



Data Article

Proteomics dataset of epididymal fluid, seminal plasma, and proteins loosely attached to epididymal and ejaculated sperm from Angus bulls



Saulo Menegatti Zoca^a, Emmalee J. Northrop-Albrecht^a, Julie A. Walker^a, Robert A. Cushman^b, George A. Perry^{a,1,*}

^a Department of Animal Science, South Dakota State University, Brookings, SD 57007, USA

^b USDA, ARS, Roman L. Hruska US Meat Animal Research Center, Clay Center, NE 68933, USA

ARTICLE INFO

Article history:

Received 9 March 2022

Revised 29 March 2022

Accepted 4 April 2022

Available online 15 April 2022

Dataset link: [Mass spec raw data files](#)
([Original data](#))

Keywords:

Bull

Ejaculation

Epididymis

Protein

Spermatozoa

Sperm longevity

ABSTRACT

Peptides and proteins were identified by liquid chromatography with tandem mass spectrometry analysis (LCMS-MS) on an Orbitrap Velos mass spectrometer to further understand biological mechanisms that regulate increased longevity in epididymis compared to ejaculated sperm. Semen from sexually mature bulls were collected and then bulls were slaughtered to collect epididymal samples from the cauda epididymis. All samples were centrifuged to separate spermatozoa from fluids. A high ionic solution was used to remove surface proteins from spermatozoa. Four unique samples were generated: (1) epididymal fluid, (2) seminal plasma (ejaculated fluid), and proteins stripped from (3) epididymal sperm, and (4) ejaculated sperm. Samples were analyzed by LCMS-MS, and data were interpreted with Protein Pilot 5. False discovery rate (FDR) was set at 1%. Unique proteins ($n = 458$) were identified in ejaculated samples, 178 proteins in seminal plasma and 298 proteins stripped from ejaculated sperm. In epididymal samples, 311 proteins were identified in the fluid, and 334 were identified among proteins stripped from epididymal sperm. This dataset can be useful

DOI of original article: [10.1016/j.theriogenology.2022.02.021](https://doi.org/10.1016/j.theriogenology.2022.02.021)

* Corresponding author.

E-mail address: George.Perry@ag.tamu.edu (G.A. Perry).

¹ Present address: Texas A&M AgriLife Research and Extension Center, Overton, TX 75684, USA.

<https://doi.org/10.1016/j.dib.2022.108150>

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

for further understand of biological mechanisms that control sperm longevity. This dataset is related to the article 'Proteomic analyses identify differences between bovine epididymal and ejaculated spermatozoa that contribute to longevity' by (Zoca et al., 2022).

© 2022 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Biological Sciences: Animal Physiology
Specific subject area	Sperm biology; sperm maturation; association of epididymal and ejaculated fluids and loosely attached to the sperm proteins to sperm function
Type of data	Tables
How data were acquired	Liquid chromatography with tandem mass spectrometry analysis (LCMS-MS) on an Orbitrap Velos mass spectrometer.
Data format	Raw Filtered Analyzed
Description of data collection	Semen was collected from nine bulls. After a resting period to renormalize epididymal reserves bulls were slaughtered. Epididymides were dissected and epididymal fluid and spermatozoa were collected from the cauda epididymis. All samples were centrifuged to separate spermatozoa and fluids (epididymal fluid or seminal plasma). Spermatozoa were washed with a high ionic solution to remove proteins attached to the sperm surface. Samples were then centrifuged to separate stripped proteins from the spermatozoa.
Data source location	Open Prairie South Dakota State University Brookings, South Dakota United States
Data accessibility	Open Prairie repository. Article with all data and analysis etd2/221: https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=2&article=1205&context=etd2&type=additional Supplemental file 1 identified proteins fluid – 3.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=3&article=1205&context=etd2&type=additional Supplemental file 2 identified proteins sperm – 4.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=4&article=1205&context=etd2&type=additional Supplemental file 3 Gene ontology fluid – 5.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=5&article=1205&context=etd2&type=additional Supplemental file 4 Gene ontology sperm – 6.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=6&article=1205&context=etd2&type=additional Supplemental file 5 KEGG analysis fluid – 7.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=7&article=1205&context=etd2&type=additional Supplemental file 6 KEGG analysis sperm – 8.1205 https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=8&article=1205&context=etd2&type=additional Mass spec raw data files 2.1205: https://openprairie.sdstate.edu/cgi/viewcontent.cgi?filename=2&article=1205&context=etd2&type=additional
Related research article	S. Menegatti Zoca, E.J. Northrop-Albrecht, J.A. Walker, R.A. Cushman, and G.A. Perry, Proteomic analyses identify differences between bovine epididymal and ejaculated spermatozoa that contribute to longevity, <i>Theriogenology</i> . 184 (2022):51-60. https://doi.org/10.1016/j.theriogenology.2022.02.021

Value of the Data

- This dataset provides a comprehensive proteomic analysis of seminal plasma and epididymal fluid of bulls, as well as proteins attached to the sperm in the epididymis or during ejacu-

lation. This dataset provides critical proteomic information related to bovine sperm function and its interaction with the environment.

- Sperm biologists and animal scientists can use this dataset as a basis for understanding the role of sperm proteins in bovine reproductive efficiency.
- This dataset can be useful for further understanding of biological mechanisms that control sperm longevity in the epididymis compared to ejaculated sperm, particularly in bovine.

1. Data Description

Liquid chromatography with tandem mass spectrometry analysis identified proteins in the fluid fractions (seminal plasma and epididymal fluid) and attached to sperm (ejaculated and epididymal). Supplemental files 1 and 2 show identified proteins and includes Protein name, species, official gene symbol, accession number, protein molecular weight, protein grouping ambiguity, spectra count, and differences between ejaculated and epididymal samples. In Supplemental file 3, there are 38 GO terms related to proteins identified in ejaculated fluid of which 26 had $FDR < 0.05$; there are 58 GO terms related to proteins present in epididymal fluid of which 26 had $FDR < 0.05$. There are also 30 GO terms that are present in both ejaculated and epididymal fluid. Supplemental file 4, contains 50 GO terms related to proteins stripped from ejaculated sperm of which 23 had $FDR < 0.05$; there are 56 GO terms related to proteins stripped from epididymal sperm. There are also 38 GO terms in common between samples from epididymal sperm and ejaculated sperm. Fluids and loosely attached sperm protein KEGG, count, protein gene symbol per KEGG and FDR are presented in Supplemental files 5 and 6, respectively.

2. Experimental Design, Materials and Methods

2.1. Experimental design

Semen (collected by electroejaculation) from 4-yr old sexually mature bulls ($n = 9$) that had previously successfully bred cows was collected weekly (2 ejaculates) and discarded. After one week of rest bulls were collected for the samples utilized in the analysis. Before being slaughtered, bulls were rested for six weeks (following third semen collection) to allow the epididymal reserve to renormalize. Testes and epididymides were collected at the local abattoir and transported back to the laboratory where the epididymis was dissected and epididymal fluid and spermatozoa were collected from the cauda epididymis [1].

2.2. Protein isolation

Spermatozoa were separated from fluids (epididymal fluid or seminal plasma) by centrifugation ($700 \times g$ for 10 min). The supernatant was removed, snap frozen in liquid nitrogen, and stored at -80°C until analysis occurred. A high ionic solution was used to wash the spermatozoa [2] by vortexing for 1 min to remove loosely attached surface proteins. All samples were again centrifuged ($700 \times g$ for 10 min) and the supernatant was collected, snap frozen in liquid nitrogen, and stored at -80°C until analysis occurred. In total four sample types were collected: (1) epididymal fluid, (2) ejaculated fluid, and proteins stripped from (3) epididymal sperm, and (4) ejaculated sperm.

2.3. Liquid chromatography mass spectrometry analysis

Samples from each of the nine bulls were pooled together by sample type [(1) epididymal fluid, (2) ejaculated fluid, (3) epididymal sperm stripped proteins (epididymal sperm), and (4)

ejaculated sperm stripped proteins (ejaculated sperm)]. This resulted in one pool for each sample type. Samples were shipped to the University of Minnesota Mass Spectrometry facility, and liquid chromatography with tandem mass spectrometry was conducted there. Briefly, a 5 μ g aliquot of each sample was mixed with 4X Laemmli buffer (final concentration 1X), heated at 95 °C for 5 min, and then loaded on an 8–16% BioRad Criterion™ Gel. The gel was run at 25 mA constant current for 30 min. The gel was stained with Imperial™ Protein Stain (ThermoFisher Scientific). The protein region for each sample was cut out, in-gel trypsin digested, and Stage Tip cleaned up as previously published [3] with the following change: the samples were alkylated with iodoacetamide. Peptides (~400 ng of reconstituted peptides) were then analyzed on an Orbitrap Velos mass spectrometer [4] with the following modifications: (1) capillary column diameter was 100 μ m, (2) gradient elution profile was of 8–35% B Solvent over 67 min at 330 nL/min [A Solvent was 98:2:0.01, H₂O:acetonitrile(ACN):formic acid (FA); and B Solvent was 98:2:0.01, ACN:H₂O:FA], (3) lock mass was not employed.

Dynamic exclusion settings were:

1. repeat count = 1,
2. exclusion list size was 200,
3. exclusion duration = 12 s,
4. exclusion mass width (high and low) was 15 ppm,
5. early expiration was disabled.

2.4. Database searching and criteria for protein identification

Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 2.1.0.81) was used to analyze all LCMS/MS data. The Uniprot.org bovine (taxid 9913) database of protein sequences was searched (downloaded on March 6, 2013). Both canonical and isoforms were included. Searched database was merged with a database containing the common lab contaminant proteins (thegpm.org/crap/index, 109 proteins). The search included: (1) trypsin as the digestion enzyme, (2) fragment ion mass tolerance was set at 0.100 Da, and (3) precursor tolerance was set at 50 ppm. Variable peptide modification included: (1) Oxidation and di-oxidation of methionine, (2) deamidated of asparagine, glutamine and pyroglutamic acid, and (3) N-terminal protein acetylation set as a variable modification and carbamidomethyl cysteine set as a fixed modification.

All LCMS/MS based peptide and protein identifications were validated with Scaffold (version Scaffold_5.0.0, Proteome Software Inc., Portland, OR). Peptide identifications were accepted at greater than 99.0% probability by the Scaffold Local FDR algorithm. Identified proteins that established greater than 7.0% probability, with an FDR less than 1.0%, and peptides identified were greater than or equal to 2 were accepted. Protein Prophet algorithm [5] was used to assign protein probabilities. The principles of parsimony were used to group proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone. Clusters were formed by grouping proteins sharing significant peptide evidence.

Total spectrum count was used to identify proteins that were present in each sample (≥ 0.8 peptides identified as exclusive and unique to each protein). The official gene name of proteins identified in each of the four sample types were entered into DAVID v 6.8 [6,7] with *Bos taurus* as background to determine the top Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (thresholds: count = 2, EASE = 0.1). The FDR for each Gene Ontology and KEGG was generated by DAVID v 6.8.

Ethics Statement

All procedures were approved by the South Dakota State University Institutional Animal Care and Use Committee (#14-031A) and conformed to the Guide for the Care and Use of Agriculture Animals in Research and Teaching.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Data Availability

[Mass spec raw data files \(Original data\)](#) (Open Prairie).

CRedit Author Statement

Saulo Menegatti Zoca: Formal analysis, Investigation, Writing – original draft, Writing – review & editing; **Emmalee J. Northrop-Albrecht:** Investigation, Writing – review & editing; **Julie A. Walker:** Writing – review & editing, Supervision; **Robert A. Cushman:** Conceptualization, Investigation, Writing – review & editing, Supervision; **George A. Perry:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This project was supported by the [National Science Foundation](#) BREAD EAGER [Grant No. 1446886]. Names are necessary to report factually on available data; however, neither South Dakota State University nor the USDA guarantees nor warrants the standard of the product, and the use of names by South Dakota State University and the USDA implies no approval of the product to the exclusion of others that may also be suitable. South Dakota State University and USDA are equal opportunity providers and employers. The authors acknowledge the Center for Mass Spectrometry and Proteomics at the University of Minnesota whose supporting agencies are listed here: <https://cbs.umn.edu/cmstp/about>.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2022.108150](https://doi.org/10.1016/j.dib.2022.108150).

References

- [1] S.M. Zoca, E.J. Northrop-Albrecht, J.A. Walker, R.A. Cushman, G.A. Perry, Proteomic analyses identify differences between bovine epididymal and ejaculated spermatozoa that contribute to longevity, *Theriogenology* (2022) 51–60, doi:[10.1016/j.theriogenology.2022.02.021](https://doi.org/10.1016/j.theriogenology.2022.02.021).
- [2] J.M. Rifkin, G.E. Olson, Characterization of maturation-dependent extrinsic proteins of the rat sperm surface, *J. Cell Biol.* 100 (1985) 1582–1591.
- [3] Y.M. Thu, S.K. Van Riper, L. Higgins, T. Zhang, J.R. Becker, T.W. Markowski, et al., Slx5/Slx8 promotes replication stress tolerance by facilitating mitotic progression, *Cell Rep.* 15 (2016) 1254–1265.
- [4] Y. Lin-Moshier, P.J. Sebastian, L. Higgins, N.D. Sampson, J.E. Hewitt, J.S. Marchant, Re-evaluation of the role of calcium homeostasis endoplasmic reticulum protein (CHERP) in cellular calcium signaling, *J. Biol. Chem.* 288 (2013) 355–367.

- [5] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem.* 75 (2003) 4646–4658.
- [6] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* 4 (2008) 44–57.
- [7] D.W. Huang, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.* 37 (2009) 1–13.