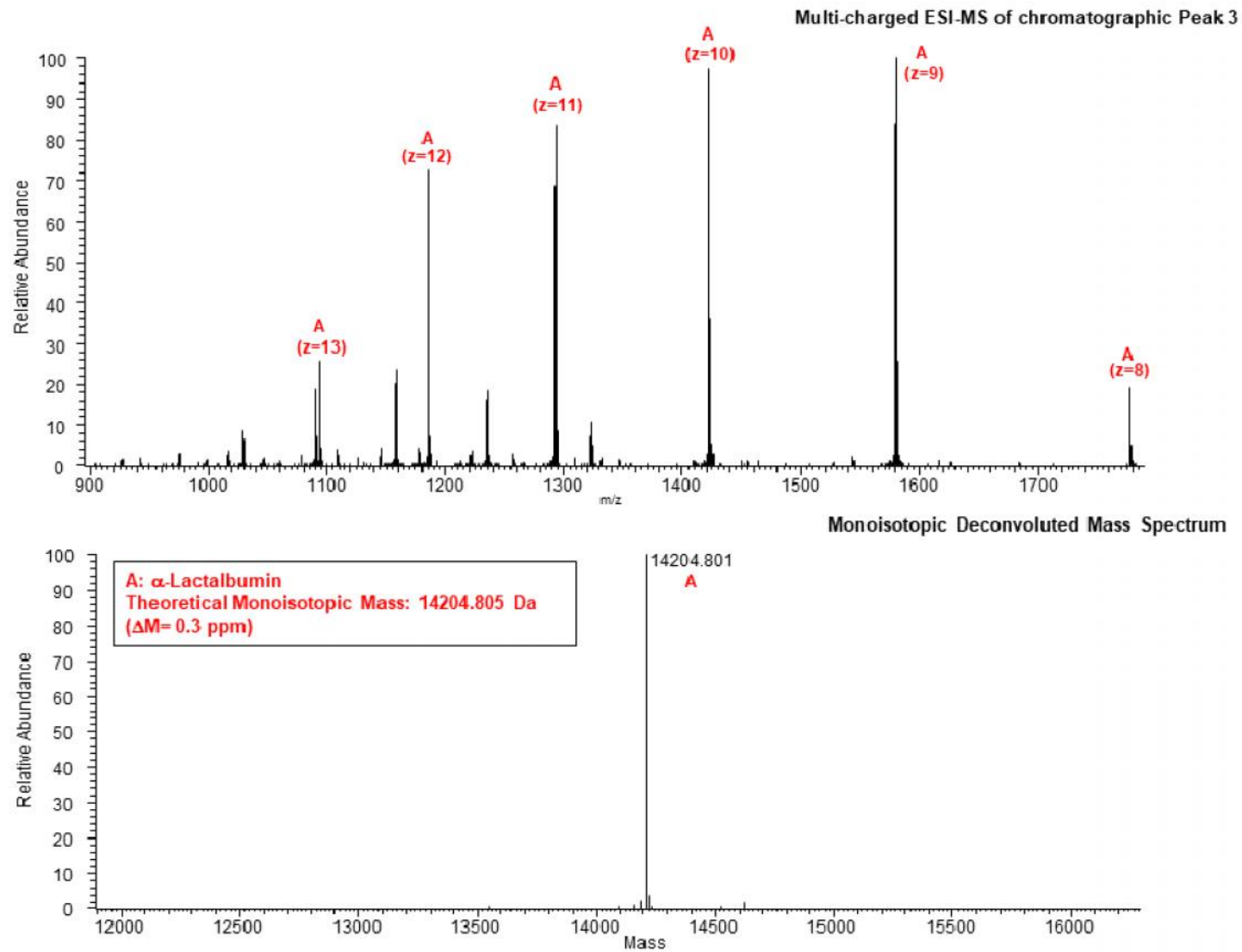
Supplementary Figure S4. Amino acid sequence of donkey β -CN (UniProt Acc. No. P86273). The Lys-X bonds cleaved by the action of plasmin and originating the polypeptide fragments corresponding to the components detected in peaks 2 and 2' are indicated by an arrow. As known, caseins show a very flexible tertiary structure and therefore are easily susceptible to proteolysis. In this respect, the β -CN appears to be the most susceptible to the action of the endogenous milk protease plasmin. Plasmin preferentially cleaves at Lys-X bonds producing a series of complementary C-terminal and N-terminal polypeptide fragments which are soluble in acidic condition and therefore can be found in the whey protein fraction [1, 2, 3].

Polypeptide Fragment	Theoretical M_{mi} (Da)	Experimental deduced (Da) M_{mi}	aM error (ppm)	Chromatographic Peak
β -CN f(Arg ¹¹² -Lys ¹⁷⁵)	7289.992	7290.013	2.8	2
β -CN f(Val ¹¹⁴ -Lys ¹⁷⁵)	7005.795	7005.797	0.3	2'
β -CN f(Ser ¹²⁰ -Lys ¹⁷⁵)	6290.387	6290.397	1.6	2'
Unknown	-	8281.398	-	2'

-
- [1] Eigel WN, Hofmann CJ, Chibber BA, Tomich JM, Keenan TW, Mertz ET (1979) Plasmin-mediated proteolysis of casein in bovine milk. Proc Natl Acad Sci USA, 76:2244.
[2] Bastian ED, Brown RJ (1996) Plasmin in milk and dairy products: an update. Int Dairy J, 6:435.
[3] Andrews AT (1983) Proteinases in normal bovine milk and their action on caseins. J Dairy Res, 50:45.

SUPPLEMENTARY INFORMATIONS

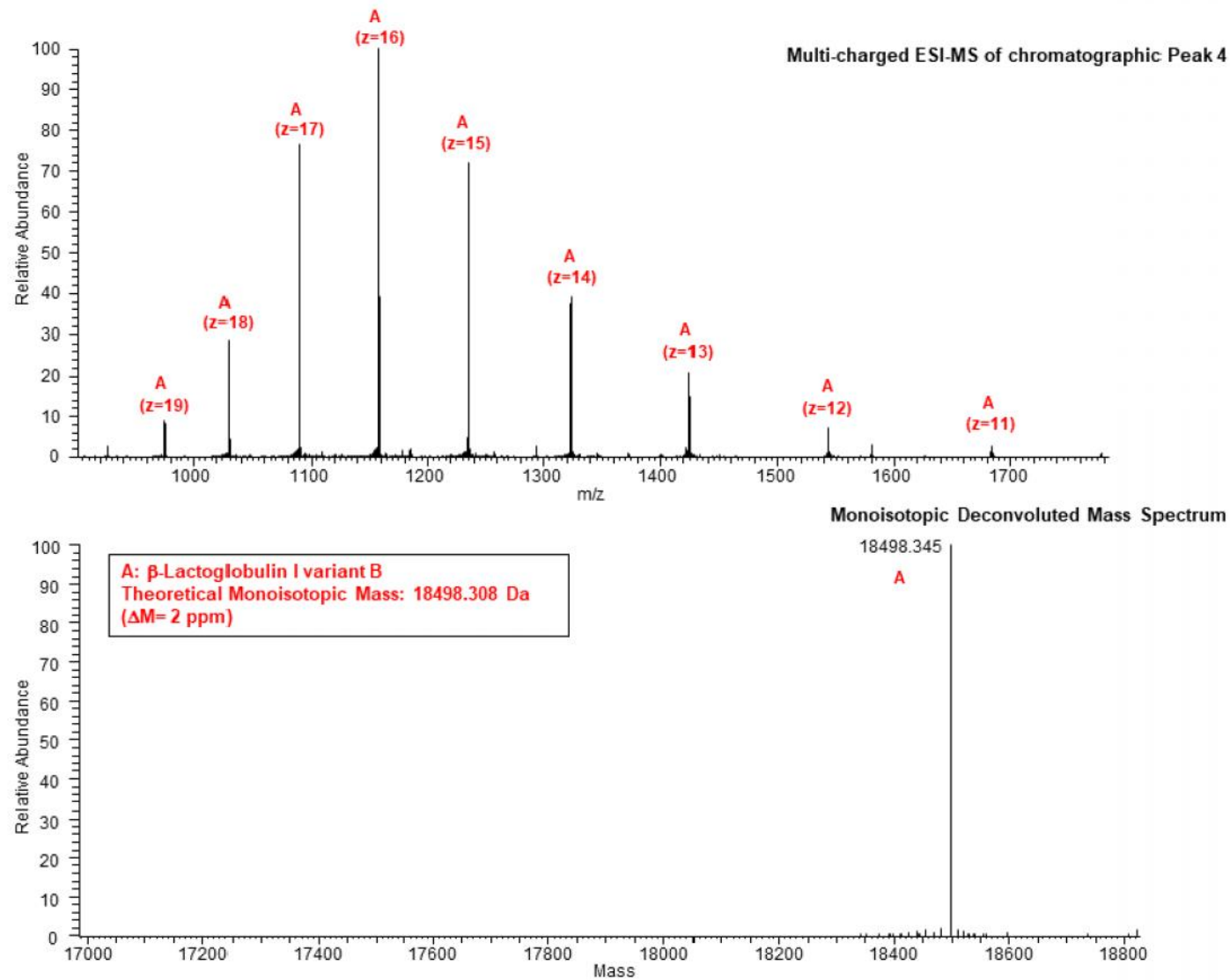
Polymorphism at donkey α -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S5. Multi-charged ESI mass spectrum and monoisotopic deconvoluted mass spectrum of chromatographic Peak 3.

SUPPLEMENTARY INFORMATION

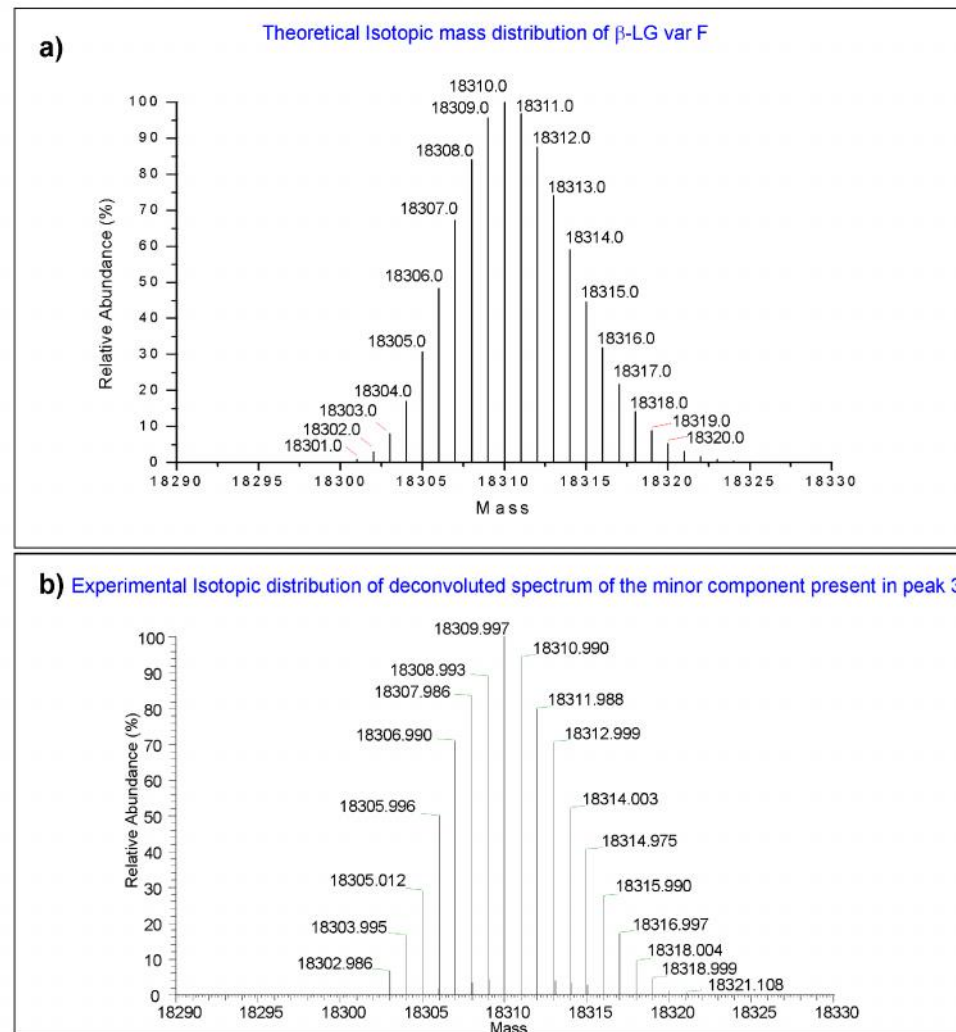
Polymorphism at donkey β -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S6. Multi-charged ESI mass spectrum and monoisotopic deconvoluted mass spectrum of chromatographic Peak 4.

SUPPLEMENTARY INFORMATIONS

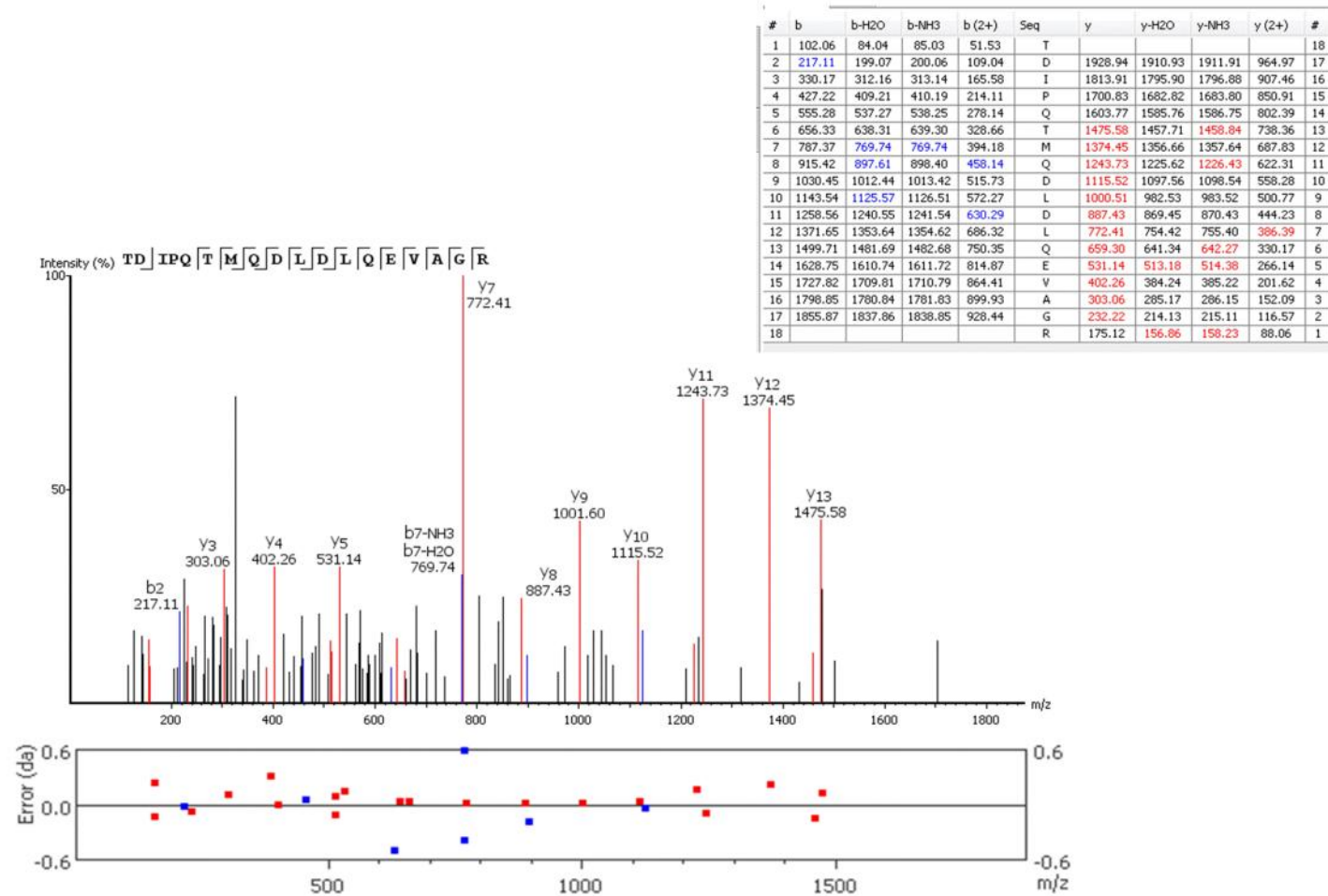
Polymorphism at donkey β -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S7. Comparison of the (a) theoretical isotopic mass distribution of β -LG F, obtained using the GPMW software, and (b) the experimentally determined isotopic distribution obtained by deconvolution of the ESI multi-charged pattern observed for the minor component eluting in the chromatographic Peak 3'

SUPPLEMENTARY INFORMATIONS

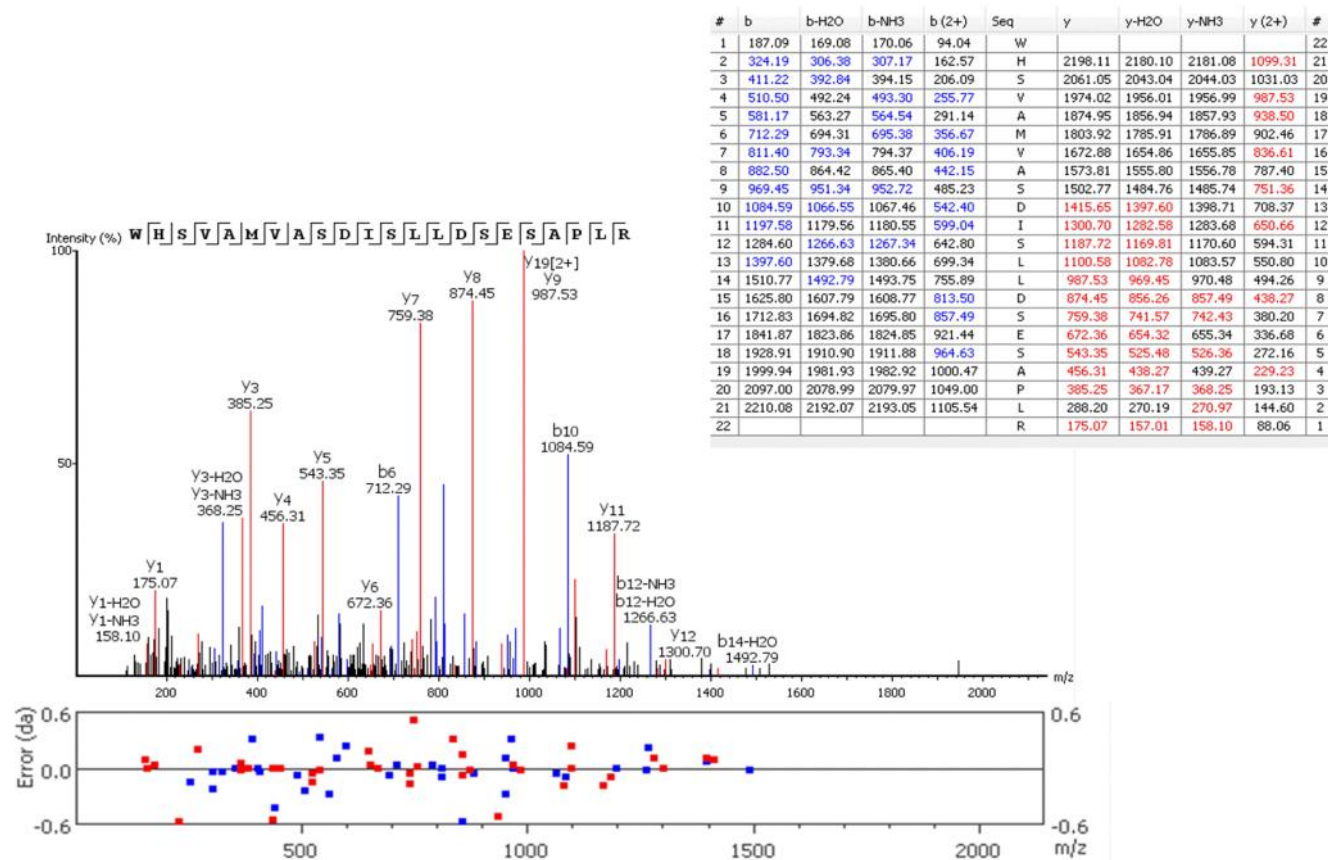
Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S8. ESI-MS/MS spectrum of the doubly-charged molecular ion at m/z 1015.4958 (see Table 5 in the paper). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

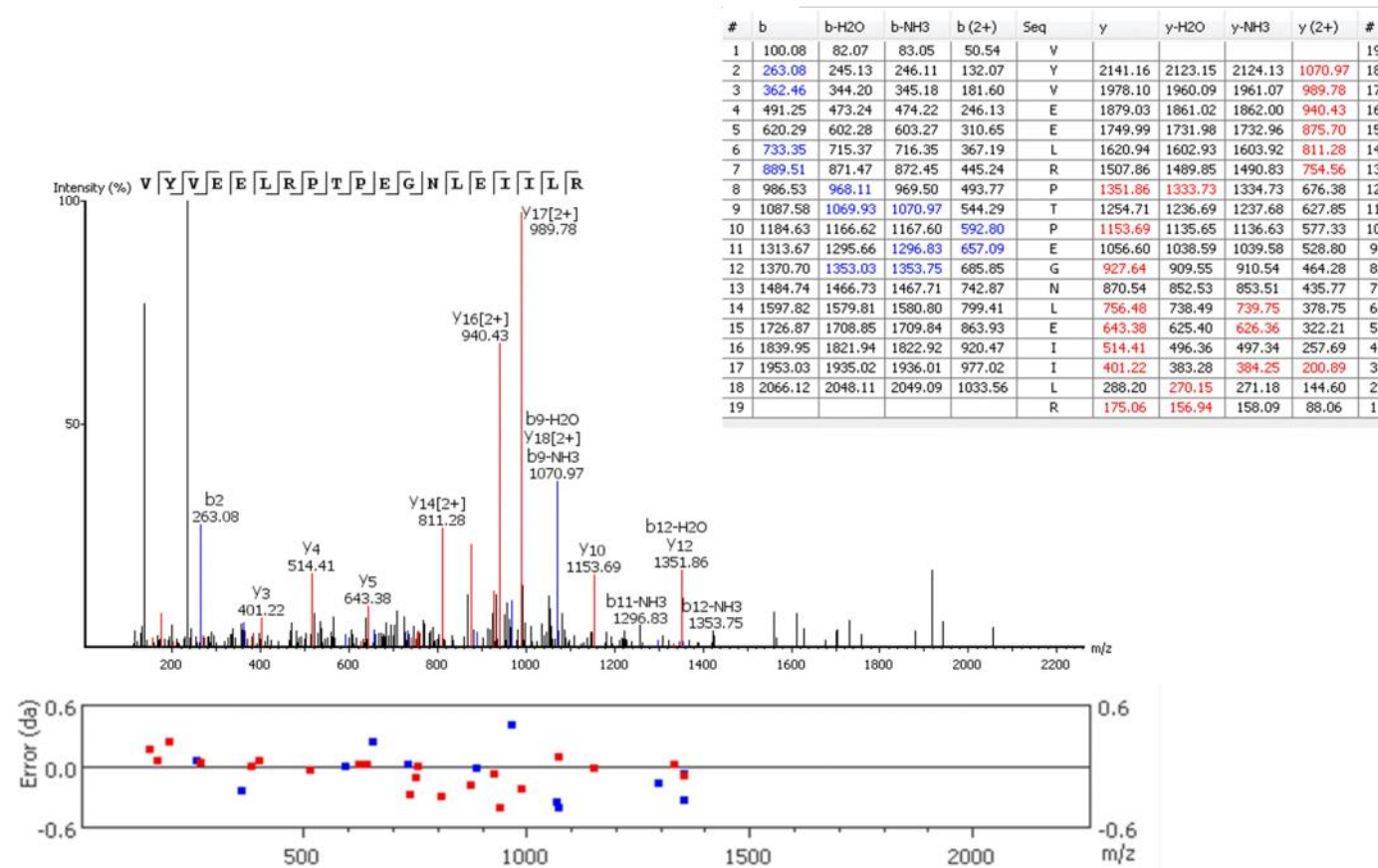
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S9. ESI-MS/MS spectrum of the triply-charged molecular ion at m/z 745.9013 (see Table 5 in the paper). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

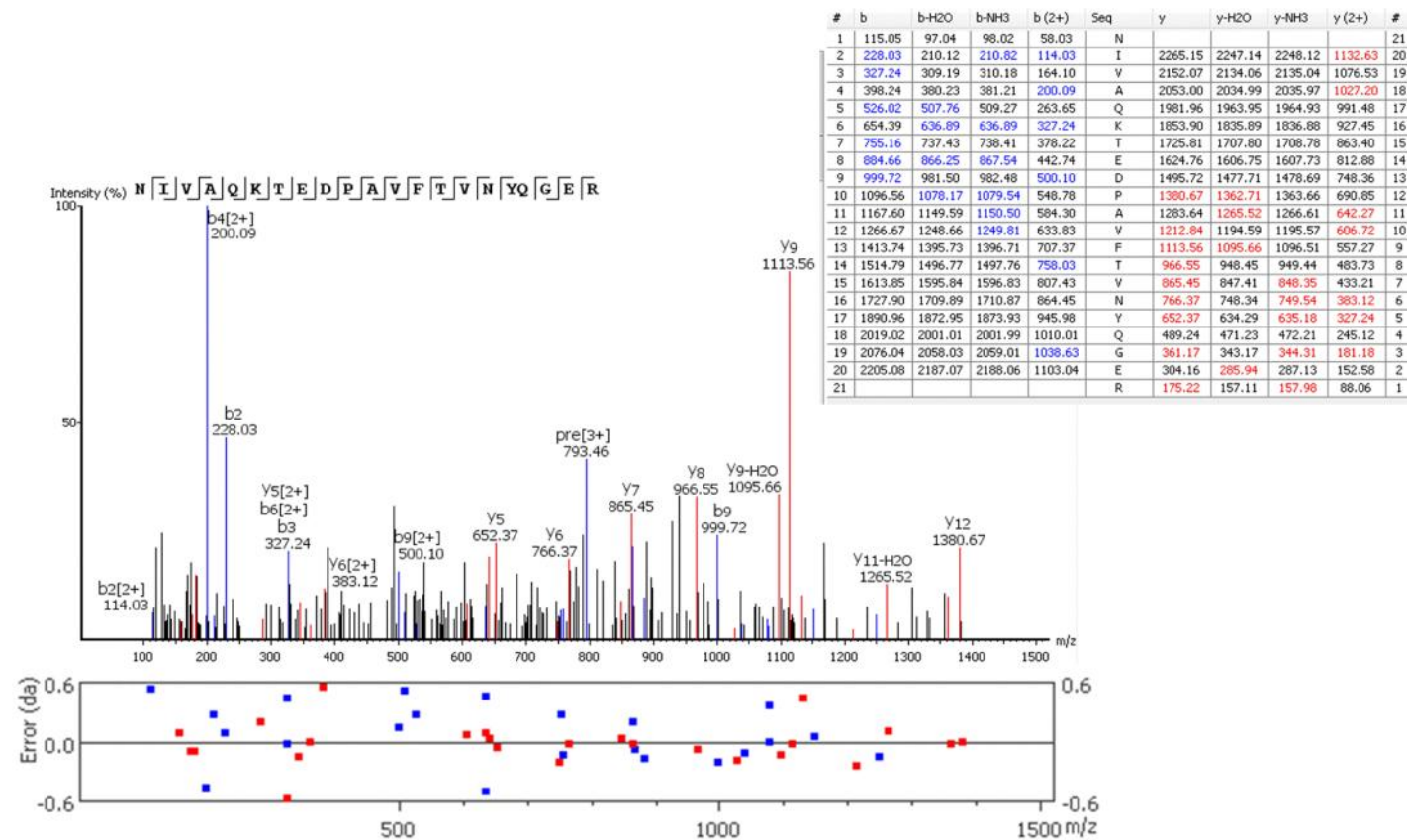
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S10. ESI-MS/MS spectrum of the triply-charged molecular ion at m/z 747.4136 (see Table 5 in the paper). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

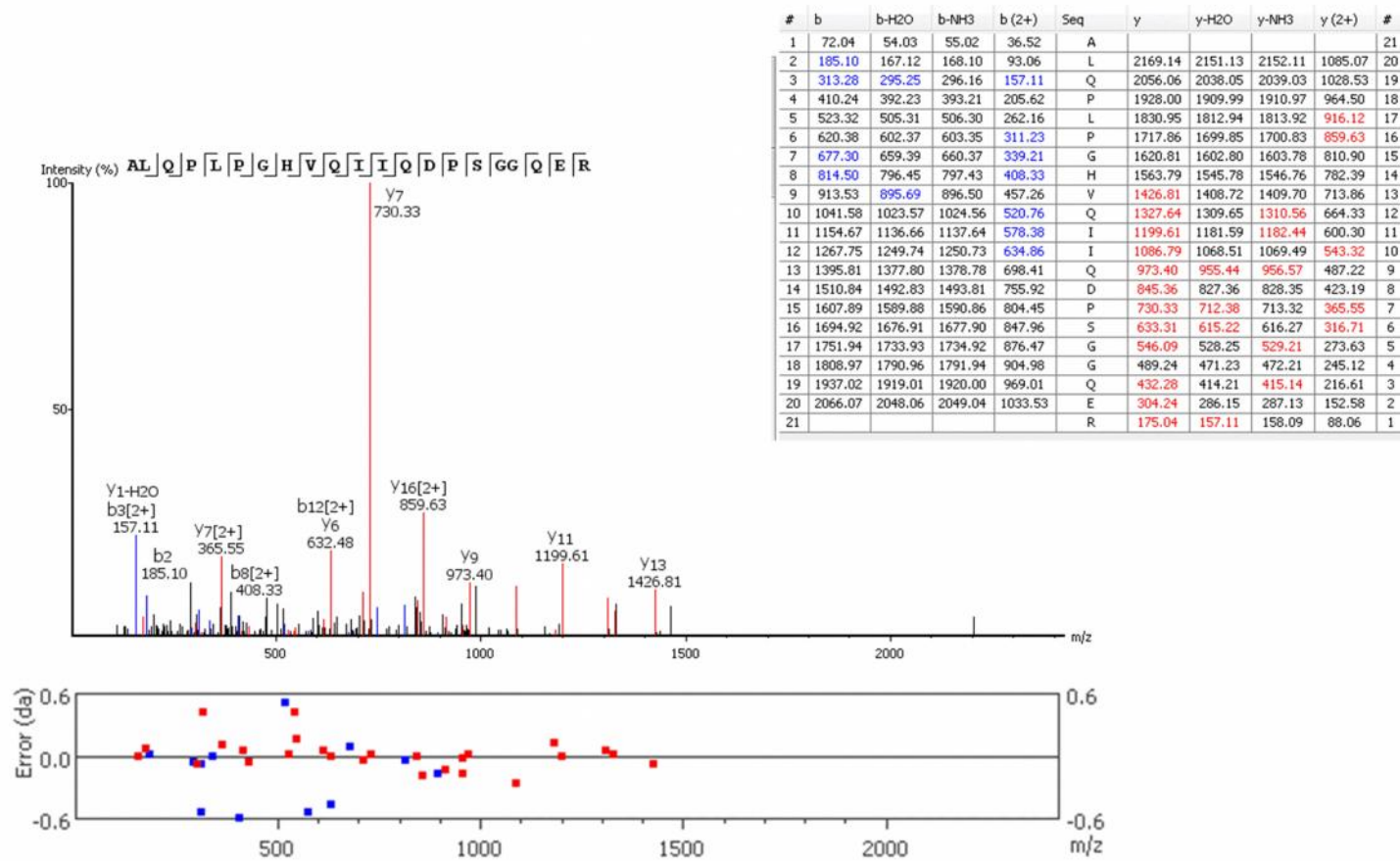
Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S11. ESI-MS/MS spectrum of the triply-charged molecular ion at m/z 793.7373 (see Table 5 in the paper). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S12. ESI-MS/MS spectrum of the triply-charged molecular ion at m/z 747.3977 (see Table 5 in the paper). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

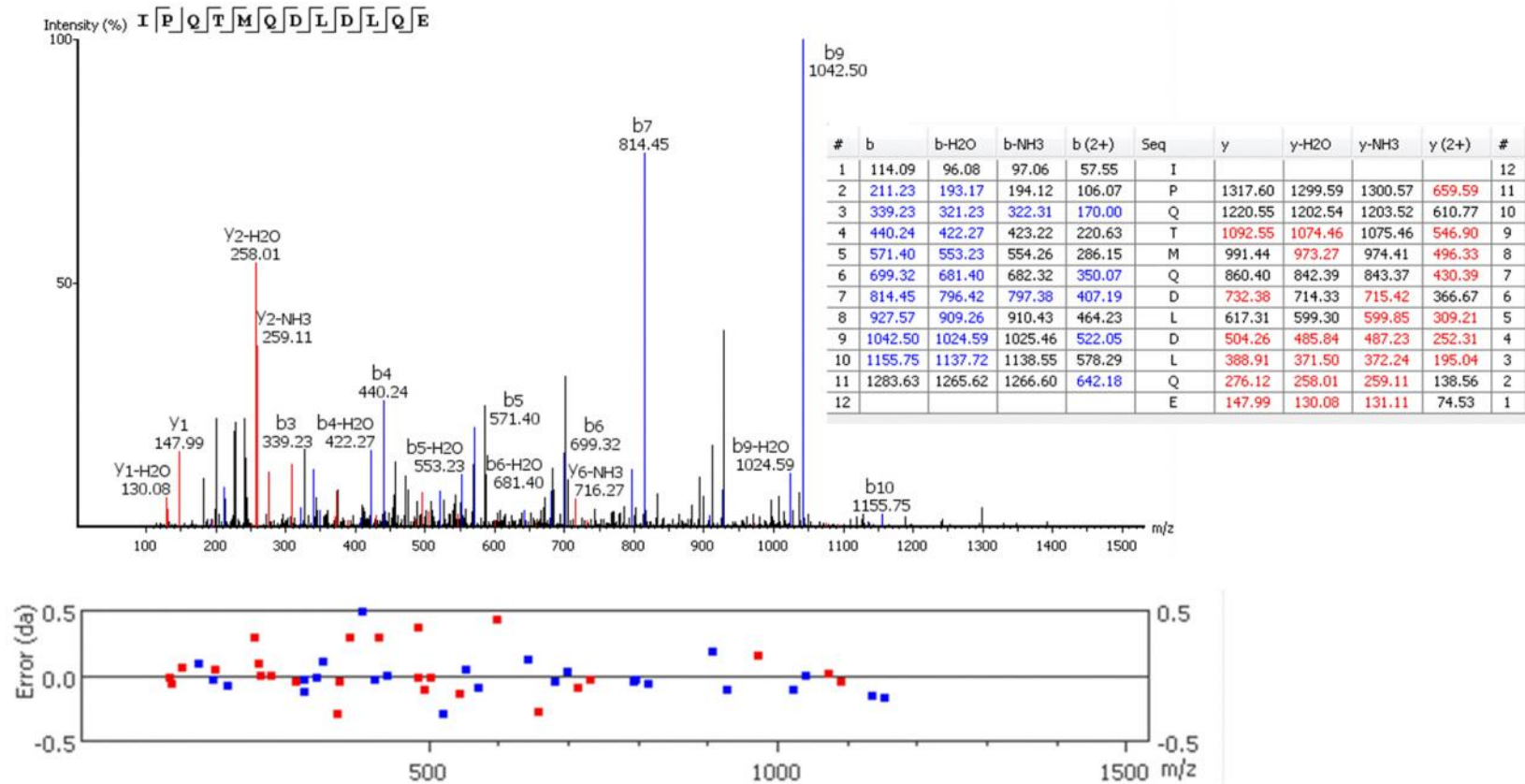
Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression

Table S1. Amino acid position, theoretical monoisotopic mass, theoretical monoisotopic doubly- or triply-charged ions, experimentally measured mass and error of enzymatic fragments from γ -LG eluting in peaks 3' are reported (Figure 1). The cDNA derived sequence of γ -LG variant F (Table 2) was taken as reference. The listed peptides were obtained using as cleavage enzyme a mixture 1:1 of porcine trypsin and endoproteinase Glu-C. This Table doesn't include the fragment 116-125, which is described in the paper and reported in Table 5.

Position	Theoretical monoisotopic mass	Theoretical monoisotopic m/z (z)	Experimental m/z (z)	Error (ppm)	Reference Sequence (cDNA derived sequence of γ -LG variant F)
3-14	1429.6759	715.8452 (2)	715.8466 (2)	1.9	IPQTMQDLDLQE
19-35	1858.8771	930.4458 (2)	930.4327 (2)	-14.1	WHSVAMVASDISLLDSE
46-55	1124.5826	563.2986 (2)	563.3050 (2)	11.3	LRPTPEGNLE
76-90	1724.8006	863.4076 (2)	863.3971 (2)	-12.2	TEDPAVFTVNYQGER
140-160	2239.1709	747.3976 (3)	747.3972 (3)	-0.5	ALQPLPGHVQIIQDPSGGQER

SUPPLEMENTARY INFORMATIONS

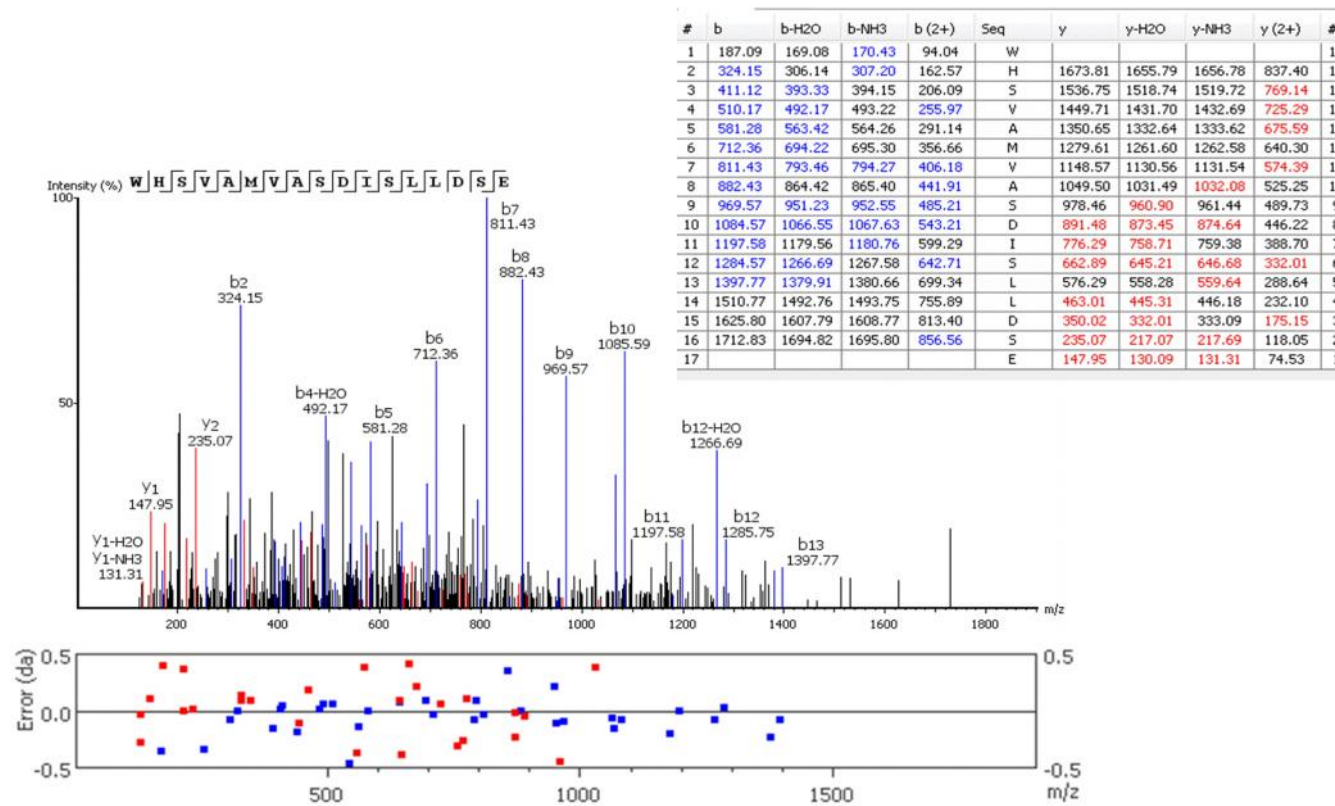
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S13. ESI-MS/MS spectrum of the doubly-charged molecular ion at m/z 715.8466 (see Table S1). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

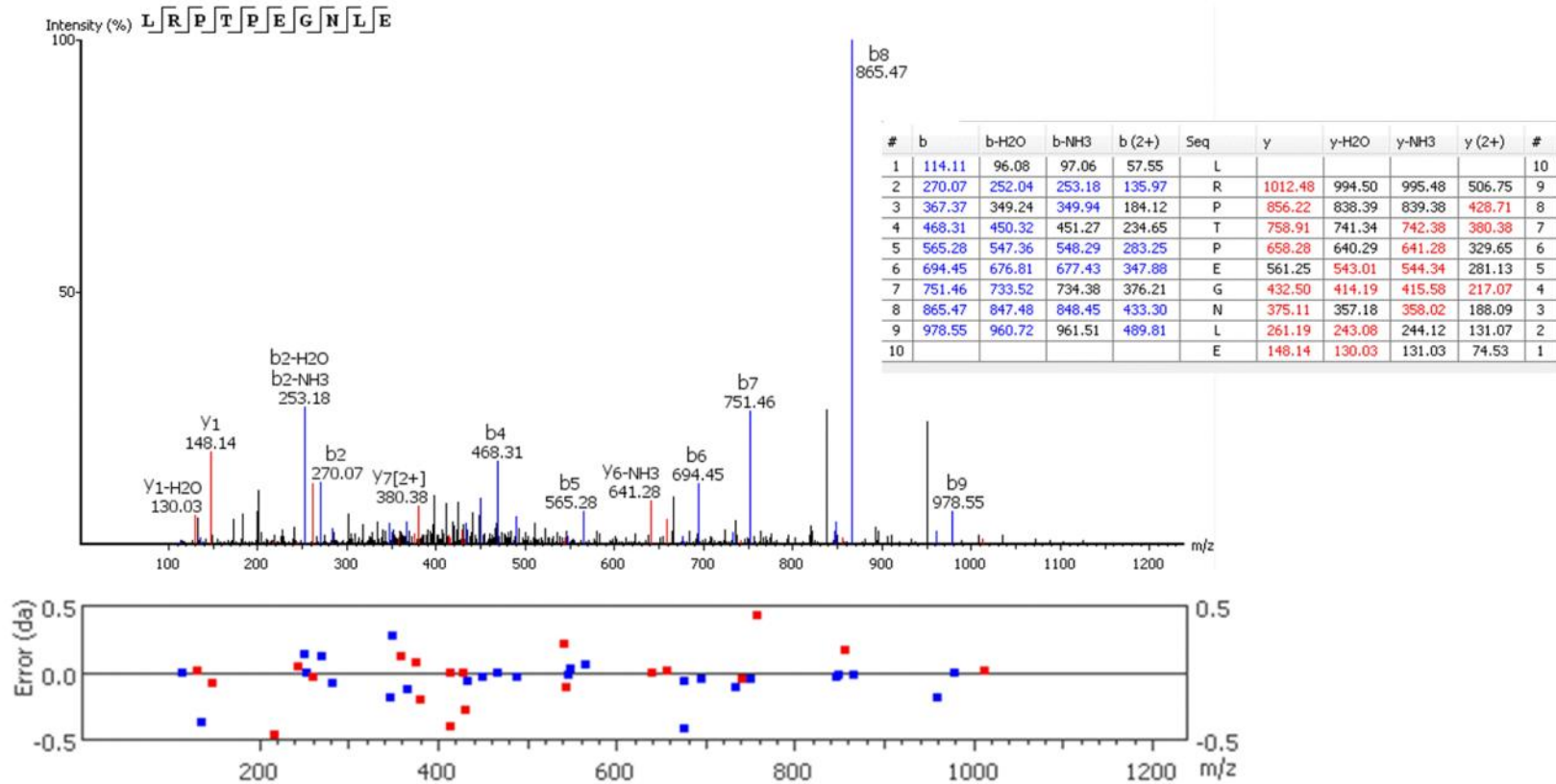
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S14. ESI-MS/MS spectrum of the doubly-charged molecular ion at m/z 930.4327 (see Table S1). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

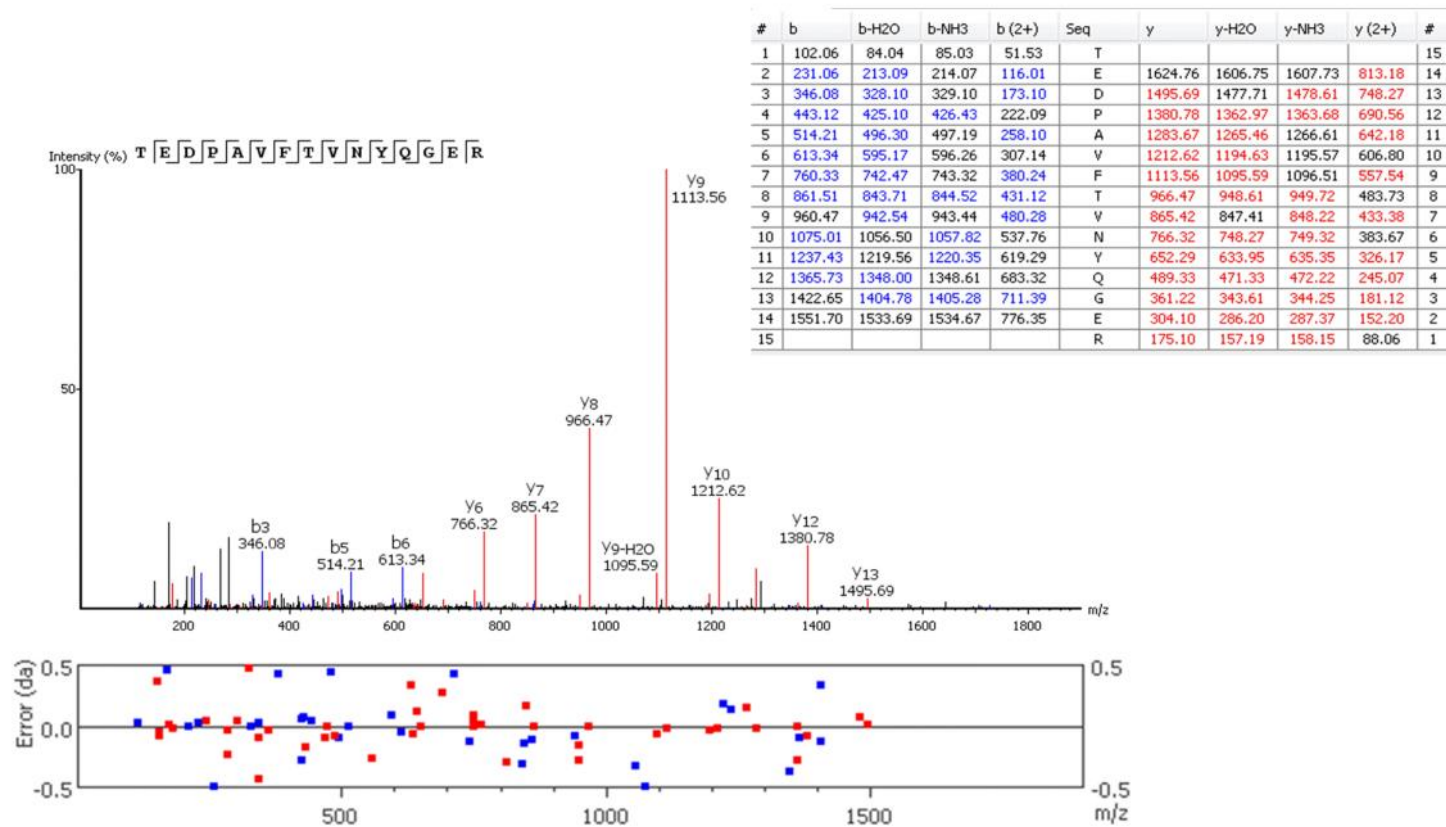
Polymorphism at donkey -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S15. ESI-MS/MS spectrum of the doubly-charged molecular ion at m/z 563.3050 (see Table S1). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

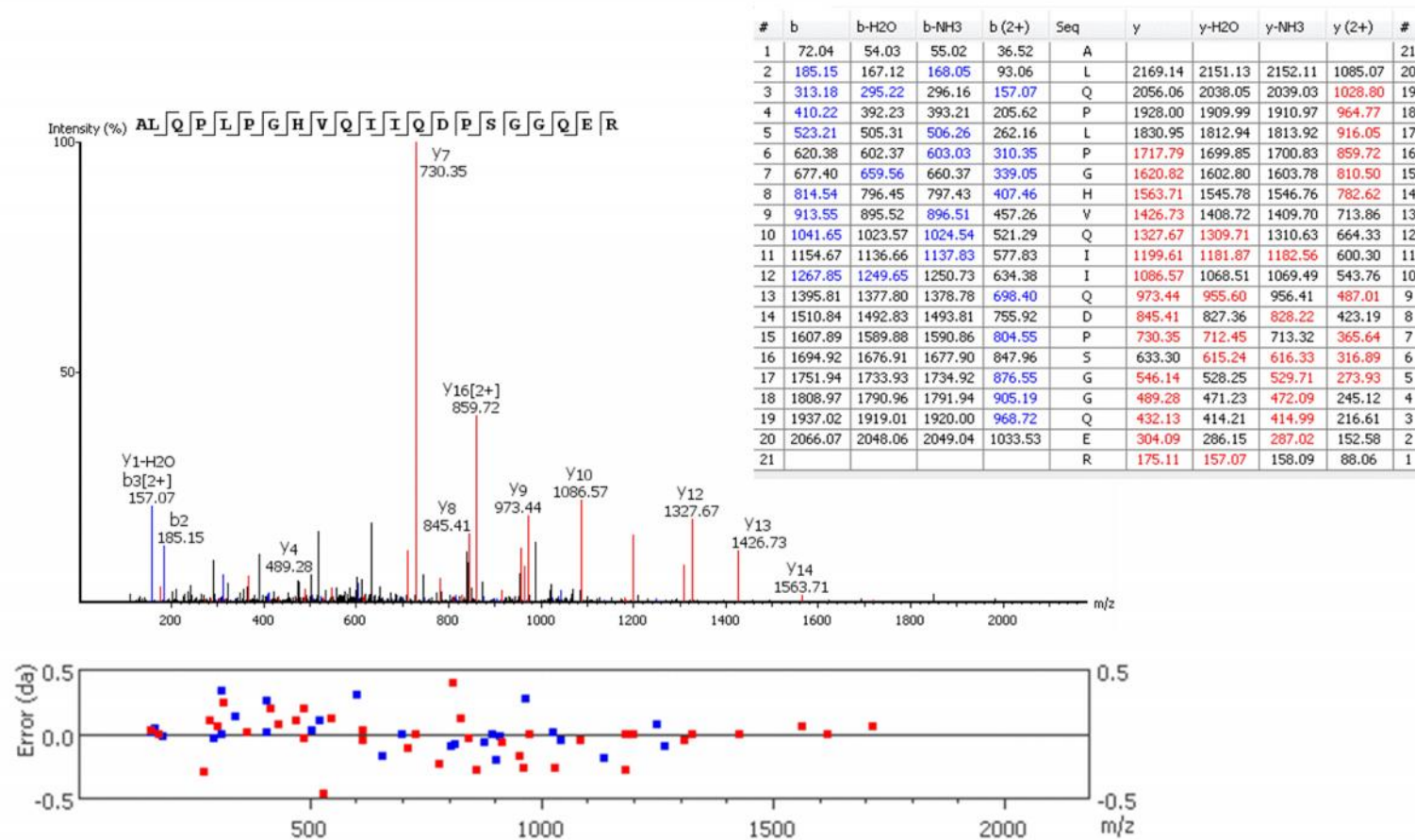
Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S16. ESI-MS/MS spectrum of the doubly-charged molecular ion at m/z 863.3971 (see Table S1). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.

SUPPLEMENTARY INFORMATIONS

Polymorphism at donkey γ -Lactoglobulin II locus: identification and characterization of a new variant with a very low expression



Supplementary Figure S17. ESI-MS/MS spectrum of the triply-charged molecular ion at m/z 747.3972 (see Table S1). The sequence of the peptide is displayed with the fragment ions observed in the spectrum. The inset Table reports the m/z values of the matching ion fragments. The lower panel shows the Δ mass error (Da) with respect to the theoretical expected ion fragments.