^{1}H NMR metabolic profiling for the differentiation of fish species

Melike Saglam[1,2], Nele Paasch[2], Anna Lena Horns[2], Markus Weidner[2], René Bachmann[2]

[1] University of Hamburg, Hamburg School of Food Science

[2] Landeslabor Schleswig-Holstein, 22437 Neumünster Germany

Corresponding author: Dr. René Bachmann, Landeslabor Schleswig-Holstein, Max-Eyth-Straße 5, 24537 Neumünster, Germany E-Mail: rene.bachmann@lsh.landsh.de

ABSTRACT

The determination of fish species is an important task for food control, especially in the case of very similar and processed fish or fish products. Typically, this analysis is carried out by DNA-based methods. However, especially for small sample series the effort for sample preparation and the required analysis time are not satisfactory. We show that by means of a simple extraction and analysis of the ^{1}H NMR spectra of these extracts a fast discrimination of fish species is possible. For this purpose, 69 samples consisting of five sample groups iridescent shark, European plaice, rock sole, witch flounder and common sole were analysed. The large differences in the ^{1}H spectra could be further highlighted with a principal component and led already to a graphical differentiation of most sample groups. The performed classification of the samples using a subspace discriminant algorithm showed a recognition rate of 100% for all samples, regardless of additives like citric acid or lysine. These results show that ^{1}H NMR spectroscopy is a robust, simple and rapid method for the recognition of the fish species studied.

Keywords: Fish differentiation, ^{1}H NMR spectroscopy, common sole, metabolomics, arsenobetaine

INTRODUCTION

Changing environmental conditions and growing demand for fish are having a major impact on the food chain. In the context of climate change fishing grounds and catch quantities are changing fundamentally.[1,2] In addition, the global demand for fish is constantly increasing due to the increasing wealth and growing of the global population.[3,4] Continued overfishing leads to a reduction in catches of many fish species and to the commercial or unintended marketing of new fish species.[5,6] In addition to fundamental ecological problems, however, this also increases the potential for adulteration and food fraud. In the case of seafood in particular, up to 40% of sold products are considered to be falsely declared.[7] Common sole (S*olea solea*) is considered to be very tasty and expensive fish. Their distribution and fishing area extends across the North Sea, the Atlantic and the Mediterranean. Common soles are mainly caught as bycatch when targeting other fish, such as plaice (*Pleuronectes platessa*).[8] Only rarely succeeds the breeding of sole in breeding facilities.[9] Due to the regulated catch limits, the described popularity as an edible fish and the low catch quantities, the common sole achieves very high prices. Thus, high profits can be made by falsely declaring another, less expensive species of fish as common sole. In particular, processed sole is difficult to distinguish from other flatfishes such as European plaice or witch flounder (*Glyptocephalus cynoglossus*). Due to the very similar appearance, however, fillets of iridescent sharks or cod have also been declared as sole.[10,11] The detection of the different fish species is normally done by PCR.[12,13] Furthermore, there are also NGS or proteome based methods.[14–16] NMR spectroscopy has the advantage of a fast and simple measurement methodology which, in contrast to other spectroscopic methods, offers molecular resolution. Typical applications in food analysis are the detection of contaminants, the authentication of the geographical origin or the identification of admixtures of other foods.[17–19] High recognition rates have already been described for the differentiation of meat species.[20,21] We demonstrate the benefit of ^{1}H NMR spectroscopy for fish species discrimination using five different sample groups and prove the robustness of this method in

the presence of typical additives such as citrate, even when their signals stay are included in the data analysis.

2 MATERIALS AND METHODS

2.1 Reagents and chemicals

Deuteriumoxide (99.9%) was purchased from Eurisotop (Saint-Aubin Cedex, France) and Chloroform-*d* (99.8%) was purchased from Deutero (Castellaun, Germany). Sodium azide (99.5%), potassium phosphate monobasic anhydrous (>99%), potassium phosphate dibasic anhydrous (>98%) from VWR (VWR International GmbH, Darmstadt, Germany).

2.2 Samples

In total 69 fish samples were analysed. Of these, 13 samples were declared as European plaice (*Pleuronectes platessa*). 11 samples were grouped into the rock sole group, with the declaration of the samples differing between *Lepidopsetta bilineata* (4 samples) and *Lepidopsetta polyxystra* (7 samples). Another 20 samples were declared as iridescent shark (*Pangasianodon hypophthalmus*). All these samples were not further analysed for authenticity, as the potential for adulteration was considered low. The other samples consisted of 11 witch flounders (*Glyptocephalus cynoglossus*) and 14 samples of common sole (*Solea solea*). The authenticity of these samples was verified by DNA sequencing. All samples were purchased from different retailers. The samples were very heterogeneous in terms of the form of supply (fresh/frozen) and in the presence of additives. A detailed listing of the samples is given in Table S1. For all samples, the skin and intestines were removed if present and the samples were stored at -20°C until analysis.

2.3 Sample Preparation NMR

Samples for statistical analysis were prepared by adding 1000 µL phosphate buffer (200mM, pH 7.0) to 100 mg sample in a 2.0 ml sample tube (Eppendorf, Germany). The samples were vortexed for 30 seconds and centrifuged for 10 minutes at 18.000 g. 600 µL of the supernatant were taken and transferred to a NMR tube (Deutero, Castellaun, Germany). For acquisition of 2D spectra the procedure were done accordingly, but using 1 g sample.

2.4 NMR acquisition

All spectra were acquired on a Bruker Ascend 400 MHz spectrometer (Bruker Biospin, Rheinstetten, Germany) operating at 400.13 MHz equipped with an Avance III console. The *noesygppr1d* pulse sequence was used for acquisition of water suppressed ^{1}H NMR spectra applying the digitization mode baseopt. The spectra were recorded at 301.8 K, with 16 scans, 65536 complex data points, D1 of 4s, and a spectral width of 8417.5 Hz. The receiver gain (RG) was set to 16 and the transmitter frequency offset was set to 1880.6 Hz. The pulssequence *hsccedetgpsisp2.3* was used for the acquisition of HSQC spectra. The spectra were recorded at 300 K, 32 Scans, 2048 (F2) and 256 (F1) data points, D1 of 1.5 s and a spectral width of 5197.5 (F2) and 16611.3 (F1) Hz. The RG was set to 207 and the transmitter frequency offset was set to 1880.6 (F2) and 7546.0 (F1) Hz. For acquisition of HMBC spectra the pulssequence *hmbcetgpl3nd* with NS=32, 4096 (F2) and 512 (F1) data points, D1 of 1.5 s and a spectral width of 6009.6 (F2) and 22123.9 (F1) Hz and a RG of 207 was used. JRES spectra were acquired using the pulssequence *jresprprqf* with NS=4, 8192 (F2) and 40 (F1) data points, D1 of 2s and a spectral width of 6684.5 (F2) and 12820.6 (F1) Hz and a RG of 16.3. TOCSY spectra were acquired using the pulssequence *mlevphpr.2* with NS=16, 2048 (F2) and 256 (F1) data points, D1 of 2 s and a spectral width of 4000 Hz (F2+F1) and a RG of 207.

2.5 NMR data processing and analysis

For data processing the FIDs were Fourier transformed with a line broadening factor of 0.3 Hz, baseline corrected and phased with Topspin 3.6.2 (Bruker Biospin, Rheinstetten, Germany).

2.6 Statistical analysis of NMR spectra

For multivariate analysis the integrals/buckets were defined using analysis Amix 3.9.15 (Bruker Biospin, Rheinstetten, Germany). The spectra were divided in 168 buckets with a variable width. The region of the water signal from 4.9 to 4.5 was exluded. The final bucket table was exported as csv file. For calculation of the principal component analysis and the following multivariate analysis the software Matlab 2021b (Mathworks) was used. T-test and Pearson correlation were done by using the respective function of Excel 2016 (Microsoft Corporation, Redmond, USA).

RESULTS AND DISCUSSION

The differentiation of fish species is a widely considered topic and is particularly relevant for processed flatfish due to high species similarities. DNA-based approaches are mostly used for analysis. For smaller sample series a faster method with less sample processing would be desirable. We therefore extracted samples from five fish groups European plaice, rock sole, iridescent shark, witch flounder and common sole and analysed the extracts using ^{1}H NMR spectroscopy. It was very surprising for us that the metabolite spectra were already markedly different at first glance. Even though no unambiguous metabolites could be identified for a single fish species, very clear differences in concentration can be seen even. In general, the NMR spectra are dominated by ubiquitously occurring amino acids and a variety of osmolytes. For identification of the metabolites, database searches were made and the results were subsequently checked for plausibility using HSQC, HMBC, TOCSY and JRES spectra. It

should be mentioned, however, that in contrast to the measurements for the statistical analysis, the concentration of the samples was increased by a factor of 10. So some signals can be seen more clearly than in the measurements for fish species differentiation. The identified metabolites including the remarkable values of the normalized concentration are shown in table 1. The differences in the region from 3.0-3.5 ppm are particularly remarkable as figure 1 shows. The evaluations of the 2D spectra indicate that these are almost exclusively signals from trimethylamine derivatives. The comparison of the spectra shows that the witch flounders contain the highest concentrations of trimethylamine-*N*-oxide (TMAO) and betaine. The samples of the rock soles show the second highest TMAO concentration, but almost no betaine. The common soles and European plaices have significant lower concentrations of these osmolytes while creatine concentration is relatively constant across all fish species. The qualitative analysis of the spectra shows additional great diversity of osmolytes. The samples of rock sole exhibit the widest diversity followed by the European plaice. Trimethylamine derivatives thus seem to differ strongly between the different sample groups.

In the next step, Bonferroni corrected t-tests were performed for all sample groups to identify significant buckets. The overview of the significant buckets is shown in table S2. It should be noted that the normalized intensities were used here because many samples had an addition of water. This leads to the fact that for the rock soles almost all buckets had a significantly lower concentration relative to the total integral since the spectra are characterized by a very high TMAO and betaine concentration. With these t-test an additional evaluation of the previously identified buckets could be performed.

Identification of Arsenobetaine

Another interesting metabolite we identified is arsenobetaine. Arsenobetaine (AsB) is reported as major species of arsenic in marine animals.[22] So far the role of AsB it is not finally known. For example, a role as a metabolic end product is assumed, which serves to make inorganic arsenic compounds harmless.[23] On the other hand, an active role as an osmolyte has

also been described.[24] As expected, we could not find any signals of AsB in the freshwater fish samples of iridescent sharks. The absolute intensities show that the rock soles have the highest concentrations, followed by the common sole. However, one of the common sole samples had by far the highest concentration of AsB. These results support the hypothesis that AsB also plays a role as an osmolyte. For a more in-depth analysis, a Pearson correlation with the signal of AsB at 3.367 ppm was performed using the absolute signal intensities. This signal was chosen because the more intense singlet at 1.964 ppm is overlaid by lysine signals in some samples. The correlation of the two AsB signals showed a correlation coefficient of 0.79, which is caused by the mentioned signal overlap. Correlation coefficients above 0.6 are obtained for many signals in the 3.5-3.0 ppm range, where we were able to locate many trimethylamine derivatives. The highest coefficients of 0.71-0.74 are obtained for the signals of betaine, TMAO and choline. Thus, AsB seems to play a role as an osmolyte, although other factors must presumably be taken into account in its formation as shown by the high individual finding in a common sole.

Multivariate Data Analysis

With regard to a time-efficient analysis, an automated and easy to use evaluation is desirable, which is why we also performed a multivariate data analysis. Challenging here is the addition of additives in some of the fish samples, which provides further variance and thus complicates multivariate analysis. However, due to the wide differences in the samples, discrimination is possible even without excluding the citric acid signals. The principal component analysis shows very clearly the differences between the sample groups. Here, all sample groups cluster together and can already be almost completely separated graphically (Figure 2). The largest variance describes principal component (PC) 1, which is also responsible for the greatest proportion in the clustering of the sample groups. Determining metabolites for PC1 are TMAO, betaine as well as an unknown trimethylamine derivate causing a singlet at 3.302 ppm. PC 2 is mainly

influenced by betaine. PC 3 shows additional clustering of samples of rock sole. Samples decelared as *L. bilineata* are seperated from the samples declared as *L. polyxystra*. Dominant metabolites in PC 3 are creatine as well as lactate. Principal component 4 (not shown) shows the increased variance due to the additions of citric acid in eight of the iridescent shark samples. It is noteworthy that in the PCA of the samples, the common sole and European plaice lay very close together, although the two fish species belong to different families. In contrast, rock sole and witch flounder, which also belong to family *Pleuronectidae*, are well separated from these two sample groups. When looking at the preferred habitat, it can be seen that the common sole and European plaice not only inhabit the similar habitats, but also populate similar depths, namely about 10-50m. Rock soles prefer to live at depths of 0-183m and witch flounders even at depths of 45-366m.[25] This matches very well with the graphical clustering, as well as with the trimethylamine derivative dominated region of the NMR spectra from figure 1.

Since PCA is not sufficient for unambiguous discrimination of all fish species, an additional algorithm-based classification was performed. For the classification, a training set was created using 11 European plaice samples, 9 samples of rock soles, 17 iridescent shark samples (with and without added citric acid) and 9 witch flounders and 9 common sole samples. Classification was performed using a subspace linear discriminant algorithm and five-fold cross-validation. The training set was recognized with an accuracy of 100% for each sample group. In the next step, the remaining test set samples, consisting of 2 European plaice samples, 2 rock soles, 3 iridescent shark, 2 witch flounders and 5 common sole samples (including 2 fresh fish and 2 frozen samples without additives and one frozen sample with additives) were assigned by the model created. Here, despite the heterogeneous test set, a recognition rate of 100% was also shown for all samples. These results show that the selected algorithm reliably recognizes the different fish species, regardless of the common additives and the type of the supply.

CONCLUSION

The results of this study show that even with a small sample quantity, a robust classification of the five fish groups investigated can be performed in a short measurement time. For the analysis of fish samples, the described method can provide a major time advantage, as the same measurement can be used to investigate various parameters such as the presence of additives and the authenticity of the sample as well. The metabolites relevant for the graphical differentiation of the sample groups were mainly osmolytes, so further differentiation of other marine fish species seems promising. The high accuracies achieved clearly demonstrate the potential of ^{1}H NMR spectroscopy as a screening method or in addition to classical methods for the authentication of fish species.

TABLES

Table 1: List of identified metabolites in the analysed fish extracts. Only distinguishable signals are listed.

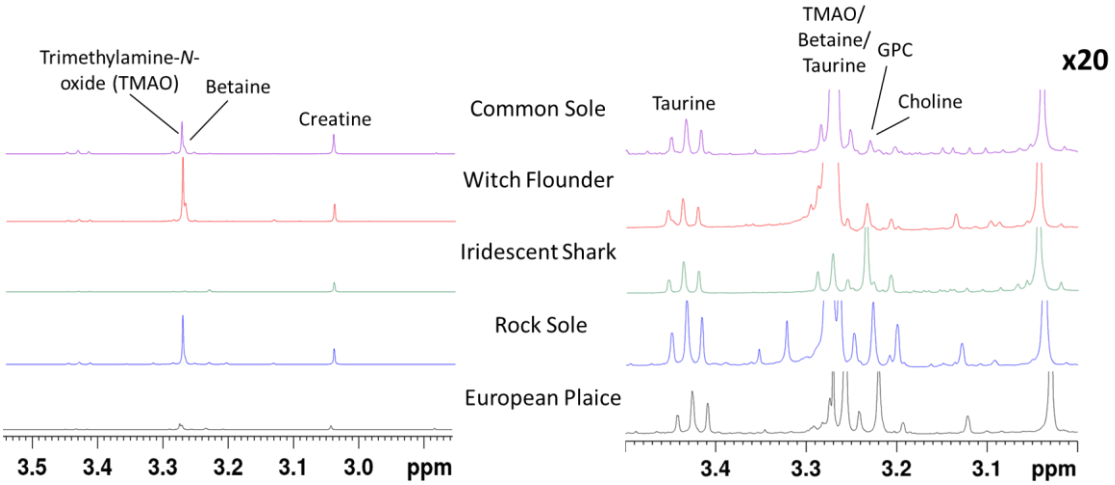
Metabolite	**Signals [ppm]**	**Significant for**	**Comments**
Acetate	1.9253, s	*P. hypophthalmus (high)* *G. cynoglossus (low)*	
Adenine	8.217, s 8.195, s	*P. platessa (high)* *P. hypophthalmus (low)*	
Alanine	1.489, d (7.3 Hz)	*P. hypophthalmus (low)*	
Arsenobetaine	3.367, s 1.964, s	*Lepidopsetta (high)* *P. hypophthalmus (low)*	
Betaine	3.901, s 3.266, s	*G. cynoglossus (high)*	Very high concentration in *G. cynoglossus*
Creatine	3.937, s 3.044, s	*P. hypophthalmus (high)* *G. cynoglossus (low)*	
Cystathionine	2.247- 2.131, m 2.766- 2.705, m 3.158, dd (4.6, 15.0 Hz) 3.093, dd (7.1, 15.0 Hz)	*G. cynoglossus (low)*	Not visible in all spectra

Dimethylamine	2.725, s	-	Not visible in all spectra
Ethanol	1.190, t (7.1 Hz)	-	Not visible in all spectra
Formiate	8.448, s	-	Not visible in all spectra
Fumerate	6.513, s	-	Not visible in all spectra
Glucose	5.246, d (3.7 Hz) 4.657, d (7.8 Hz)	*G. cynoglossus (low)*	
Glycerol	3.667, dd (4.3, 12.0 Hz) 3.571, dd, (6.5, 12.0 Hz)	*G. cynoglossus (low)*	
Glycerophosphocholine	3.234, s 3.700-3.624, m	*Lepidopsetta (high)* *P. hypophthalmus (low)*	
Glycine	3.572, s	*P. platessa (high)* *P. hypophthalmus (low)*	
Histidine	7.136, d (1.0 Hz) 7.989, d (1.0 Hz)	*P. platessa (high)*	
Inosine	8.350, s 8.239, s 6,101, d (5.7 Hz) 4.448, dd (4.0, 5.3 Hz) 4.350-4.316, m	*P. hypophthalmus (high)* *G. cynoglossus (low)*	
Inosinemonophosphate	8.563, s 8.235, s 6.149, d (5.7 Hz) 4.525, dd (3.7, 5.3 Hz) 4.399- 4.365, m 4.071- 4.026, m	*P. platessa (low)* *P. hypophthalmus (high)*	
Isoleucine	1.001, d (6.4 Hz)	*Lepidopsetta (high)* *G. cynoglossus (low)*	
Lactate	4.124, q (6.8 Hz) 1.337, d (6.9Hz)	*P. hypophthalmus (high)* *Lepidopsetta (low)*	
Leucine	1.775-1.628, m 0.959, d (4.4Hz) 0.943, d (4.4Hz)	*G. cynoglossus (low)*	
Lysine	3.061- 3.011, m 1.960- 1.869, m 1.785- 1.694, m 1.573- 1.400, m	*P. hypophthalmus (high)* *G. cynoglossus (low)*	
Methionine	2.145, s	*G. cynoglossus (low)*	
Phenylalanine	7.466- 7.304, m	*G. cynoglossus (low)*	Not visible in all spectra

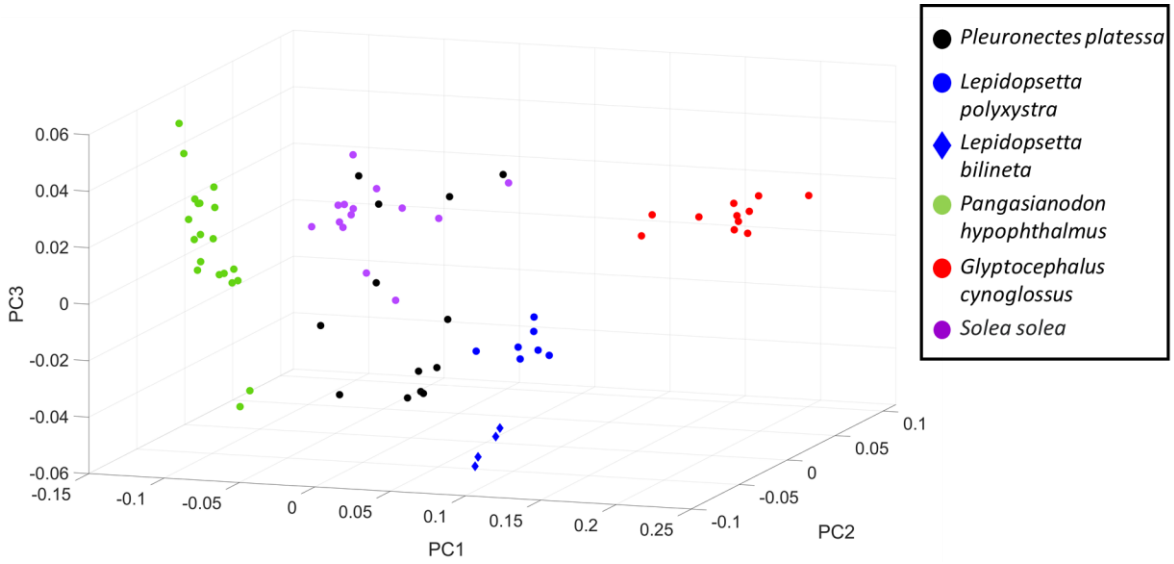
Taurine	3.433, t (6.6 Hz) 3.272, t (6.6 Hz)	*Lepidopsetta (high)* *P. hypophthalmus (low)*	
Trimethylamine	2.890, s	-	
Threonine	4.301- 4.247, m 3.602, d (4.9 Hz)	*S. solea (high)* *G. cynoglossus (low)*	
Tyrosine	7.199, d (8.4 Hz) 6.904, d (8.4 Hz)	*P. hypophthalmus (high)* *G. cynoglossus (low)*	
Valine	1.034, d (7.0 Hz) 0.982, d (7.1 Hz)	*P. hypophthalmus (high)* *G. cynoglossus (low)*	

218

219 FIGURES



220



221

222

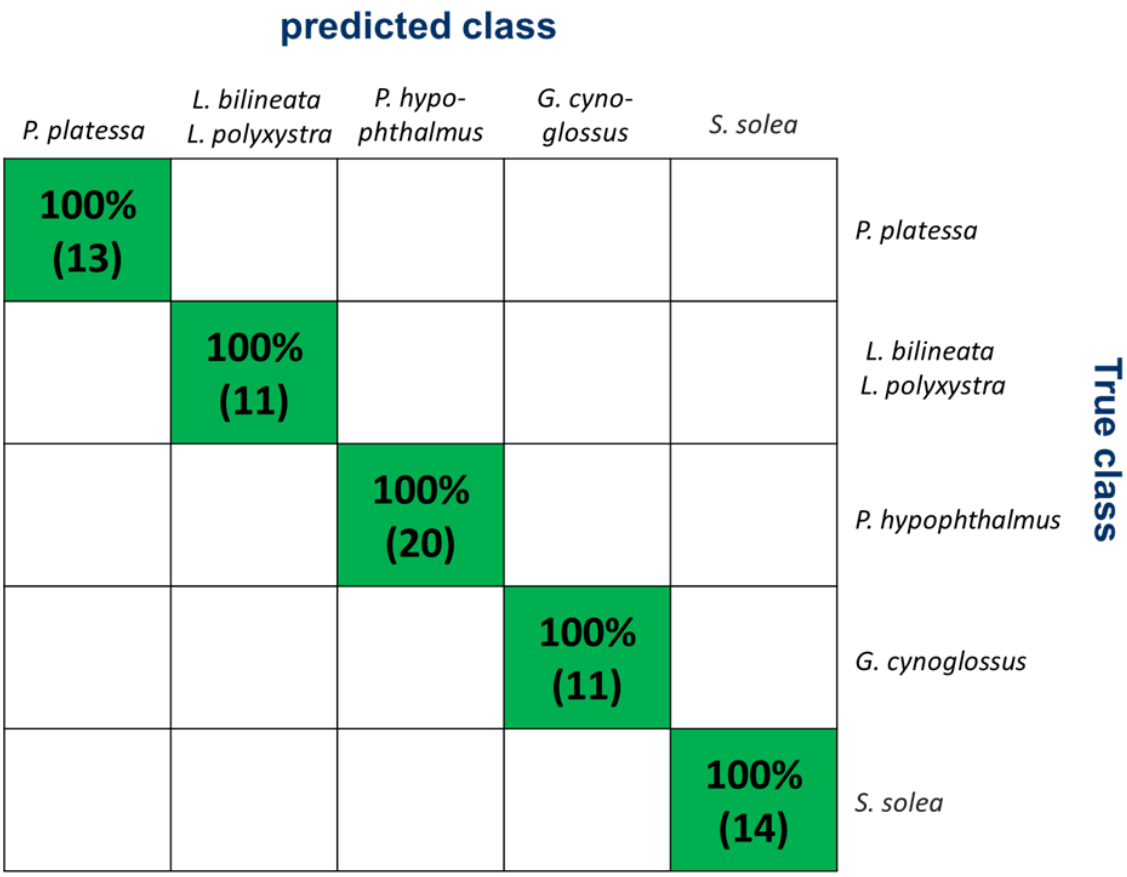


FIGURE CAPTIONS

Figure 1: Excerpts from NMR spectra of the polar extracts of the five analysed sample groups.

Figure 2: PCA Plot of all analysed samples. From right to left is a clustering of the samples that correlates with the respective depth of the habitat. Iridescent Shark as a freshwater fish is on the far left, while common sole and European plaice inhabit depths from 10 to about 60m. Rock soles and witch flounders live at much greater depths and are clearly delineated from these.

Figure 3: Confusion matrix of the combined trainings and test set results.

DECLARATION OF INTEREST

The authors declare that there is no conflict of interest.

ACKNOWLEDGEMENT

The authors thank Ole Sindt and Dr. Hubert Wonnemann from the State Laboratory of Schleswig-Holstein for DNA-sequencing of the witch flounder and common sole samples and their support.

1. Frainer, A. *et al.* Climate-driven changes in functional biogeography of Arctic marine fish communities. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 12202–12207 (2017).

2. Palomares, M. L. D. *et al.* Fishery biomass trends of exploited fish populations in marine ecoregions, climatic zones and ocean basins. *Estuar. Coast. Shelf Sci.* **243**, 106896 (2020).

3. Statista. Global fish production from 2002 to 2022. https://www.statista.com/statistics/264577/total-world-fish-production-since-2002/ (2023).

4. Han, K., Leem, K., Choi, Y. R. & Chung, K. What drives a country's fish consumption? Market growth phase and the causal relations among fish consumption, production and income growth. *Fish. Res.* **254**, 106435 (2022).

5. Johnson, J. W. & Wilmer, J. W. Epinephelus fuscomarginatus (Perciformes: Epinephelidae), a new species of grouper from off the Great Barrier Reef, Australia. *Zootaxa* **4674**, 329–348–329–348 (2019).

6. Froese, R., Kesner-Reyes, K. Impact of Fishing on the Abundance of Marine Species (2002).

7. Leahy, S. Revealed: seafood fraud happening on a vast global scale. *The Guardian* (2021).

8. The International Council for the Exploration of the Sea. *ICES FishMap species factsheet-solea*. https://www.ices.dk/about-ICES/projects/EU-RFP/EU Repository/ICES FIshMap/ICES FishMap species factsheet-sole.pdf (2014).

9. Imsland, A. K. *et al.* A review of the culture potential of Solea solea and S. senegalensis. *Rev. Fish Biol. Fish.* **13**, 379–407 (2004).

10. Holtbecker. *Seezunge, Tropenzunge oder Pangasius? Untersuchungen am CVUA Karlsruhe weisen Verbrauchertäuschung nach*. https://www.ua-bw.de/pub/beitrag.asp?subid=0&Thema_ID=2&ID=1153&Pdf=No (2006).

11. Deconinck, D. *et al.* A high-quality genetic reference database for European commercial fishes reveals substitution fraud of processed Atlantic cod (Gadus morhua) and common sole (Solea solea) at different steps in the Belgian supply chain. *Food Chem. Toxicol.* **141**, 111417 (2020).

12. Asensio Gil, L. PCR-based methods for fish and fishery products authentication. *Trends Food Sci. Technol.* **18**, 558–566 (2007).

13. Herrero, B., Lago, F. C., Vieites, J. M. & Espiñeira, M. Real-time PCR method applied to seafood products for authentication of European sole (Solea solea) and differentiation of common substitute species. *http://dx.doi.org/10.1080/19440049.2011.623682* **29**, 12–18 (2011).

14. Ceruso, M., Mascolo, C., Anastasio, A., Pepe, T. & Sordino, P. Frauds and fish species authentication: Study of the complete mitochondrial genome of some Sparidae to provide specific barcode markers. *Food Control* **103**, 36–47 (2019).

15. Kappel, K., Haase, I., Käppel, C., Sotelo, C. G. & Schröder, U. Species identification in mixed tuna samples with next-generation sequencing targeting two short cytochrome b gene fragments. *Food Chem.* **234**, 212–219 (2017).

16. Mazzeo, M. F. & Siciliano, R. A. Proteomics for the authentication of fish species. *J. Proteomics* **147**, 119–124 (2016).

17. Schmitt, C., Schneider, T., Rumask, L., Fischer, M. & Hackl, T. Food Profiling: Determination of the Geographical Origin of Walnuts by 1H NMR Spectroscopy Using the Polar Extract. *J. Agric. Food Chem.* **68**, 15526–15534 (2020).

18. Lachenmeier, D. W. *et al.* NMR-spectroscopy for nontargeted screening and simultaneous quantification of health-relevant compounds in foods: The example of melamine. *J. Agric. Food Chem.* **57**, 7194–7199 (2009).

19. Bachmann, R. *et al.* Minor metabolites as chemical marker for the differentiation of cane, beet and coconut blossom sugar. From profiling towards identification of adulterations. *Food Control* **135**, 108832 (2022).

20. Decker, C., Krapf, R., Kuballa, T. & Bunzel, M. Nontargeted Analysis of Lipid Extracts Using 1H NMR Spectroscopy Combined with Multivariate Statistical Analysis to Discriminate between the Animal Species of Raw and Processed Meat. *J. Agric. Food Chem.* **70**, 7230–7239 (2022).

21. Decker, C., Krapf, R., Kuballa, T. & Bunzel, M. Differentiation of meat species of raw and processed meat based on polar metabolites using 1H NMR spectroscopy combined with multivariate data analysis. *Front. Nutr*. 9, 2293 (2022).

22. Grinham A, Kvennefors C and Fisher PL et al. Baseline arsenic levels in marine and terrestrial resources from a pristine environment: Isabel Island, Solomon Islands. *Mar Pollut Bull*; **88**: 354–360 (2014)

23. Luvonga C, Rimmer CA, Yu LL, Lee SB. Organoarsenicals in Seafood: Occurrence, Dietary Exposure, Toxicity, and Risk Assessment Considerations - A Review. *J Agric Food Chem.*, **68**, 943-960 (2020).

24. Hoffmann T, *et al*. Arsenobetaine: an ecophysiologically important organoarsenical confers cytoprotection against osmotic stress and growth temperature extremes. *Environ Microbiol.*, **20**, 305-323 (2018).

25. Froese, R. and D. P. E. FishBase. www.fishbase.org (2023).