

Postprint version

Characterization of the porcine seminal plasma proteome comparing ejaculate portions

Cristina Perez-Patino, Isabel Barranco, Inmaculada Parrilla, M. Luz Valero, Emilio A. Martinez, Heriberto Rodriguez-Martinez and Jordi Roca

Published version: Publication: <http://dx.doi.org/10.1016/j.jprot.2016.04.026>

Characterization of the porcine seminal plasma proteome comparing ejaculate portions

Cristina Perez-Patiño¹, Isabel Barranco¹, Inmaculada Parrilla¹, M. Luz Valero^{2,3}, Emilio A. Martinez¹, Heriberto Rodriguez-Martinez⁴ and Jordi Roca¹

¹Department of Medicine and Animal Surgery, Faculty of Veterinary Science, University of Murcia, Spain; ²Proteomics Section, Central Service for Experimental Research, University of Valencia, Spain; ³Department of Biochemistry and Molecular Biology, University of Valencia, Spain; and ⁴Department of Clinical & Experimental Medicine (IKE), Linköping University, Sweden

Correspondence:

Jordi Roca, Department of Medicine and Animal Surgery, Faculty of Veterinary Science, University of Murcia, E-30100 Murcia, Spain. E-mail: roca@um.es

Abstract

Full identification of SP-proteins remains challenging, particularly in some livestock species such as porcine. This experimental study aims to provide an extensive proteomic analysis of boar SP and to generate a public accessible database of boar SP-proteome. A SP-pool from 33 entire ejaculates from 11 boars (3 ejaculates per boar) was analyzed to characterize the boar SP-proteome. Moreover, 20 ejaculates collected in fractions (P1: first 10 mL of sperm rich ejaculate fraction (SRF), P2: rest of SRF and P3: post-SRF) from 5 boars (4 ejaculates per boar) were analyzed to evaluate differentially expressed SP-proteins among portions. SP-samples were subjected to a combination of SEC, 1-D SDS PAGE and NanoLC-ESI-MS/MS followed by functional bioinformatics analysis. The identified proteins were quantified from normalized LFQ intensity data. A total of 33,557 spectra corresponding to 8,189 peptides and 536 SP-proteins were identified with $\geq 95\%$ Confidence (Unused Score > 1.3) and a false discovery rate (FDR) $\leq 1\%$. Of the 536 SP-proteins, 409 were identified in *Sus Scrofa* taxonomy and 374 of them were

validated with $\geq 99\%$ Confidence. The GO bioinformatics analysis revealed that barely 20 of the identified SP-proteins were specifically implicated in reproductive processes, albeit other SP-proteins could be indirectly involved in functional features and fertilizing ability of boar spermatozoa. The comparative analysis of the SP-proteome revealed 34 proteins (16 identified in *Sus Scrofa* taxonomy) differentially expressed among ejaculate portions, 16 being over-expressed and 18 under-expressed in P1-P2 regarding to P3. In conclusion, this major proteome mapping of the boar SP provides a complex inventory of proteins with potential roles as sperm function- and fertility- biomarkers.

Biological Significance: This proteomic study provides the major characterization of the boar SP-proteome with more than 250 proteins first reported. The boar SP-proteome is described so that a spectral library can be built for relative 'label free' protein quantitation with SWATH approach. This proteomic profiling allows the creation of a publicly accessible database of the boar SP-proteome, as a first step for further understanding the role of SP-proteins in reproductive outcomes as well as for identification of biomarkers for sperm quality and fertility.

Keywords: porcine, ejaculate, seminal plasma, proteome

1. Introduction

Artificial insemination (AI) is used worldwide in the swine industry [1]. Nowadays, commercial semen AI-doses are used as liquid state, but increasing research is done for the practical incorporation of frozen-thawed semen and even sex-sorted spermatozoa into commercial swine AI-programs [2]. Boar ejaculates are selected if most relevant sperm attributes, such as motility, morphology or membrane sperm integrity, are within acceptable limits when evaluated using innovative technologies such as computer assisted analysis and flow-cytometry based procedures [3]. Despite this pre-screening, not all selected ejaculates perform well, neither fertilizing or overcoming technological handling [4]. Consequently, complementary criteria for boar ejaculate selection are currently being explored, specially focusing on seminal plasma (SP) a composite fluid that surrounds and interacts with sperm from ejaculation to deposition into the uterus of the sow.

Seminal plasma is a complex mixture composed of exocrine secretions from the testis, epididymis and the male accessory sexual glands, containing a wide variety of both

organic and inorganic components, being proteins the main one and probably the major conditioners of the reproductive success involving the use of processed semen [5,6]. Certainly, in different mammalian species some SP-proteins have been evidenced to stabilize sperm membrane [7], regulate sperm motility [8] and capacitation [7,9], and further facilitate the colonization of the sperm reservoirs in the female reproductive tract [10]. Moreover, some SP-proteins, once into the female genital tract after natural mating or AI, modulate the uterine immune response against the foreign semen, facilitating sperm transport [11], sperm-oocyte fusion [12] and the development of a healthy embryo [13]. These findings provide clear evidence of the relevance of SP-proteins for sperm functional performance and further highlight the possible use of SP-proteins as potential markers for ejaculate selection. As indicated by Park et al. [14], extensive and comprehensive knowledge of SP-proteome is an essential prerequisite before SP-proteins could be used as biomarkers of sperm functionality, or even fertility. Unfortunately, the boar SP-proteome is still far from being completely decoded, requiring further research [5,15]. Despite the economic relevance, porcine is currently the livestock species with the fewer SP-proteins identified, less than 100 [15,16], which contrasts with other livestock species, such as poultry or ovine with more than 600 SP-proteins identified. This shortage in porcine is surprising, considering boar ejaculates yield a large SP-volume, usually above 150 mL. Consequently, the first aim of the current study will be to improve the knowledge of the boar SP-proteome. To achieve this, SP-samples from boars with proven fertility would be examined using 1-D SDS PAGE and Nano LC-ESI-MS/MS followed by bioinformatics of Gene Ontology (GO) to evaluate cellular location, molecular function and biological process of the identified proteins. In addition, the boar SP-proteome will be described so that a spectral library can be built for relative 'label free' protein quantitation with SWATH approach [17].

The boar ejaculate is fractionally emitted, alike other mammalian species such as equine and even human [5]. Then, it is possible to visually differentiate two main fractions during the ejaculation, a first fraction named sperm rich fraction (SRF), containing most of the total sperm, and a second named post-SRF, largest in volume but containing very few sperm. As boar ejaculates are traditionally hand-collected, the SRF is the only fraction sampled, discarding the post-SP and thereby most of the total SP. However, the ejaculate collection method is currently shifting from manual to semi-automated, for hygienic and labour reasons, moving from collecting just the SRF to the entire ejaculate [18]. The

change increases the relevance of SP, as the large SP-volume of post-SRF fraction will now become a part of the collected ejaculate. There are yet no studies evaluating how this procedural change can affect functional sperm performance. However, in this regard, it is well known that sperm from the SRF freezes better than those from the entire ejaculate, suggesting that the SP from the post-SRF, owing to its rich protein composition, impairs sperm freezability [19]. Furthermore, large proportions of SP in boar semen samples impair sperm sortability and sorting efficiency in ejaculates subjected to flow cytometry for sperm sex-sorting [20]. In addition, Saravia et al. (2009) demonstrated that sperm from the first 10 mL of the SRF sustain cryopreservation better than those from the rest of the ejaculate, which was attributable to differences in SP-proteins between the ejaculate fractions. Accordingly, this study also aims to identify possible qualitative and quantitative differences in SP-proteome among the most relevant boar ejaculate portions: P1: the first 10 mL of SRF, P2: the rest of SRF and P3: the post-SRF.

2. Material and methods

2.1 Boars and ejaculates

All procedures involving boars were performed according to international guidelines (Directive 2010/63/EU), following approval by the Bioethics Committee of Murcia University (research code: 639/2012).

Healthy and sexually mature Large White and Landrace boars with proven fertility and undergoing regular semen collection for commercial AI (AIM Iberica, Spain) were used as ejaculate donors. Boars were subjected to diet and housing conditions characteristics of an AI-center. All the ejaculates used fulfilled the standards of sperm quantity and quality thresholds for the preparation of semen doses for use in AI programs. Thirty-three entire ejaculates collected using the semi-automatic system Collectis® (IMV Technologies, L'Aigle, France) from 11 boars (3 ejaculates per boar), were used to characterize the boar SP-proteome. In order to evaluate differences in SP-protein composition among ejaculate portions, twenty ejaculates were collected fractionated (P1-3) using the gloved hand method, from 5 boars (4 ejaculates per boar).

2.2 Preparation of seminal plasma samples for proteome analysis

Immediately after collection, fully filled 15 mL tubes of semen from each ejaculate/portion were centrifuged twice at 1,500 xg for 10 min (Rotofix 32A; Hettich Zentrifugen, Tuttlingen, Germany). The second SP-supernatant recovered was microscopically verified as sperm free. A protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA) was added to the SP-samples (1%, vol/vol) and then they were stored at -80 °C (Ultra Low Freezer; Haier, Schomberg, Ontario, Canada) until proteomic analysis.

The proteomic analyses were carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform). Seminal plasma samples were thawed at room temperature, centrifuged at 16,100 xg at 4 °C for 1 min. The protein concentration of the SP-sample was measured using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) according to manufacturer instructions. Before starting the proteome analysis, the 33 SP-samples collected for characterization of boar SP-proteome were mixed in a single pool. Similarly, the 4 SP-samples of each ejaculate portion collected from each boar to identify and quantify differentially expressed proteins among ejaculate portions were also mixed in a single pool, thereby generating a total of 5 SP-pools for each ejaculate portion, one per boar.

A SEC was carried out in an ETTAN LC system (GE Healthcare Life Sciences, Little Chalfont, United Kingdom) using a Superdex 200 5/150 GL column (GE Healthcare Life Sciences) controlled by an AKTA ETTAN LC system (GE Healthcare Life Sciences), which furthermore would ease a large-scale proteomic study. Fifty µL of the SP-pool was injected into the column, equilibrated with 200 mM ammonium bicarbonate and 1 mM DTT (Sigma-Aldrich) at a flow rate of 0.18 mL/min at 4 °C. The eluent was collected in 0.2 mL fractions. It was immediately evident that it was not possible to achieve only one fraction with all proteins except the dominant ones. Therefore, 2.5 to 5 µg of protein (depending of the total amount of sample) from the fractions were collected after SEC step (A3-A10), dried in a rotatory evaporator and loaded onto different wells of 12% Tris-HCl precast 1-D SDS PAGE (Bio-Rad, Richmond, CA, USA). The gel was run at a constant voltage of 200 mV for 30 min at room temperature including a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences), and Coomassie Brilliant R250 Blue stain (Bio-Rad) was used to visualize protein bands on the gel. Thereafter, the gel was sliced at 38 kDa, as indicated in Fig. 1, and the top of the gel used to analyze less abundant proteins by in-gel digestion processing. The 1D Gel

fraction between 38 kDa and 20 kDa was discarded for the analysis because in a previous analysis (data not shown) this gel area was highly contaminated by the more abundant SP-proteins, providing little information to the global study of the boar SP-proteome. The more abundant proteins were identified from an aliquot of the mixed SP-sample analyzed by in-solution processing.

2.3 Building a MS/MS library for SWATH analysis of boar SP

2.3.1 Complete proteome. In-solution digestion processing

The more abundant proteins were analyzed using in-solution digestion. Three μL of the pooled SP, containing 10 μg of proteins, were digested with Sequencing Grade Trypsin (Promega Corporation, Madison, USA) to generate peptides of each individual protein according to the following steps: (1) the proteins were reduced using 2 mM DTT reducing reagent in 50 mM NH_4HCO_3 (Sigma-Aldrich) to a final volume of 25 μL and the mixture incubated for 20 min at 60 °C; (2) the proteins were alkylated using 5.5 mM iodoacetamide (IAM) (Sigma-Aldrich) alkylating reagent in 50 mM NH_4HCO_3 to a final volume of 30 μL and incubated for 30 min in the dark; (3) 10 mM DTT in 50 mM NH_4HCO_3 were added to a final volume of 60 μL and sample was vortex and incubated for 30 min at room temperature; (4) trypsin was added in a 1:20 ratio (Trypsin:Protein, w:w) to a final volume of 70 μL , sample was carefully mixed and digestion carried out over night at 37 °C. The digestion was stopped on the following morning by acidifying the solution ($\text{pH}<6$) with 7 μL 10% TFA (Fisher Scientific, Madrid, Spain). The final concentration of protein in the digested sample was 0.13 $\mu\text{g}/\mu\text{L}$.

2.3.2 Less abundant proteins. In-gel digestion processing

The portion of the 1-D SDS PAGE containing proteins with a molecular weight higher than 38 kDa was processed as a unique sample. Following washing in MilliQ water (Merck Millipore, Darmstadt, Germany), the gel was dehydrated in ACN (Fisher Scientific), reduced with DTT and alkylated with IAM. Each slide was cut and small pieces of approximately 1 mm^2 in size were transferred into 1.5 mL Eppendorf tubes. Sequencing Grade Trypsin digestion of the sliced gel was performed following the protocol used by Shevchenko et al. [22]. Briefly, for protein digestion 200 μL of trypsin digestion buffer were added to each dried gel piece and digestion was set at 37 °C overnight. The trypsin digestion was stopped with 10% TFA and the supernatant,

containing the non-extracted digests, was carefully removed, leaving behind the sliced gels in the Eppendorf tube. Then, for peptide extraction, 200 μ L of pure ACN, were added to each tube and incubated for 15 min at 37 °C in a shaker. The new supernatant containing the peptide mixture was carefully withdrawn. Both supernatants were mixed in a tube and dried in a speed vacuum (ISS 110 SpeedVac System, Thermo Savant, ThermoScientific, Langensfeld, Germany) for 20 min and re-suspended in 7 μ L of 2% ACN and 0.1% TFA prior to LC and MS analysis.

2.3.3 LC-MS/MS analysis

The peptides recovered from in-gel and in-solution digestion processing were examined by LC using a NanoLC Ultra 1-D plus Eksigent (Eksigent Technologies, Dublin, CA, USA) which was directly connected to an AB SCIEX TripleTOF 5600 mass spectrometer (AB SCIEX, Framingham, MA, USA) in direct injection mode. Briefly, 5 μ L from each digested sample were trapped on a NanoLC pre-column (3 μ m particles size C18-CL, 350 μ m diameter x 0.5 mm long; Eksigent Technologies) and desalted with 0.1% TFA at 3 μ L/min during 5 min. Then, the digested peptides present in the samples were separated using an analytical LC column (3 μ m particles size C18-CL, 75 μ m diameter x 12 cm long, Nikkyo Technos Co®, Tokyo, Japan) equilibrated in 5% ACN 0.1% formic acid (FA) (Fisher Scientific). Peptide elution was performed by applying a mixture of solvents A and B; solvent A being 0.1% FA in water and solvent B being 0.1% FA in ACN. The peptides were eluted from the column with a linear gradient from 5% to 35% of solvent B at a constant flow rate of 300 nL/min over 90 min.

The eluted peptides were thereafter direction-ionized using an ESI Nanospray III (AB SCIEX) and then analyzed on an AB SCIEX TripleTOF 5600 mass spectrometer coupled to the NanoLC system. The samples were ionized applying 2.8 kV to the spray emitter and the TripleTOF was operated in data-dependent mode, in which a TOF MS scan was made from 350 to 1259 m/z, accumulated for 250 ms TOF followed by 150 ms TOF with the same scan range for MS, and the 25 most abundant multiply charged (2+, 3+, 4+ or 5+) precursor peptide ions automatically selected. Ions with 1+ and unassigned charge states were rejected from the MS/MS analysis. Collision energy was automatically set by the instrument according to the equation $|CE|=(\text{slope})\times(m/z)+(\text{intercept})$ with Charge=2; Slope= 0.0575 and Intercept=9.

The proteomics data and result-files from the analysis have been deposited to the ProteomeXchange Consortium [23] via the PRIDE partner repository, with the dataset identifier PXD003579 and DOI: 10.6019/PXD003579.

2.3.4. LC-SWATH-MS acquisition

For SWATH-MS-based experiments the instrument (SCIEX TripleTOF 5600) was configured as described by Gillet et al. [17]. Briefly, the mass spectrometer was operated in a looped product ion mode. In this mode, the instrument was specifically tuned to allow a quadrupole resolution of Da/mass selection. The stability of the mass selection was maintained by the operation of the Radio Frequency (RF) and Direct Current (DC) voltages on the isolation quadrupole in an independent manner. Using an isolation width of 16 Da (15 Da of optimal ion transmission efficiency and 1 Da for the window overlap), a set of 37 overlapping windows was constructed covering the mass range 450–1000 Da. Consecutive swaths need to be acquired with some precursor isolation window overlap to ensure the transfer of the complete isotopic pattern of any given precursor ion in at least one isolation window and, thereby, to maintain optimal correlation between parent and fragment isotopes peaks at any LC time point. In this way, 5 μ L of each sample were loaded onto a trap column (NanoLC Column, 3 μ m C18-CL, 75 μ m x 15 cm; Eksigent) and desalted with 0.1% TFA at 3 μ L/min during 5 min. The peptides were loaded onto an analytical column (LC Column, 3 μ m C18-CL, 75 μ m x 12cm, Nikkyo) equilibrated in 5% acetonitrile 0.1% FA. Peptide elution was carried out with a linear gradient of 5 to 40% B in 90 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nL/min. Eluted peptides were infused in the spectrometer nanoESI qTOF (SCIEX TripleTOF 5600). The TripleTOF was operated in swath mode, in which a 0.050 s TOF MS scan from 350 to 1250 m/z was performed, followed by 0.080 s product ion scans from 230 to 1800 m/z on the 37 defined windows (3.05 sec/cycle). Collision energy was set to optimum energy for a 2+ ion at the center of each SWATH block with a 15 eV collision energy spread. The mass spectrometer was always operated in high sensitivity mode.

2.5 Protein identification, validation and quantification

After LC-MS/MS, The SCIEX.wiff data-files were processed using ProteinPilot v5.0 search engine (AB SCIEX). The Paragon algorithm (4.0.0.0, 4767) of ProteinPilot was used to search against the National Center for Biotechnology Information non-redundant (NCBI nr; 70353186 proteins searched) protein sequence database with the following

parameters: trypsin specificity, cys-alkylation (IAM), no taxonomy restricted, and the search effort set to through. The same analysis was also made for *Sus scrofa* taxonomy. To avoid using the same spectral evidence in more than one protein, the identified proteins were grouped based on MS/MS spectra by the Protein-Pilot Pro Group™ Algorithm, regardless of the peptide sequence assigned. The protein within each group that could explain more spectral data with confidence was depicted as the primary protein of the group. The resulting Protein-Pilot group file was loaded into PeakView® (v2.1, AB SCIEX) and peaks from SWATH runs were extracted with a peptide confidence threshold of 99% confidence (Unused Score ≥ 1.3) and a false discovery rate (FDR) less than 1%. It was not set up a minimum number of peptides for the identification. However, the proteins identified from a single peptide matched were validated manually (see Supplementary data 1). For this, the MS/MS spectra of the assigned peptides were extracted by Protein-Pilot, and were only validated the proteins that fulfilled the following criteria: (1) peptide mass tolerance less than 10 ppm, (2) 99% of confidence level in peptide identification, and (3) complete b/y ions series found in the MS/MS spectrum. The identified proteins were quantified using PeakView® from normalized label-free quantification (LFQ) intensity data.

2.6 Gene ontology and bioinformatic analysis

Bioinformatic analysis of identified and validated SP-proteins was manually performed using the comprehensive bioinformatics tool for functional annotation UniProt KB database (www.uniprot.org) in combination with PANTHER (www.pantherdb.org). Both databases downloaded 22/03/2016, containing 34,272 and 21,483 entries in UniProt KB and PANTHER, respectively. This analysis allowed elucidation of the different functions and processes in which the identified and validated proteins would be putatively involved. Three independent sets of ontology were used in the annotation: “the molecular function”, “the biological processes”, in which the proteins participate, and their “cellular component”. Proteins without similarity to database entries were not considered for collation.

2.7 Statistical analysis

The quantitative data obtained by PeakView® were analyzed using MarkerView® (v1.2, AB SCIEX). First, areas were normalized by total areas summa. Principal Component Analysis (PCA) was performed to evaluate the discriminative ability of proteins in

ejaculate portions. Multiexperiment Viewer (MeV) (<http://www.tm4.org/mev.html>) was used to identify the protein differentially expressed among ejaculate portions using Student's t-test subjected to Bonferroni correction. Proteins were considered significantly differentially expressed among ejaculate portions with an adjusted P-value < 0.05. The mean quantity of proteins in each ejaculate portion and the fold-change (ratio between the mean of ejaculate portions) were calculated. The results of Hierarchical clustering analysis of proteome profile of different samples were shown by mean of heat map after z-score normalization, using Euclidean distances.

3. Results

3.1 Characterization of boar seminal plasma proteome

The complete spectral library included 33,557 spectral corresponding to 8,189 distinct peptides and 536 SP-proteins with a FDR \leq 1% is showed in Supplementary Table 1. A total of 31,237 spectra corresponding to 7,840 distinct peptides and 409 SP-proteins were identified in *Sus Scrofa* taxonomy from the nrNCBI databank with a FDR \leq 1%, and 374 of them were validated with \geq 99% Confidence (Unused Score \geq 1.3). A total of 359 proteins were identified from in-gel digestion and 70 from in-solution digestion processing. Of the latter, just 17 proteins were not identified in the in-gel approach. The complete list of the 374 well-identified boar SP-proteins, their gi Accession Number, Protein Name, Gene Name, Species, Matched Peptides, Unused Score, % of Sequence Coverage (% Cov), protein previously reported indication, Detection Method of each protein, and the band sec to which belong (if applicable) are provided in Supplementary Table 2. The complete list with the Peptide sequence, Confidence and Contribution from all matched peptides for the 374 proteins identified is showed in Supplementary Table 3.

3.2 Bioinformatics Analysis

Of the 374 proteins identified and validated in *Sus Scrofa* taxonomy, a total of 195 proteins were mapped to 201 IDs in UniProt KB and 115 proteins to 119 IDs in PANTHER for protein enrichment. The results are showed in Fig. 2. For molecular function (Fig. 2a), a total of 255 hits were found. Most proteins (117, 46%) were engaged in catalytic activity, while 81 (32%) and 32 (12%) were linked to binding- and regulatory activity, respectively. Only two proteins were classified as antioxidants. Of the proteins

identified with catalytic activity, 78 belonged to the hydrolases. Most of those with binding activity showed ion binding activity (49). Regarding to biological process (Fig. 2b), 126 out of 601 identified hits (21%) were related to metabolic processes and 101 (17%) related to cellular processes. Noticeably, 29 proteins (5%) would be implicated in immune system processes and solely 20 (3%) were specifically implicated in reproductive processes. Finally, a total of 400 hits belonging to cellular components (Fig. 2c). Most of them were related to intracellular processes, predominantly from cell organelles (95, 24%) and cell part (99, 25%). Others proteins belong to extracellular region (104, 26%) and extracellular matrix (13, 3%). Sixty-five proteins (16%) originated from the plasma membrane.

3.3 Comparison of SP-proteome profiles among ejaculate portions

A total of 447 proteins were identified and all of them were identified and quantified in the three-ejaculate portions. Of them, 223 belong to *Sus Scrofa* taxonomy. The full list of these proteins, including relative amount of each protein in each one of the three-ejaculate portions, is showed in Supplementary Table 4. PCA was used to determine whether expression of SP-proteins would differ among ejaculate portions. The biplot, displaying the first and the second components (accounting for 53.5% and 12.4% of variance), did not shown a clear separation based in protein expression between P1 and P2 in the five boars, but P3 samples were clearly separated from the P1-P2 in component 1 of the PCA (Figure 3a). The most discriminating protein spots between P1-P2 and P3 were identified in Volcano plots. A total of 34 proteins were differentially ($p < 0.05$) expressed between P1-P2 and P3 samples (Figure 3b). Among them, 16 were identified in *Sus Scrofa* taxonomy. Of the 34 differentially expressed proteins, 16 proteins were over-expressed and 18 proteins under-expressed in P1-P2 regarding to P3 (Tables 1 and 2). The hierarchical clustering of SP-samples, based on the expression of these 34 differentially expressed proteins, separated the fifteen SP-samples (5 of each ejaculate portion) into two different main clusters, corresponding to P1-P2 samples and to P3 samples, respectively.

4. Discussion

To the best of our knowledge this study generates the largest library of SP-proteins reported to date in boars as more than 250 proteins were first time identified in boar SP.

One of the most important contributions of this study would be the identification of a large number of proteins present in small quantities in the boar SP. Probably the approach followed with SEC, 1-D SDS PAGE and post digestion processing in both gel and solution, was decisive for the emergence of these less abundant proteins, which are usually hidden by the most abundant proteins. Three hundred fifty-nine of the 374 identified proteins were spotted in gel analysis due to the chromatographic fractionation and the in gel depletion of the more abundant SP-proteins. In solution digestion of the SP allowed the identification of other 70 proteins, 17 of them were not identified in the gel analysis. These results show that both approaches are complementary for a more deep description of boar SP-proteome. The correct identification of these less abundant proteins could be essential to understand the reproductive response of ejaculates spermatozoa.

A full decrypting of the boar SP-proteome is challenging since most total protein mass is taken up by a few number of low molecular weight proteins, the so-called most abundant proteins. Previous proteomic studies in boar SP using gel electrophoresis allowed identifying simultaneously a limited number of proteins [15,16]. It is well known that the gel-based techniques presents a limited resolution and most abundant proteins typically mask the identification of less abundant ones [24]. To overcome these limitations and to identify the largest possible number of SP-proteins, a pre-fractionation step by SEC is needed to fractionate proteins according to molecular weight and thereby improving accuracy of later LC-MS/MS analysis, as SEC is a chromatographic method allowing to separate molecules in solution by their size, and in some cases by their molecular weight. The chromatographic fractionation of our SP-samples did not allow achieving SEC fractions without the presence of low molecular weight proteins (most abundant proteins). Thereby, a subsequent analysis by 1-D SDS PAGE proved necessary to achieve an adequate SP fractionation for spectral library construction.

Regarding the bioinformatics analysis, it is noticeable that only 20 of the 374 proteins identified in boar SP are -until date- recorded in GO as directly associated with reproductive processes. This is absolutely surprising for a heterogeneous fluid as SP, a secretion composite closely related to spermatozoa from their origin to their transport along the female genitalia. It is also surprising that many of the hereby-identified proteins in boar SP have no yet an attributed specific role in the GO. Despite of this, we are aware that many identified SP-proteins, included or not in GO, are indirectly involved in

reproductive processes, such as those regulating immune responses by the female genital tract. Certainly, some SP-proteins such as Complement factor H, Plasma protease C1 inhibitor, Cathepsin H, Dipeptidyl peptidase 4 and Transforming growth factor- β_2 regulate immune responses in the female genital tract [11,25], facilitating sperm transport and fertilization. Many of the identified SP-proteins show catalytic activity, specifically hydrolase activity, and some of them would also indirectly be related with reproductive functions, as occurs with Acrosin. It is known that Acrosin is involved in the proacrosin-acrosin system necessary for boar sperm capacitation [26]. Cathepsins (Pro-cathepsin H, Cathepsin L1, Cathepsin H, Cathepsin Z and Cathepsin D) are also SP-proteins with hydrolase activity, specifically protease activity, some secreted in the epididymis [27,28]. This origin suggests that they may rule remodeling sperm membranes during epididymal sperm transit. Moreover, SP-Cathepsins could also interact with sperm membrane proteins affecting sperm fertilizing ability, as previously suggested in bull by Moura et al. [29]. A large number of the boar SP-proteins identified in the present study showed binding activity, which would also show important reproductive functions because are able to bind to the sperm plasma membranes modulating sperm functionality [30]. Among them, the most abundant in boar SP are spermadhesins, including AWN-1, AWN-3, AQN, PSP-I and PSP-II, with proved implication in the fertilizing ability of boar spermatozoa [31]. Heat shock 70 kDa protein and Lysosomal type alpha-D-mannosidase are other two SP-proteins encoded in GO as binding activity that have proven impact on sperm functionality [32,33]. Boar SP includes a large number of glycoproteins, which also includes the aforementioned spermadhesins. Their facility for glycosylation gives them an important role in sperm functionality [34], particularly in regulating sperm motility, capacitation, acrosome reaction and sperm-oocyte binding [35]. Some identified SP-proteins show ion-binding properties, being particularly relevant for sperm performance those showing zinc- and calcium-binding properties. Zinc-binding proteins play an important role to preserve sperm chromatin stability and to regulate sperm motility [36]. Moreover, they have antibacterial and antioxidant functions and participate in the regulatory immune response in the female reproductive tract [37,38]. Some calcium-binding proteins have also been identified in boar SP. Calcium, especially extracellular calcium, is necessary for sperm exocytosis processes such as the acrosome reaction [39]. Therefore, the presence of calcium-binding proteins in boar SP could limit the availability of free calcium by sperm, preventing premature sperm capacitation. Just a few number of boar SP-proteins are encoded in GO as antioxidants. However, identified

SP-proteins have proven antioxidant activity. For instance, Albumin, a common SP-protein encoded as binding activity, which has proved capacity in protecting sperm against oxidative stress caused by lipid peroxidation [16]. Likewise, some identified SP-proteins encoded in GO as metabolic function show oxidoreductase activity, for instance Epididymal secretory glutathione peroxidase and Superoxide dismutase. Boar spermatozoa are particularly vulnerable to oxidative stress because their plasma membrane is particularly rich in polyunsaturated fatty acids and the above-mentioned SP-enzymes have proved effective protecting boar sperm against oxidative stress [40]. Regarding to cellular location, it is also noticeable large proportion of SP-proteins had an intracellular localization. One possible explanation would be they come from SP-exosomes, whose content is rich in proteins [41]. Fifty-three proteins identified in boar SP have been encoded in GO as originated from exosomes. Prevailing data demonstrate that proteins in the epididymal fluid associated with post-testicular sperm maturation are transferred to the sperm by exosomes and consequently the protein content of these extracellular vesicles have been shown to be involved in post-testicular sperm maturation, including sperm motility acquisition and reduction of oxidative stress and acrosome reaction [41-42].

In addition to the different proteomic analytical approach used, another reason for discrepancy among studies about the number of SP-protein identified in boar ejaculates could be related to the source of SP. The boar SP samples evaluated by Druart et al. [15] and Gonzalez-Cadavid et al. [16] seem to come from SRF as they were collected using glove-hand method, while SP-samples evaluated in the present study come from the entire ejaculate, including SRF and post-SRF. In this context, it is well known that the total protein composition differs among boar ejaculate fractions [43]. As in human, the boar ejaculate is expelled in fractions, the SRF and the post-SRF being the two main fractions, with relevant differences in SP-composition [43] including cytokine-contents [44] and antioxidant enzymes [45]. Further research is required to verify possible divergences in the SP-proteome among ejaculate fractions/portions, particularly in the light of the ongoing discussions as to whether to use only the SRF or the entire ejaculate for both AI and sperm technologies, such as cryopreservation and sex-sorting [2]. Noticeable all expressed SP-proteins were present in each of the three ejaculate portions, which would indicate that the possible influence of SP-proteins on boar sperm performance, both AI-fertility and aid overcome sperm technologies, would be more linked to quantitative rather

than qualitative differences. In this regard, it was also noticeable that there were no statistically quantitative differences between P1 and P2 in any of the 447 expressed SP-proteins and that only 34 of them showed statistically quantitative differences between P1-P2 and P3. The protein quantitative data were subjected to Bonferroni correction before statistical analysis. The Bonferroni sequential correction is considered conservative as it has the effect of "cleaning" data from false positives resulting from the random nature of sampling process. Cleaning data before statistical analysis deserves special consideration when analyzing biological compounds undergoing to large natural quantitative variations among samples, as in the case of SP-proteins. Consequently, the quantitative differences on SP-proteins among ejaculate portions showed in the present study should be considered statistically robust, but the conservative statistic procedure developed might overlook some other expressed proteins that could also show quantitative differences between ejaculate portions, although to a lesser degree and always open to misinterpretation. The absence of quantitative differences in SP-protein expression between P1 and P2 may be understandable, as P1 and P2 together comprise the first naturally occurring ejaculate fraction (SRF), whose SP has a common origin, albeit having amount differences [5].

Looking at the 34 differentially expressed SP-proteins between P1-P2 and P3, 16 of them were identified in *Sus Scrofa* taxonomy and eight of them were more expressed in P1-P2 than P3. Six of these eight proteins were previously identified in SP of livestock species and all of them have a testicular (Pancreatic secretory granule membrane major glycoprotein [GP2], Galactosidase beta 1-like 3 [GLB1L3] and Golgi apparatus protein 1 [GLG1]) or epididymal (Hexosaminidase B [HEXB], Epididymal-specific lipocalin-5 [LCN5] and Arylsulfatase A [ARSA]) origin. Many of these proteins have been related to some specific sperm functionalities, as membrane stability and permeability (GLB1L3, [46]; ARSA, [47]), capacitation (HEXB, [48]), acrosome reaction (GP2, [49]), fertilizing ability (LCN-5, [50]) or sperm-zone pellucid binding (ARSA, [51]).

The other eight SP-proteins identified in *Sus Scrofa* taxonomy were less expressed in P1-P2 than in P3. Four of them (Fibronectin [FN1], Sulfhydryl oxidase 1, Angiotensin-converting enzyme and Epididymal secretory protein E1) have also a testicular or epididymal origin and all of them have been related to sperm maturation. Other of these SP-proteins more expressed in P3, as Alpha-enolase or Alkaline phosphatase (ALP), have been related to sperm motility [52-53]. Noticeable, deoxyribonuclease -2-alpha, an acid

endonuclease secreted by male accessory glands, is involved in the degradation of exogenous DNA [54] and it also provide a bactericide activity protecting sperm in the transit along female genital tract [55]. Nucleobinding-1 belongs to a family of proteins with calcium and DNA binding properties [56], but its reproductive function is still unknown. In addition, some of the above related proteins, as HEXB [48], FN1 [57] or AP [58], have been linked to the ability of sperm to overcome cryopreservation.

Eighteen of the 34 proteins differently expressed between P1-P2 and P3 were not identified in *Sus Scrofa* taxonomy but some of them have been related to spermatogenesis, sperm maturation or mature sperm functionality in other mammalian species. For instance, Myelin protein zero-like protein 1 [59], Guanine nucleotide-binding protein subunit alpha-11 [60], Plastin-3 isoform 1 [61] would be related to spermatogenesis; and Polypeptide N-acetylgalactosaminyltransferase 2 and Beta-galactosidase-1-like protein 2-like with sperm maturation along epididymal transit [62] (Hall and Killian, 1987). Other five of these 18 SP-proteins (EGF-like repeat and discoidin I-like domain-containing protein 3, Putative phospholipase B-like 2 isoform 1, Prominin-2, Ezrin and Syntaxin-binding protein 2) have been related with sperm capacitation [63-67]. Heat shock cognate 71 kDa is a protein related to environmental or physiological stresses [68] synthesized by sperm cells [69] that has been recently related to boar sperm cryotolerance [70]. Finally, Ectonucleotide pyrophosphatase/phosphodiesterase family member is an ATP-hydrolyzing enzyme described in epididymis suggesting that this protein could be involved in the modulation of the epididymal environment [71].

Based on the foregoing, it is clear that most of the SP-proteins differentially expressed among the ejaculate portions show a clear connection with the development and/or functionality of spermatozoa. However, it does not seem clear that increased expression of a particular functional SP-protein was linked to a best performance by sperm. For instance, some SP-proteins, as FN1, have been positively related to sperm cryoresistance, as sperm from boars with high FN1 in SP showed best freezability [57]. This would indicate that sperm belonging to the ejaculate portion with more FN1 expression (more in P3 than P1+P2) should cryopreserve better. This does not seem be the case, as sperm from the entire ejaculate (P1+P2+P3) cryopreserve worse than those from SRF (P1+P2) [19]. This example could also be applied to other differentially expressed SP-proteins as AP, which has been related to sperm motility. That a SP-protein carries out its maximal function when expressed within a range, declining it when under- or overexpressed, could

explain these apparent inconsistencies. All together, the presented data clearly demonstrate the current limited knowledge of the number and role of boar SP-proteins, which, in turn, provides exciting research opportunities. Clarifying the role of SP-proteins on reproductive processes, including ability of sperm to successfully overcome sperm technologies, will possibly allow the identification of SP-protein biomarkers but also to promote the design of semen extenders able to maintain sperm fertilizing capacity for longer.

5. Conclusion

In summary, this experimental study provided an extensive proteomic analysis of boar SP, which still remained poorly understood, and generated a public accessible database of boar SP-proteome. The boar SP-proteome is described so that a spectral library can be built for relative 'label free' protein quantitation with SWATH approach. The analytical approach used, including SEC, 1-D SDS PAGE and post digestion processing in both gel and solution, proved crucial for the identification of the large number of proteins present in small quantities, in the boar SP. Additionally, the comparative SP-proteome study of the boar ejaculate portions resulted in the identification of a number of differentially expressed proteins, many of which are linked to sperm reproductive performance. Overall, the results of the present study intend to provide an elementary core for further studies focused on understanding the role of SP-proteins in reproductive outcomes of boar semen as well as for identification of biomarkers for sperm quality and fertility.

Acknowledgements

This experimental study was supported by MINECO (AGL2012-39903) Madrid (Spain), FEDER funds (EU), Seneca Foundation (19892/GERM/15) Murcia (Spain), The Swedish Research Council (VR, 521-2011-6353), the Swedish Research Council Formas (221-2011-512) and the Research Council in Southeast Sweden (FORSS, 378091/312971), Sweden. C. Perez-Patiño and I. Barranco and were financially supported by the Seneca Foundation (Murcia, Spain) and MECD (Madrid, Spain), respectively. The authors are grateful to AIM Iberica (Topigs Norsvin Iberica) for supplying the boar ejaculates.

References

- [1] Riesenbeck A. Review on international trade with boar semen. *Reprod Dom Anim* 2011;46(S2):1-3.
- [2] Roca J, Parrilla I, Bolarin A, Martinez EA, Rodriguez-Martinez H. Will artificial insemination in pigs become more efficient? *Theriogenology* 2015; S0093-691X(15)00651-2.
- [3] Broekhuijse ML, Šoštarić E, Feitsma H, Gadella BM. Relationship of flow cytometric sperm integrity assessments with boar fertility performance under optimized field conditions. *J Anim Sci.* 2012;90(12):4327-36.
- [4] Parrilla I, del Olmo D, Sijes L, Martinez-Alborcia MJ, Cuello C, Vazquez JM, Martinez EA, Roca J. Differences in the ability of spermatozoa from individual boar ejaculates to withstand different semen-processing techniques. *Anim Reprod Sci.* 2012;132(1-2):66-73
- [5] Rodríguez-Martínez H, Kvist U, Ernerudh J, Sanz L, Calvete JJ. Seminal plasma proteins: what role do they play? *Am J Reprod Immunol* 2011;66(1):11-22.
- [6] Jonáková V, Manásková P, Tichá M. Separation, characterization and identification of boar seminal plasma proteins. *J Chromatogr B Analyt Technol Biomed Life Sci* 2007;849(1-2):307-14.
- [7] Caballero I, Vazquez JM, Mayor GM, Almiñana C, Calvete JJ, Sanz L, Roca J, Martinez EA. PSP-I/PSP-II spermadhesin exert a decapacitation effect on highly extended boar spermatozoa. *Int J Androl* 2009;32(5):505-13.
- [8] Luna C, Yeste M, Rivera Del Alamo MM, Domingo J, Casao A, Rodriguez-Gil JE, Pérez-Pé R, Cebrián-Pérez JA, Muiño-Blanco T. Effect of seminal plasma proteins on the motile sperm subpopulations in ram ejaculates. *Reprod Fertil Dev.* 2015;doi: 10.1071/RD15231.
- [9] Vadnais ML, Roberts KP. Seminal plasma proteins inhibit in vitro- and cooling-induced capacitation in boar spermatozoa. *Reprod Fertil Dev* 2010;22(6):893-900.
- [10] Wagner A, Ekhlesi-Hundrieser M, Hettel C, Petrunkina A, Waberski D, Nimtz M, Töpfer-Petersen E. Carbohydrate-based interactions of oviductal sperm reservoir formation-studies in the pig. *Mol Reprod Dev* 2002;61(2):249-57.
- [11] Rodriguez-Martinez H, Saravia F, Wallgren M, Martinez EA, Sanz L, Roca J, Vazquez JM, Calvete JJ. Spermadhesin PSP-I/PSP-II heterodimer induces migration of polymorphonuclear neutrophils into the uterine cavity of the sow. *J Reprod Immunol* 2010;84(1):57-65.
- [12] Töpfer-Petersen E, Ekhlesi-Hundrieser M, Kirchhoff C, Leeb T, Sieme H. The role of stallion seminal proteins in fertilisation. *Anim Reprod Sci* 2005;89(1-4):159-70.

- [13] O'Leary S, Jasper MJ, Warnes GM, Armstrong DT, Robertson SA. Seminal plasma regulates endometrial cytokine expression, leukocyte recruitment and embryo development in the pig. *Reproduction* 2004;128(2):237-47.
- [14] Park Y-J, Kim J, You Y-A, Pang M-G. Proteomic revolution to improve tools for evaluating male fertility in animals. *J Proteome Res* 2013;12:4738-4747.
- [15] Druart X, Rickard JP, Mactier S, Kohnke PL, Kershaw-Young CM, Bathgate R, Gibb Z, Crossett B, Tsikis G, Labas V, Harichaux G, Grupen CG, de Graaf SP. Proteomic characterization and cross species comparison of mammalian seminal plasma. *J Proteomics* 2013;91:13-22.
- [16] Gonzalez-Cadavid V, Martins JA, Moreno FB, Andrade TS, Santos AC, Monteiro-Moreira AC, Moreira RA, Moura AA. Seminal plasma proteins of adult boars and correlations with sperm parameters. *Theriogenology* 2014;82(5):697-707.
- [17] Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics*. 2012;11(6):O111.016717.
- [18] Roca J, Broekhuijse ML, Parrilla I, Rodriguez-Martinez H, Martinez EA, Bolarin A. Boar Differences In Artificial Insemination Outcomes: Can They Be Minimized? *Reprod Domest Anim* 2015; 50(S2):48-55.
- [19] Alkmin DV, Perez-Patiño C, Barranco I, Parrilla I, Vazquez JM, Martinez EA, Rodriguez-Martinez H, Roca J. Boar sperm cryosurvival is better after exposure to seminal plasma from selected fractions than to those from entire ejaculate. *Cryobiology* 2014;69(2):203-10.
- [20] Alkmin DV, Parrilla I, Tarantini T, Del Olmo D, Vazquez JM, Martinez EA, Roca J. Seminal plasma affects sperm sex sorting in boars. *Reprod Fertil Dev* 2014;doi: 10.1071/RD14088.
- [21] Saravia F, Wallgren M, Johannisson A, Calvete JJ, Sanz L, Peña FJ, Roca J, Rodríguez-Martínez H. Exposure to the seminal plasma of different portions of the boar ejaculate modulates the survival of spermatozoa cryopreserved in MiniFlatPacks. *Theriogenology*. 2009h;71(4):662-75.
- [22] Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci* 1996;93(25):14440-5.
- [23] Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnol* 2014; 30: 223-6.
- [24] Rabilloud T, Lelong C. Two-dimensional gel electrophoresis in proteomics: a tutorial. *J Proteomics* 2011;74(10):1829-41.
- [25] Sakaue T, Takeuchi K, Maeda T, Yamamoto Y, Nishi K, Ohkubo I. Factor H in porcine seminal plasma protects sperm against complement attack in genital tracts. *J Biol Chem* 2010;285(3):2184-92.
- [26] Puigmulé M, Fàbrega A, Yeste M, Bonet S, Pinart E. Study of the proacrosin-acrosin system in epididymal, ejaculated and in vitro capacitated boar spermatozoa. *Reprod Fertil Dev*. 2011;23(7):837-45.

- [27] Okamura N, Tamba M, Uchiyama Y, Sugita Y, Dacheux F, Syntin P, Dacheux JL. Direct evidence for the elevated synthesis and secretion of procathepsin L in the distal caput epididymis of boar. *Biochim Biophys Acta*. 1995;1245(2):221-6.
- [28] Moura AA, Chapman DA, Koc H, Killian GJ. Proteins of the cauda epididymal fluid associated with fertility of mature dairy bulls. *J Androl*. 2006;27(4):534-41.
- [29] Moura AA, Souza CE, Stanley BA, Chapman DA, Killian GJ. Proteomics of cauda epididymal fluid from mature Holstein bulls. *J Proteomics*. 2010;73(10):2006-20.
- [30] Jelínková P, Manásková P, Tichá M, Jonáková V. Proteinase inhibitors in aggregated forms of boar seminal plasma proteins. *Int J Biol Macromol*. 2003;32(3-5):99-107.
- [31] Caballero I, Vazquez JM, García EM, Parrilla I, Roca J, Calvete JJ, Sanz L, Martínez EA. Major proteins of boar seminal plasma as a tool for biotechnological preservation of spermatozoa. *Theriogenology*. 2008;70(8):1352-5.
- [32] Jin YZ, Dacheux F, Dacheux JL, Bannai S, Sugita Y, Okamura N. Purification and properties of major alpha-D-mannosidase in the luminal fluid of porcine epididymis. *Biochim Biophys Acta*. 1999;1432(2):382-92.
- [33] Spinaci M, Volpe S, Bernardini C, De Ambrogi M, Tamanini C, Seren E, Galeati G. Immunolocalization of heat shock protein 70 (Hsp 70) in boar spermatozoa and its role during fertilization. *Mol Reprod Dev*. 2005;72(4):534-41.
- [34] Calvete JJ, Solís D, Sanz L, Díaz-Mauriño T, Töpfer-Petersen E. Glycosylated boar spermadhesin AWN-1 isoforms. Biological origin, structural characterization by lectin mapping, localization of O-glycosylation sites, and effect of glycosylation on ligand binding. *Biol Chem Hoppe Seyler*. 1994;375(10):667-73.
- [35] Pang PC, Chiu PC, Lee CL, Chang LY, Panico M, Morris HR, Haslam SM, Khoo KH, Clark GF, Yeung WS, Dell A. Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. *Science*. 2011;333(6050):1761-4.
- [36] Strzezek J and Hopfer E. Zinc ion-dependent protein in boar Semen. I. Egg-yolk precipitating activity and some biochemical properties. *Anim. Reprod. Sci*. 1987;13: 117-131.
- [37] Vivacqua A, Siciliano L, Sabato M, Palma A, Carpino A. Prostatomes as zinc ligands in human seminal plasma. *Int J Androl*. 2004;27(1):27-31.
- [38] Edström AM, Malm J, Frohm B, Martellini JA, Giwercman A, Mörgelin M, Cole AM, Sørensen OE. The major bactericidal activity of human seminal plasma is zinc-dependent and derived from fragmentation of the semenogelins. *J Immunol*. 2008;181(5):3413-21.
- [39] Gadella BM, Harrison RA. The capacitating agent bicarbonate induces protein kinase A-dependent changes in phospholipid transbilayer behavior in the sperm plasma membrane. *Development*. 2000;127(11):2407-20.
- [40] Bathgate R. Antioxidant mechanism and their benefit on post-thaw boar sperm quality. *Reprod Domest Anim* 2011;46(2):23-5.
- [41] Siciliano L, Marciánò V, Carpino A. Prostatome-like vesicles stimulate acrosome reaction of pig spermatozoa. *Reprod Biol Endocrinol* 2008;6:5.
- [42] Machtinger R, Lauren LC, Baccarelli AA. Extracellular vesicles: roles in gamete maturation, fertilization and embryo implantation. *Human Reprod Update* 2016;22(2):182-93.

- [43] Rodríguez-Martínez H, Kvist U, Saravia F, Wallgren M, Johannisson A, Sanz L, Peña FJ, Martínez EA, Roca J, Vázquez JM, Calvete JJ. The physiological roles of the boar ejaculate. *Soc Reprod Fertil* 2009;66:1-21.
- [44] Barranco I, Rubér M, Perez-Patiño C, Atikuzzaman M, Martinez EA, Roca J, Rodriguez-Martinez H. The Seminal Plasma of the Boar is Rich in Cytokines, with Significant Individual and Intra-Ejaculate Variation. *Am J Reprod Immunol* 2015;doi: 10.1111/aji.12432.
- [45] Barranco I, Tvarijonaviciute A, Perez-Patiño C, Parrilla I, Ceron JJ, Martinez EA, Rodriguez-Martinez H, Roca J. High total antioxidant capacity of the porcine seminal plasma (SP-TAC) relates to sperm survival and fertility. *Sci. Rep.* 2015;doi: 10.1038/srep18538.
- [46] Dacheux JL, Dacheux F, Paquignon M. Changes in sperm surface membrane and luminal protein fluid content during epididymal transit in the boar. *Biol Reprod* 1989; 40(3):635-51.
- [47] Gadella BM, Colenbrander B, Van Golde LM, Lopes-Cardozo M. Boar seminal vesicles secrete arylsulfatases into seminal plasma: evidence that desulfation of seminolipid occurs only after ejaculation. *Biol Reprod.* 1993;48(3):483-9.
- [48] Wysocki P, Orzolek A, Strzeżek J, Koziorowska-Gilun M, Zasiadczyk Ł, Kordan W. The activity of N-acetyl-β-hexosaminidase in boar seminal plasma is linked with semen quality and its suitability for cryopreservation. *Theriogenology.* 2015;83(7):1194-202.
- [49] Guyonnet B, Marot G, Dacheux JL, Mercat MJ, Schwob S, Jaffrézic F, Gatti JL. The adult boar testicular and epididymal transcriptomes. *BMC Genomics.* 2009;10:369.
- [50] Watanabe H, Takeo T, Tojo H, Sakoh K, Berger T, Nakagata N, Mak TW, Kondoh G. Lipocalin 2 binds to membrane phosphatidylethanolamine to induce lipid raft movement in a PKA-dependent manner and modulates sperm maturation. *Development.* 2014;141(10):2157-64.
- [51] Carmona E, Weerachayanukul W, Soboloff T, Fluharty AL, White D, Promdee L, Ekker M, Berger T, Buhr M, Tanphaichitr N. Arylsulfatase a is present on the pig sperm surface and is involved in sperm-zona pellucida binding. *Dev Biol.* 2002;247(1):182-96.
- [52] Nakamura N, Dai Q, Williams J, Goulding EH, Willis WD, Brown PR, Eddy EM. Disruption of a spermatogenic cell-specific mouse enolase 4 (eno4) gene causes sperm structural defects and male infertility. *Biol Reprod.* 2013;88(4):90.
- [53] Zhang HY, Lu JC, Feng RX. Correlations of 24 biochemical markers in seminal plasma with routine semen parameters. *Zhonghua Nan Ke Xue.* 2015;21(12):1087-92.
- [54] Carballada R, Esponda P. Regulation of foreign DNA uptake by mouse spermatozoa. *Exp Cell Res.* 2001;262(2):104-13.
- [55] Cheng GZ, Li JY, Li F, Wang HY, Shi GX. Human ribonuclease 9, a member of ribonuclease A superfamily, specifically expressed in epididymis, is a novel sperm-binding protein. *Asian J Androl.* 2009;11(2):240-51.
- [56] Gonzalez R, Mohan H, Unniappan S. Nucleobindins: bioactive precursor proteins encoding putative endocrine factors? *Gen Comp Endocrinol.* 2012;176(3):341-6.
- [57] Vilagran I, Yeste M, Sancho S, Castillo J, Oliva R, Bonet S. Comparative analysis of boar seminal plasma proteome from different freezability ejaculates and identification of Fibronectin 1 as sperm freezability marker. *Andrology.* 2015;3(2):345-56.

- [58] Bucci D, Giaretta E, Spinaci M, Rizzato G, et al. Characterization of alkaline phosphatase activity in seminal plasma and in fresh and frozen-thawed stallion spermatozoa. *Theriogenology* 2016;85(2):288-295.
- [59] Nakata H, Wakayama T, Adthapanyawanich K, Nishiuchi T, Murakami Y, Takai Y, Iseki S. Compensatory upregulation of myelin protein zero-like 2 expression in spermatogenic cells in cell adhesion molecule-1-deficient mice. *Acta Histochem Cytochem.* 2012;45(1):47-56.
- [60] Hu Y, Xing J, Chen L, Guo X, Du Y, Zhao C, Zhu Y, Lin M, Zhou Z, Sha J. RGS22, a novel testis-specific regulator of G-protein signaling involved in human and mouse spermiogenesis along with GNA12/13 subunits. *Biol Reprod.* 2008;79(6):1021-9.
- [61] Li N, Mruk DD, Wong CK, Lee WM, Han D, Cheng CY. Actin-bundling protein plastin 3 is a regulator of ectoplasmic specialization dynamics during spermatogenesis in the rat testis. *FASEB J.* 2015;29(9):3788-805.
- [62] Hall JC, Killian GJ. Changes in rat sperm membrane glycosidase activities and carbohydrate and protein contents associated with epididymal transit. *Biol Reprod.* 1987;36(3):709-18.
- [63] Furuya S, Endo Y, Oba M, Ito T, Suzuki S, Nozawa S. Effect of epidermal growth factor (EGF) on mouse sperm capacitation. *Nihon Sanka Fujinka Gakkai Zasshi.* 1993;45(11):1313-9.
- [64] Douard V, Gassama-Diagne A, Hermier D, Blesbois E. Activity of phospholipases A and lysophospholipase in turkey semen and oviducal fluid. *Poult Sci.* 2004;83(8):1385-93.
- [65] Florek M, Bauer N, Janich P, Wilsch-Braeuninger M, Fargeas CA, Marzesco AM, Ehninger G, Thiele C, Huttner WB, Corbeil D. Prominin-2 is a cholesterol-binding protein associated with apical and basolateral plasmalemmal protrusions in polarized epithelial cells and released into urine. *Cell Tissue Res.* 2007;328(1):31-47.
- [66] Salvolini E, Buldreghini E, Lucarini G, Vignini A, Lenzi A, Di Primio R, Balercia G. Involvement of sperm plasma membrane and cytoskeletal proteins in human male infertility. *Fertil Steril.* 2013;99(3):697-704.
- [67] Tomes CN, Michaut M, De Blas G, Visconti P, Matti U, Mayorga LS. SNARE complex assembly is required for human sperm acrosome reaction. *Dev Biol.* 2002;243(2):326-38.
- [68] Lowe DG, Moran LA. Proteins related to the mouse L-cell major heat shock protein are synthesized in the absence of heat shock gene expression. *Proc Natl Acad Sci U S A.* 1984;81(8):2317-21.
- [69] Dun MD, Aitken RJ, Nixon B. The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. *Hum Reprod Update.* 2012;18(4):420-35.
- [70] Vilagran I, Yeste M, Sancho S, Casas I, Rivera del Álamo MM, Bonet S. Relationship of sperm small heat-shock protein 10 and voltage-dependent anion channel 2 with semen freezability in boars. *Theriogenology.* 2014;82(3):418-26.
- [71] Belleannée C, Labas V, Teixeira-Gomes AP, Gatti JL, Dacheux JL, Dacheux F. Identification of luminal and secreted proteins in bull epididymis. *J Proteomics.* 2011;74(1):59-78.

Figure Captions

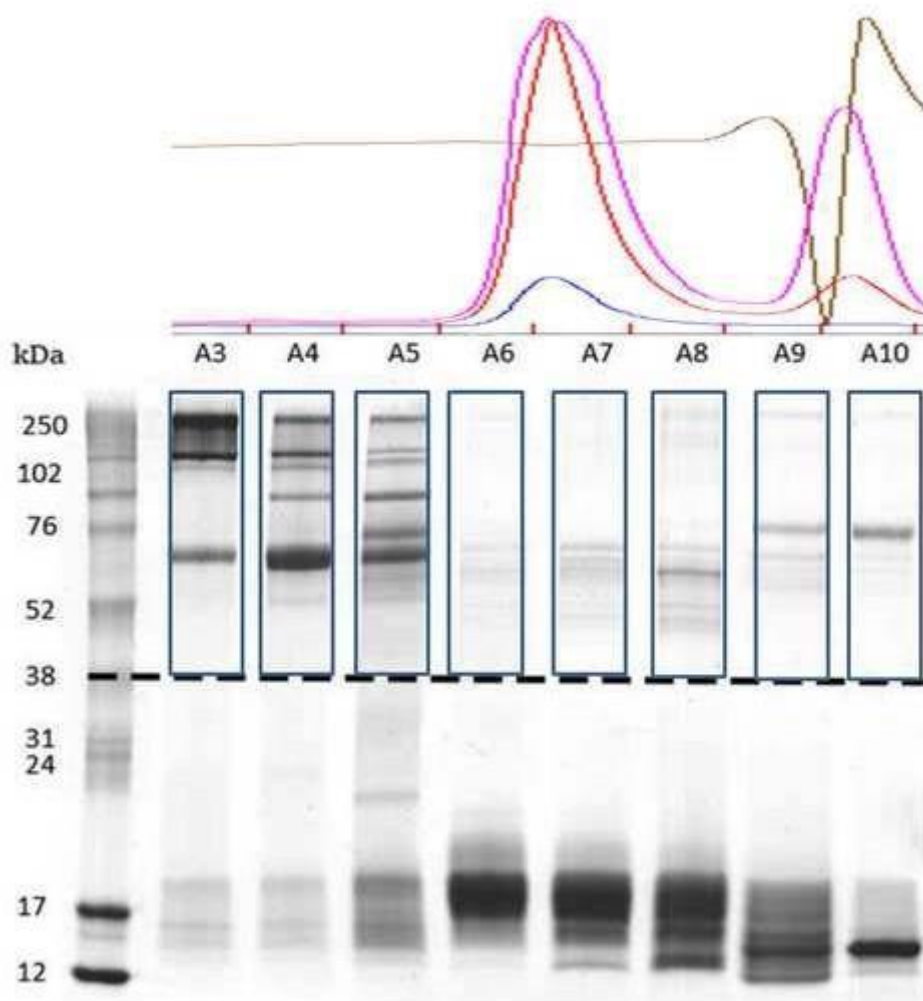


Fig.

1 - Representative 1-D SDS PAGE resulting from the electrophoresis of SEC fractions obtained from a pool of boar seminal plasma (SP, 33 ejaculates). The first column corresponds to a molecular weight marker (ECL Plex Fluorescent Rainbow Marker, GE Healthcare Life Sciences) and the rest of columns (A3-A10) correspond to different times eluted fractions obtained by SEC when the pooled boar SP was injected in an ETTAN LC system (GE Healthcare Life Sciences). The gel was sliced at 38 kDa in order to reject dominant proteins (low molecular weight proteins, <38 kDa) and the top of gel used to analyze minor proteins (high molecular weight proteins, >38 kDa) by in-gel digestion processing.

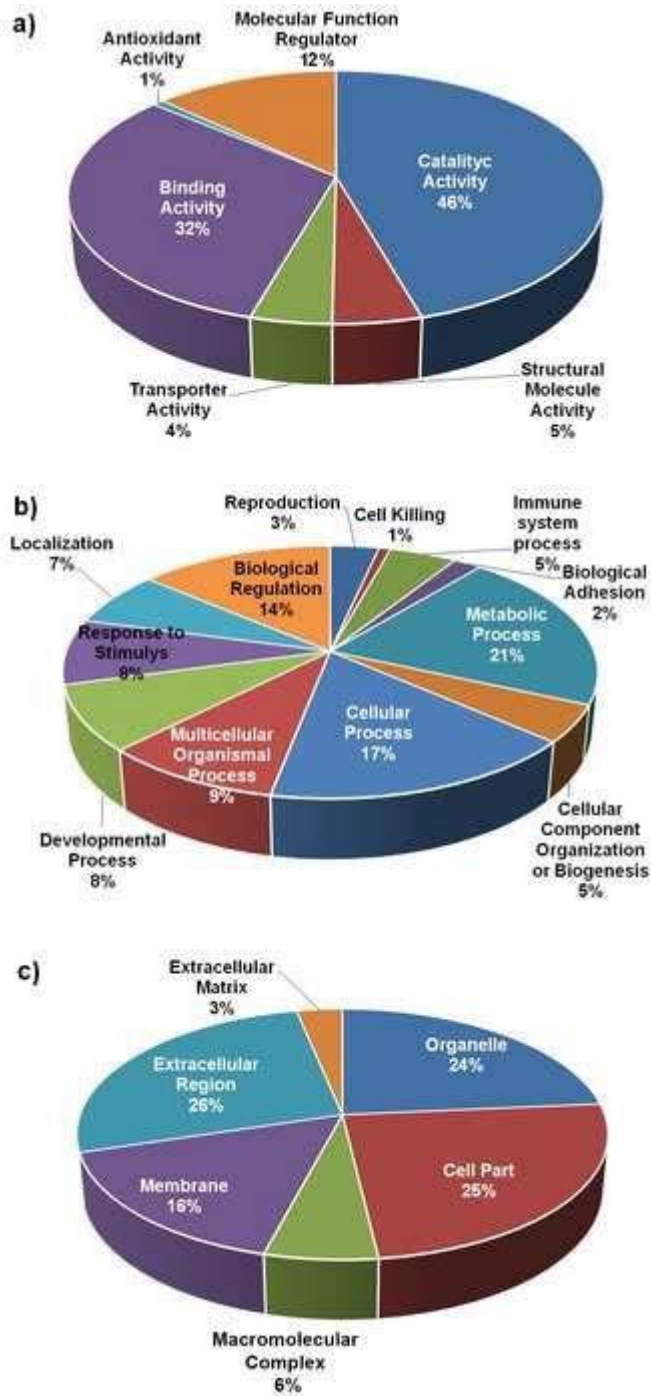


Fig. 2 - Pie charts representing the distribution of the proteins identified in boar seminal plasma according to (a) molecular function; (b) biological process and (c) cellular component, using UniProt KB database (www.uniprot.org) in combination with PANTHER (www.pantherdb.org).

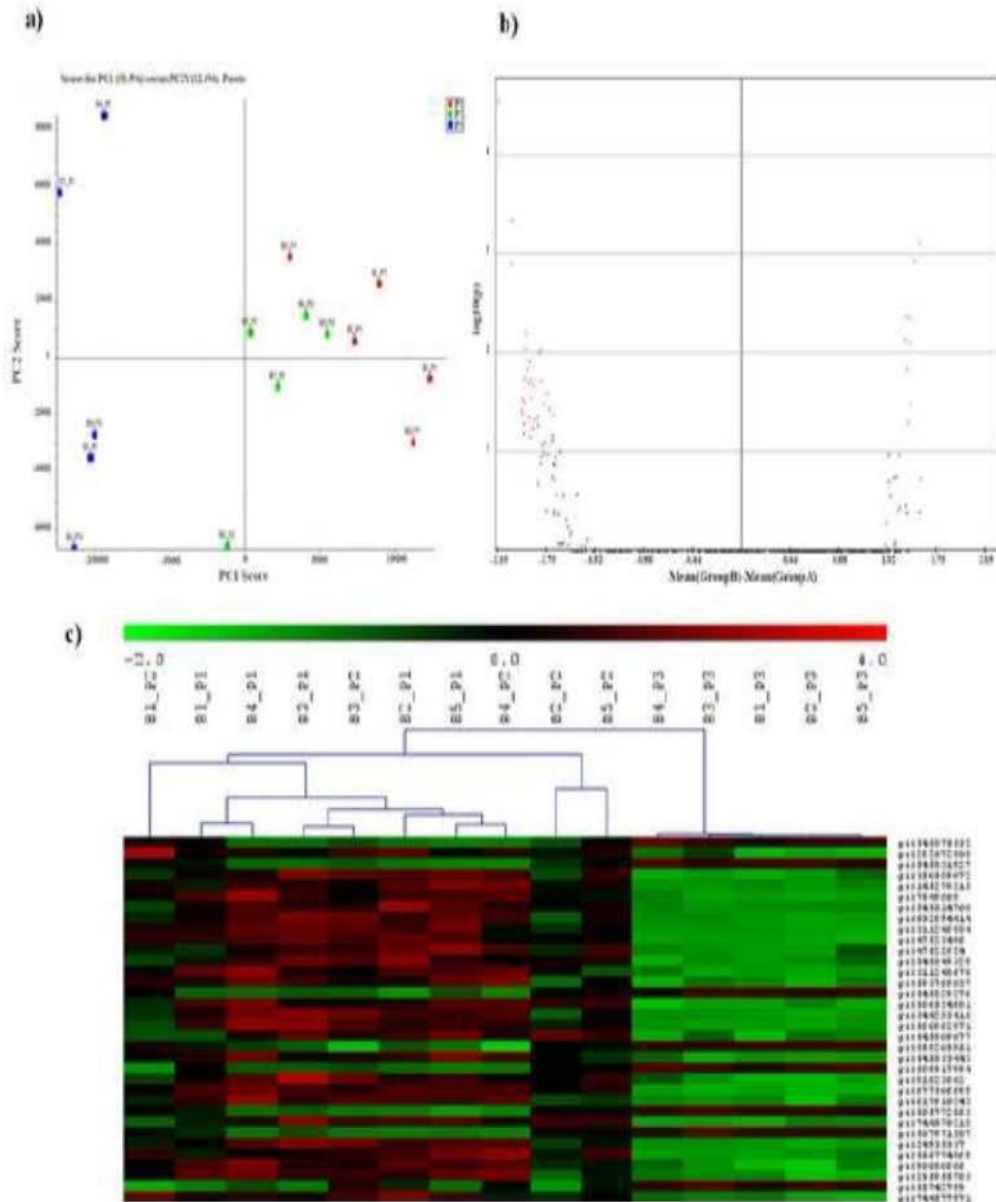


Fig.

3. Bioinformatic analysis of proteins identified and quantified using label-free LC-MS/MS analysis in 15 boar seminal plasma samples of three different ejaculate portions: (P1: first 10 mL of sperm rich ejaculate fraction (SRF), P2: rest of SRF, and P3: post-SRF) from 5 boars. (A) PCA showing the separation between P1-P2 samples and P3 samples in each of the five boars, based on relative protein amount. (B) Volcano plot showing differences in protein expression between P1-P2 and P3 according to magnitude and p-value (t-test). 34 proteins were differentially ($p < 0.05$) expressed between P1-P2 and P3 (indicated as red points). (C) Heat map, showing levels of differentially expressed proteins among ejaculate portions, and hierarchical clustering, showing two main clusters comprising P1-P2 samples and P3 samples, respectively.

Table 2. List of differentially expressed boar seminal plasma proteins, not included in *Sus Scrofa* taxonomy, between sperm rich ejaculate fraction (SRF) and post-SRF.

gi Accession number	Protein Name	Specie	Gene Id	FC (log2)*	p value (T-Test)
Over-expressed proteins in SRF					
gi 677306699	Heat shock cognate 71 kDa protein	Apaloderma vittatum	HSPA8	0.645	<0.001
gi 507971587	Putative phospholipase B-like 2	Condylura cristata	PLBD2	0.886	0.004
gi 617610243	Guanine nucleotide-binding protein subunit alpha-11	Erinaceus europaeus	GNA11	1.984	0.047
gi 90080606	Unnamed protein product	Macaca fascicularis	PGK1	0.652	0.006
gi 558205951	Polypeptide N-acetylgalactosaminyltransferase 2	Myotis lucifugus	GALNT2	0.762	0.014
gi 589917994	Fibronectin	Peromyscus maniculatus bairdii	FN1	0.449	<0.001
gi 593766037	Ezrin	Physeter catodon	EZR	1.873	0.018
gi 505772863	Fibronectin	Sorex araneus	FN1	1.399	0.004
Down-expressed proteins in SRF					
gi 744570218	EGF-like repeat and discoidin I-like domain-containing protein 3	Camelus dromedarius	EDIL3	-0.582	0.012
gi 744577971	Myelin protein zero-like protein 1	Camelus dromedarius	MPZL1	-1.819	0.009
gi 7549809	Plastin-3 isoform 1	Homo Sapiens	PLS3	-0.834	0.018
gi 91823602	Ectonucleotide pyrophosphatase/phosphodiesterase	Homo Sapiens	ENPP2	-1.266	0.044
gi 602694414	Alkaline phosphatase	Lipotes vexillifer	ALPL	-1.112	0.031
gi 344283415	Alkaline phosphatase	Loxodonta africana	ALPL	-0.964	0.040
gi 556774869	Beta-galactosidase-1-like protein 2-like	Pantholops hodgsonii	GLB1L2	-0.002	0.020
gi 255955783	Pc21g16370	Penicillium rubens Wisconsin 54-1255	Pc21g16370	-1.155	0.019
gi 586524681	Syntaxin-binding protein 2	Pteropus alecto	STXBP2	-2.021	<0.001
gi 560962971	Prominin-2	Vicugna pacos	PROM2	-0.565	0.037

*FC: Fold change

SRF includes two ejaculate portions: first 10 ml of SRF (P1 in text) and rest of SRF (P2 in text).

Post-SRF is P3 in text.

Supplementary Table Captions

Supplementary Table 1 - Complete list of 389 boar seminal plasma proteins validated with $\geq 95\%$ Confidence (Unused Score ≥ 1.3) and a false discovery rate (FDR) $\leq 1\%$.

Supplementary Table 2 - Complete list of all matched peptides for 389 proteins identified in boar seminal plasma.

Supplementary Table 3

Supplementary Table 4

