

Quantitative Metabolomic Dataset of Avian Eye Lenses

Ekaterina A. Zelentsova ^{1,*}, Sofia S. Mariasina ^{2,3,4} , Vadim V. Yanshole ^{1,5} , Lyudmila V. Yanshole ¹ , Nataliya A. Osik ^{1,5}, Kirill A. Sharshov ⁶ and Yuri P. Tsentlovich ^{1,*} 

¹ Laboratory of Proteomics and Metabolomics, International Tomography Center SB RAS, Institutskaya 3a, Novosibirsk 630090, Russia; vadim.yanshole@tomo.nsc.ru (V.V.Y.); lucy@tomo.nsc.ru (L.V.Y.); n.osik@tomo.nsc.ru (N.A.O.)

² Faculty of Fundamental Medicine, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russia; sofia.mariasina@yandex.ru

³ Institute of Functional Genomics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119992, Russia

⁴ Department of Chemistry, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russia

⁵ Faculty of Physics, Novosibirsk State University, Pirogova Street 1, Novosibirsk 630090, Russia

⁶ Laboratory of Molecular Epidemiology and Biodiversity of Viruses, Federal Research Center of Fundamental and Translational Medicine, Timakova Street 2, Novosibirsk 630117, Russia; sharshov@yandex.ru

* Correspondence: zelentsova@tomo.nsc.ru (E.A.Z.); yura@tomo.nsc.ru (Y.P.T.); Tel.: +7-383-330-3136

Abstract: Metabolomics is a powerful set of methods that uses analytical techniques to identify and quantify metabolites in biological samples, providing a snapshot of the metabolic state of a biological system. In medicine, metabolomics may help to reveal the molecular basis of a disease, make a diagnosis, and monitor treatment responses, while in agriculture, it can improve crop yields and plant breeding. However, animal metabolomics faces several challenges due to the complexity and diversity of animal metabolomes, the lack of standardized protocols, and the difficulty in interpreting metabolomic data. The current dataset includes quantitative metabolomic profiles of eye lenses from 26 bird species (111 specimens) that can aid researchers in developing new experiments, mathematical models, and integrating with other “-omics” data. The dataset includes raw ¹H NMR spectra, protocols for sample preparation, and data preprocessing, with the final table containing information on the abundance of 89 reliably identified and quantified metabolites. The dataset is quantitative, making it relevant for supplementing with new specimens or comparison groups, followed by data mining and expected new interpretations. The data were obtained using the bird specimens collected in compliance with ethical standards and revealed potential differences in metabolic pathways due to phylogenetic differences or environmental exposure.

Dataset: The dataset, including NMR raw data, metabolite concentrations, and experimental protocols, has been deposited in the MetaboLights repository, study identifier MTBLS7739 (<https://www.ebi.ac.uk/metabolights/MTBLS7739> (accessed on 30 June 2023)). DOI: 10.1093/nar/gkz1019, PMID: 31691833. This dataset is now also available through the laboratory’s own database, accessible via the following links: amdb.online/amdb/experiments/145/; amdb.online/amdb/experiments/199/; amdb.online/amdb/experiments/255/; amdb.online/amdb/experiments/248/.

Dataset License: CC0

Keywords: quantitative metabolomics; NMR spectroscopy; eye lens; birds



Citation: Zelentsova, E.A.; Mariasina, S.S.; Yanshole, V.V.; Yanshole, L.V.; Osik, N.A.; Sharshov, K.A.; Tsentlovich, Y.P. Quantitative Metabolomic Dataset of Avian Eye Lenses. *Data* **2023**, *8*, 125. <https://doi.org/10.3390/data8080125>

Academic Editor: Rüdiger Pryss

Received: 27 April 2023

Revised: 19 July 2023

Accepted: 20 July 2023

Published: 31 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Summary

Metabolomics is a field of study that aims to identify and quantify the small molecules, or metabolites, present in biological samples, such as cells, tissues, or fluids. These molecules are initial, intermediate, and end products of cellular processes that provide a

snapshot of the metabolic state of a biological system. Metabolomics uses many analytical techniques, such as liquid chromatography—mass spectrometry (LC–MS) and nuclear magnetic resonance (NMR) spectroscopy—to measure and identify metabolites and generate metabolomic profiles. In medicine, metabolomics can be used to identify disease biomarkers, reveal the molecular basis of a disease, and monitor treatment responses. In agriculture, metabolomics can help to improve crop yields and to identify markers for plant breeding. In environmental science, metabolomics can be used to assess the impact of pollutants on ecosystems and monitor the health of natural systems. Overall, metabolomics provides a powerful tool for understanding biological systems and for developing new diagnostic, therapeutic, and environmental solutions.

Metabolomics of animals is a rapidly developing branch of “-omics” science, but it still faces several challenges that hinder its progress. One of the main difficulties is the complexity and diversity of animal metabolomes. Different animal species have distinct metabolomes, and even conspecific individuals can exhibit variations in their metabolomic profiles. This heterogeneity requires a large sample size to draw meaningful conclusions, making the analysis time-consuming and costly. Additionally, many metabolites in animal samples are present in low abundance, making their detection and quantification challenging. The lack of standardized sample preparation and analysis protocols also presents a challenge, making it difficult to compare results across different studies. Furthermore, the interpretation of metabolomic data can be complex due to the interactions between various metabolic pathways and the effects of external factors such as diet, environment, and genetics. These challenges have limited the widespread adoption of animal metabolomics and highlighted the need for continued development of new analytical techniques, standardization of protocols, and integration with other -omics disciplines. In addition, there is an insufficient number of publicly available datasets in the field of animal metabolomics, making it difficult for researchers to build on previous work.

The current dataset is a set of quantitative metabolomic profiles of lens tissues from 26 bird species from various orders. The main method of identifying and quantifying metabolites in tissue was NMR spectroscopy (LC–MS was used for complicated identification cases and cross-identification). The dataset includes not only the final table with metabolite concentrations but also all the raw NMR data and detailed protocols for sample preparation and data preprocessing. A part of this dataset formed the basis of a pilot study on the use of quantitative metabolomic profiling for taxonomic differentiation of species [1]. Also, analysis of a subset of this dataset made it possible to discover a new molecular ultraviolet filter in the bird’s eye lens—nicotinamide adenine dinucleotide reduced (NADH) [2]. However, we are confident that the potential for using these data could be much wider, and making it available to the public will allow scientists to use it in their work. Metabolomic studies are most often represented by semi-quantitative LC–MS measurements, which are very protocol- and instrument-dependent and can hardly be re-used in other analyses. The data presented in this work are quantitative, making it as relevant as possible for supplementing with new samples or comparison groups or as a basis for completely new experiments. The data can also be integrated with other omics data, such as genomics, transcriptomics, and proteomics, to provide a more comprehensive understanding of biological processes and systems. Additionally, quantitative metabolomic data can be used to develop mathematical models of metabolic systems, which can aid in predicting metabolic responses to perturbations and designing new interventions.

2. Data Description

The dataset is represented by two main parts:

1. Raw data—¹H NMR spectra (a description of the acquisition protocol is provided below in the methods section);
2. Quantitative metabolomic data presented in the table (csv format), where the columns correspond to samples and the rows correspond to metabolites. Concentrations in the table are given in nmol per gram of wet tissue weight.

The research was carried out in line with the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research and the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes. Additionally, the study was granted ethical approval by the International Tomography Center SB RAS (ECITC-2017-02).

The bird specimens were obtained from three different sources (Table 1): (1) during hunting season with official authorization from regional Ministries of Ecology and Natural Resources (Dagestan Republic; Altay Republic; Omsk Region; Tyva Republic; Republic of Sakha; Novosibirsk Region, Russia) for the purpose of collecting biological material as part of the annual program for studying infectious diseases in wild animals, which was approved by the Biomedical Ethics Committee of FRC FTM, Novosibirsk, Russia (Protocols No. 2013-23 and 2021-10); (2) provided by the Center for the Rehabilitation of Wild Animals (CRWA, Novosibirsk, Russia) following humane euthanasia of birds that were mortally wounded; and (3) obtained through a special permit for scientific purposes from the Committee for the Protection of the World's Wild Animals of the Republic of Altay, Russia (#5, 21 August 2018).

Table 1. Species and sample descriptions.

Species	Date and Place of Catching	N of Samples
Black-headed gull (<i>Chroicocephalus ridibundus</i>)	May 2019, Tyva Republic	7 individuals
Black kite (<i>Milvus migrans</i>)	July 2019, Tyva Republic	5 individuals
Common buzzard (<i>Buteo buteo</i>)	July 2019, Novosibirsk Region	1 individual
Common magpie (<i>Pica pica</i>)	December 2018–January 2019, Altay Republic	4 individuals
Common pochard (<i>Aythya ferina</i>)	August 2020, Novosibirsk Region	7 individuals
Common raven (<i>Corvus corax</i>)	December 2018, Altay Republic; July 2019, Tyva Republic; December 2019, Novosibirsk Region	5 individuals
Eurasian coot (<i>Fulica atra</i>)	April 2019, Novosibirsk Region; May 2019, Tyva Republic	4 individuals
European herring gull (<i>Larus argentatus</i>)	July 2019, Tyva Republic	5 individuals
Godlewski's bunting (<i>Emberiza godlewskii</i>)	January 2019, Altay Republic	18 individuals, 6 samples
Great crested grebe (<i>Podiceps cristatus</i>)	May 2019, Tyva Republic	5 individuals
Great grey shrike (<i>Lanius excubitor</i>)	August 2019, Novosibirsk Region	1 individual
Great tit (<i>Parus major</i>)	December 2018, Altay Republic	12 individuals, 6 samples
Hawfinch (<i>Coccothraustes coccothraustes</i>)	December 2018, Altay Republic	6 individuals, 3 samples
Hooded crow (<i>Corvus cornix</i>)	January 2019, Altay Republic; April 2019, Novosibirsk Region	3 individuals
Horned grebe (<i>Podiceps auritus</i>)	September 2019, Novosibirsk Region	1 individual
House sparrow (<i>Passer domesticus</i>)	November 2018, Novosibirsk Region	14 individuals, 7 samples
Little grebe (<i>Podiceps ruficollis</i>)	May 2019, Tyva Republic	1 individual
Mongolian gull (<i>Larus vegae mongolicus</i>)	May 2019, Tyva Republic	3 individuals
Pale sand martin (<i>Riparia diluta</i>)	June 2022, Republic of Sakha	14 individuals, 7 samples
Red-necked grebe (<i>Podiceps grisegena</i>)	September 2019, Novosibirsk Region	1 individual
Rock dove (<i>Columba livia</i>)	September 2017, Novosibirsk Region	12 individuals
Rook (<i>Corvus frugilegus</i>)	April 2019, Novosibirsk Region	5 individuals
Short-eared owl (<i>Asio flammeus</i>)	May 2019, CRWA, Novosibirsk	1 individual
Ural owl (<i>Strix uralensis</i>)	May 2019, CRWA, Novosibirsk	1 individual
Western marsh harrier (<i>Circus aeruginosus</i>)	September 2021, Dagestan Republic	4 individuals
Wild duck (<i>Anas platyrhynchos</i>)	August 2022, Omsk Region	6 individuals

As an example of NMR spectrum identification, Figure 1 shows the ¹H NMR spectrum with signal annotations obtained for the rook (*Corvus frugilegus*) lens.

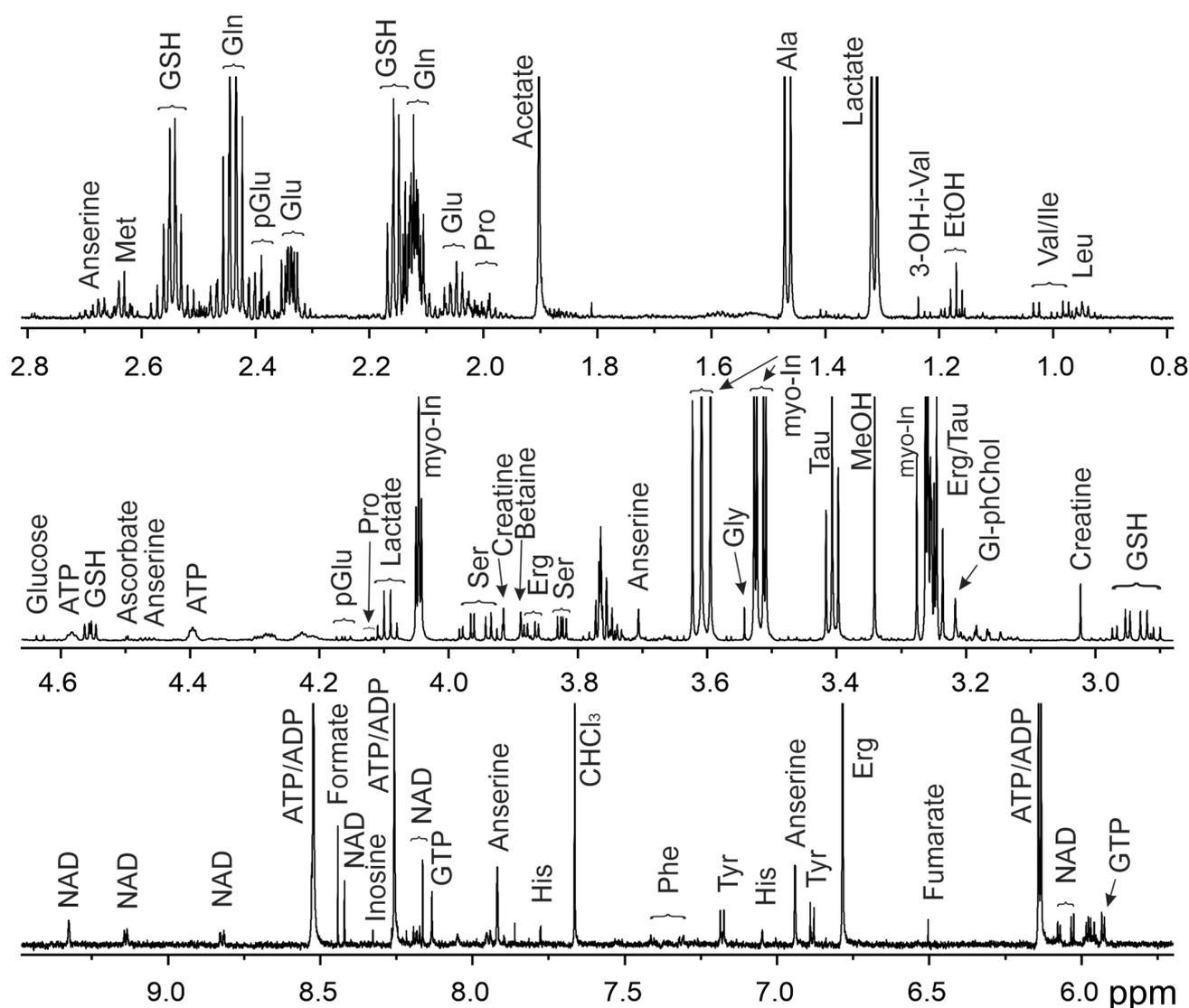


Figure 1. Representative ^1H NMR spectrum of bird lens metabolome, obtained for Rook (*Corvus frugilegus*) lens. Abbreviations: 3-OH-i-Val—3-hydroxyisovalerate, ADP—adenosine diphosphate, ATP—adenosine triphosphate, CHCl_3 —chloroform, Erg—ergothioneine, EtOH—ethanol, Gl-phChol—glycero-3-phosphocholine, GSH—glutathione reduced, GTP—guanosine triphosphate, MeOH—methanol, myo-In—myo-inositol, NAD—nicotinamide adenine dinucleotide, pGlu—pyroglutamate, Tau—taurine. For amino acids, a standard tree letter code is used.

As an example of a possible analysis of the published dataset and its visualization, we provide the results of several algorithms for dimension reduction. The most commonly used data visualization method in metabolomics is principal component analysis (PCA). Principal component analysis is a statistical technique used to reduce the complexity of a dataset by identifying and representing its important underlying variables through a smaller set of linearly uncorrelated variables, known as principal components. The advantage of PCA is its high interpretability of contributions of old components to forming new ones, but it is not suitable for analyzing a large number of sample groups. Figure 2 (right bottom) shows that the significant differences between the pale sand martin, Anatidae, and Passeridae does not allow other species to form separate clusters. Two other algorithms provide much better results for separating species groups (Figure 2, top and bottom left). UMAP (uniform manifold approximation and projection) [3] and t-SNE (t-distributed stochastic neighbor embedding) are non-linear dimension reduction algorithms used for visualizing high-dimensional data. The main difference between them is that UMAP

considers the global structure of the data and preserves more accurate distances between neighboring objects, while t-SNE tends to optimize locally and preserve higher object densities in the resulting clusters. UMAP's advantage is a faster runtime and better scalability, while t-SNE usually provides more intuitive visualizations for small datasets with many clusters. Figure 2 also demonstrates that for metabolomics data, UMAP better reflects the genetic relationship of species, i.e., better preserves the global structure, while t-SNE forms clusters by species, but they homogeneously fill the embedding space.

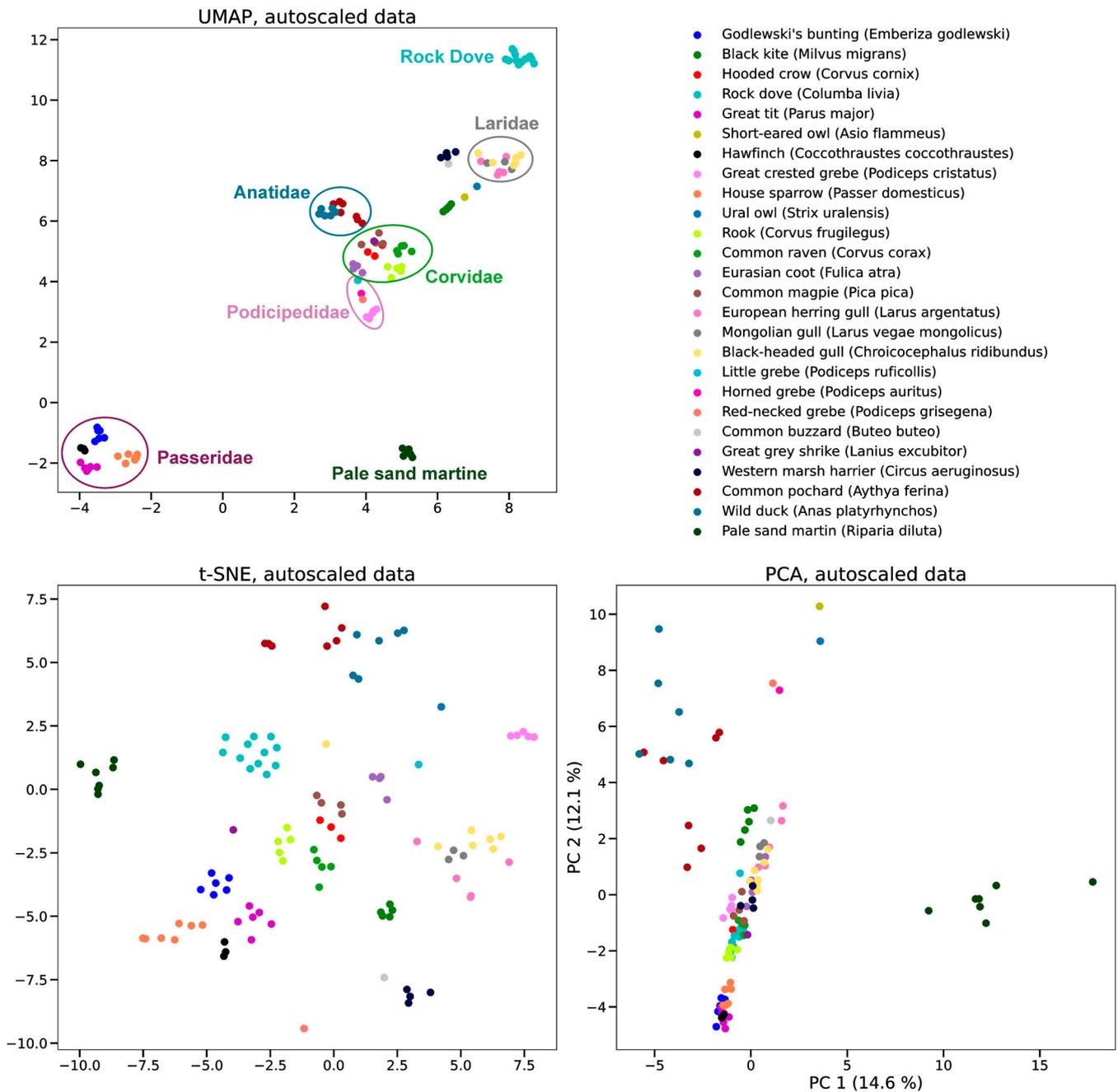


Figure 2. The figure displays the embeddings generated by three different dimensionality reduction methods: UMAP (top left), t-SNE (bottom left), and PCA (bottom right). Some species within the same

family form clusters: Passeridae (*Emberiza godlewskii*, *Passer domesticus*, *Coccothraustes coccothraustes*, *Parus major*), Laridae (*Larus argentatus*, *Larus vegae mongolicus*, *Chroicocephalus ridibundus*), Podicipedidae (*Podiceps ruficollis*, *Podiceps cristatus*, *Podiceps grisegena*, *Podiceps auritus*), Anatidae (*Anas platyrhynchos*, *Riparia diluta*), and Corvidae (*Fulica atra*, *Corvus corax*, *Pica pica*, *Corvus cornix*, *Corvus frugilegus*).

3. Methods

3.1. Sample Preparation

The process of preparing the lens samples was carried out following the detailed procedure described in [4]. The analyzed species are rather widespread in Siberia, and obtaining the samples was relatively straightforward. All individuals used in the study were adult wild-caught specimens collected between 2017 and 2022. After extraction from the eye, the lenses were cleaned, placed in individual cryotubes, frozen in liquid nitrogen, and stored at $-70\text{ }^{\circ}\text{C}$ until analyzed. For each sample, we analyzed one, two, or three lenses from different individuals depending on the lens size. Prior to homogenization, each sample was weighed. The lenses were homogenized in glass vials using a TissueRuptor II rotor-stator homogenizer (Qiagen, Venlo, The Netherlands) with 1600 μL of cold MeOH ($-20\text{ }^{\circ}\text{C}$), followed by the addition of 800 μL of water and 1600 μL of cold chloroform. The mixture was shaken for 20 min in a shaker and then left at $-20\text{ }^{\circ}\text{C}$ for 30 min. After that, the mixture was centrifuged at $16,100\times g$, $+4\text{ }^{\circ}\text{C}$ for 30 min, which resulted in two immiscible liquid layers separated by a lipid-protein layer. The upper aqueous layer (MeOH- H_2O) was collected, divided into two parts for NMR (2/3) and LC-MS (1/3) analyses, and vacuum-dried for further analysis.

3.2. NMR Measurements

For NMR measurements, the extracts were dissolved in 600 μL of D_2O containing $2\times 10^{-5}\text{ M}$ of sodium 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as an internal standard and 20 mM of deuterated phosphate buffer (pH 7.2). The ^1H NMR measurements were conducted at the “Mass Spectrometric Investigations” Center of Collective Use, SB RAS, using an AVANCE III HD 700 MHz NMR spectrometer (Bruker BioSpin, Ettlingen, Germany). The NMR spectra for each sample in a standard 5 mm glass NMR tube were acquired using a 5 mm TXI ATMA NMR probe by summing 96 transients while maintaining the sample temperature at $25\text{ }^{\circ}\text{C}$ and using a 90-degree detection pulse. To allow for the relaxation of all spins, a repetition time of 20 s was used between scans. Prior to acquisition, low-power radiation was applied at the water resonance frequency to presaturate the water signal.

3.3. LC-MS Measurements

The hydrophilic interaction liquid chromatography (HILIC) method was used for LC separation of the samples, utilizing a TSKgel Amide-80 HR column ($4.6\times 250\text{ mm}$, $5\text{ }\mu\text{m}$) on an UltiMate 3000RS chromatograph (Dionex, Germering, Germany). A diode array UV-vis detector (DAD) with a 190–800 nm spectral range was used with a flow cell. The mobile phase consisted of solvent A, a 0.1% formic acid solution in H_2O , and solvent B, a 0.1% formic acid solution in acetonitrile. The column temperature was set to $40\text{ }^{\circ}\text{C}$, and the flow rate was 1 mL/min. The injection volume for the samples was 10 μL . The gradient was as follows: 95% solvent B from 0 to 5 min, 95–65% from 5 to 32 min, 65–35% from 32 to 40 min, 35% from 40 to 48 min, 35–95% from 48 to 50 min, and 95% from 50 to 60 min. A flow splitter (1:10) directed the lesser flow after the DAD cell to an ESI-q-TOF high-resolution hybrid mass spectrometer maXis 4G (Bruker Daltonics, Bremen, Germany). The mass spectra were recorded in positive mode with a range of 50–1000 m/z . The MS setup, calibration procedure, and data processing were previously described [5].

3.4. Metabolite Identification and Quantification

Metabolite identification was conducted by analyzing their NMR spectra, sourced from literature, databases (HMDB, METLIN, BMRB, and SpectraBase), and our in-house NMR library [5–7]. In instances where NMR signal assignment was not straightforward, we verified the identification of metabolites by spiking the lens extract with commercially available standard compounds. To identify unknown signals, we fractionated the metabolomic extract by HILIC and conducted MS, MS/MS, and NMR analysis on each fraction, as detailed in [5]. Despite these efforts, several signals in the NMR spectra remained unassigned. The metabolite concentrations in the lenses were determined in nmol/g by integrating NMR signals relative to the internal standard DSS and then normalized to the wet weight of the tissue. The baseline processing, identification, and integration of spectral NMR peaks (quantification) were done using the MestReNova v12.0 (Mestrelab Research, A Coruna, Spain). On average, 60–80 compounds were identified for each species. However, quantifying some compounds was unreliable due to weak signals or overlapping with other signals. The final table contains only 89 reliably identified and quantified metabolites. The raw data could contain more information on yet unidentified metabolites.

Author Contributions: Conceptualization, E.A.Z.; methodology, Y.P.T.; validation V.V.Y. and Y.P.T.; formal analysis, E.A.Z., V.V.Y., L.V.Y., S.S.M. and Y.P.T.; investigation, E.A.Z., L.V.Y., N.A.O., S.S.M. and K.A.S.; data curation, V.V.Y. and Y.P.T.; writing—original draft preparation, E.A.Z.; writing—review and editing, V.V.Y., L.V.Y. and Y.P.T.; visualization, E.A.Z.; supervision, Y.P.T.; project administration, E.A.Z.; funding acquisition, E.A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant number 21-74-00068.

Institutional Review Board Statement: The study was conducted in accordance with the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research and the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes and with the ethical approval from the International Tomography Center SB RAS (ECITC-2017-02).

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw NMR spectra, the descriptions of specimens and samples, metabolite concentrations, and the preliminary metabolomic analysis are available at the MetaboLights repository, study identifier MTBLS7739 (<https://www.ebi.ac.uk/metabolights/MTBLS7739> (accessed on 30 June 2023)).

Acknowledgments: We thank the Ministry of Science and Higher Education of the RF for access to NMR and LC-MS equipment. Additionally, we extend our appreciation to the Novosibirsk Regional Ministry of Ecology and Natural Resources, the Center for the Rehabilitation of Wild Animals (CRWA) in Novosibirsk, and the Committee for the Protection of the World’s Wild Animals of the Republic of Altay for their help in obtaining the biological materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zelentsova, E.A.; Yanshole, L.V.; Tsentlovich, Y.P.; Sharshov, K.A.; Yanshole, V.V. The Application of Quantitative Metabolomics for the Taxonomic Differentiation of Birds. *Biology* **2022**, *11*, 1089. [[CrossRef](#)] [[PubMed](#)]
2. Osik, N.; Zelentsova, E.; Sharshov, K.; Tsentlovich, Y. Nicotinamide Adenine Dinucleotide Reduced (NADH) Is a Natural UV Filter of Certain Bird Lens. *Sci. Rep.* **2022**, *12*, 16850. [[CrossRef](#)] [[PubMed](#)]
3. McInnes, L.; Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**, arXiv:1802.03426. [[CrossRef](#)]
4. Tsentlovich, Y.P.; Yanshole, V.V.; Yanshole, L.V.; Zelentsova, E.A.; Melnikov, A.D.; Sagdeev, R.Z. Seasonal Variations and Interspecific Differences in Metabolomes of Freshwater Fish Tissues: Quantitative Metabolomic Profiles of Lenses and Gills. *Metabolites* **2019**, *9*, 264. [[CrossRef](#)] [[PubMed](#)]
5. Yanshole, L.; Zelentsova, E.; Tsentlovich, Y. Ovothioliol A Is the Main Antioxidant in Fish Lens. *Metabolites* **2019**, *9*, 95. [[CrossRef](#)] [[PubMed](#)]

6. Tsentlovich, Y.; Zelentsova, E.A.; Yanshole, L.V.; Yanshole, V.V.; Odud, I.M. Most Abundant Metabolites in Tissues of Freshwater Fish Pike-Perch (*Sander lucioperca*). *Sci. Rep.* **2020**, *10*, 17128. [[CrossRef](#)] [[PubMed](#)]
7. Tsentlovich, Y.P.; Verkhovod, T.D.; Yanshole, V.V.; Kiryutin, A.S.; Yanshole, L.V.; Fursova, A.Z.; Stepakov, D.A.; Novoselov, V.P.; Sagdeev, R.Z. Metabolomic Composition of Normal Aged and Cataractous Human Lenses. *Exp. Eye Res.* **2015**, *134*, 15–23. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.