



Detection of adulterated meat products by a next-generation sequencing-based metabarcoding analysis within the framework of the operation OPSON X: a cooperative project of the German National Reference Centre for Authentic Food (NRZ-Authent) and the competent German food control authorities

Kristina Kappel¹ · Andreas Gadelmeier² · Grégoire Denay³ · Lars Gerdes⁴ · Andrea Graff⁵ · Margit Hagen⁶ · Melanie Hassel⁷ · Ingrid Huber⁴ · Gabriele Näumann⁵ · Melanie Pavlovic⁴ · Klaus Pietsch⁸ · Barbara Stumme⁵ · Inger Völkel⁹ · Simone Westerdorf¹⁰ · Anne Wöhlke¹¹ · Rupert Hochegger¹² · Erik Brinks¹³ · Charles Franz¹³ · Ilka Haase²

Received: 1 February 2023 / Revised: 1 February 2023 / Accepted: 5 April 2023
© The Author(s) 2023

Abstract

The German National Reference Centre for Authentic Food (NRZ-Authent) and the competent German food control authorities of the federal states cooperated within the framework of the 10th joint Europol INTERPOL operation OPSON (OPSON X) in the detection of adulterated meat products. A total of 63 meat product samples were collected and analysed by the authorities using standard analytical procedures and subjected to a recently published 16S rDNA metabarcoding analysis. The sequence reads were analysed using 3 bioinformatics data processing strategies. The study aimed to gain additional data on the test samples regarding the authenticity of the declared species and to validate the 16S rDNA metabarcoding method with representative samples. The method was tested not only on 63 test samples, but also on 5 commercial samples from 2 interlaboratory comparison studies and 9 mock mixtures in parallel. The 16S rDNA metabarcoding method was able to detect species that were not target species of the used standard analytical methods, but failed, as shown previously, to detect fallow deer. Otherwise, the qualitative results of the 16S rDNA metabarcoding method were very similar to those of the methods currently in use by the German food control laboratories. Thus, the method has great potential to be used as a screening method for the authentication of mammal and poultry species in meat products.

Keywords Food authenticity · Meat products · Metabarcoding · Next-generation sequencing · Operation OPSON

1 Introduction

The detection of adulterated and counterfeit food became a focus in many countries even before European competent food control authorities were challenged by the horse meat scandal and its associated media attention in 2013. Since

2011, Europol and INTERPOL jointly coordinate concerted operations called OPSON with many participating countries, in order to detect counterfeit and substandard food and beverages on the markets, and to dismantle the involved organized crime groups.¹ OPSON means “food” or refers to “the value-giving component of the food” in ancient Greek. In the first operation, 10 mainly European countries participated. Since then, the operations have grown and operation OPSON X (spanning from 2020 to 2021) involved 72 countries from all over the world. Besides uncovering food adulteration, the joint OPSON actions aim to strengthen inter-agency cooperations between food control, consumer

Grégoire Denay, Lars Gerdes, Andrea Graff, Margit Hagen, Melanie Hassel, Ingrid Huber, Gabriele Näumann, Melanie Pavlovic, Klaus Pietsch, Barbara Stumme, Inger Völkel, Simone Westerdorf and Anne Wöhlke contributed equally to this work and are listed in alphabetical order.

✉ Kristina Kappel
kristina.kappel@mri.bund.de

Extended author information available on the last page of the article

¹ <https://www.interpol.int/Crimes/Illicit-goods/Food-crime-operations> Accessed 4.4.2023.

protection authorities, law enforcement, and customs, both at the national and international levels. Germany participated in operation OPSON X within a national priority action on meat products with regard to the declared species.² Many processed meat products are nearly impossible to be differentiated by the naked eye, taste or smell with regard to the processed animal species. Because of different prices and availabilities of meat from different species, there is a significant potential for fraud, such as substitution of meat from an expensive species with a cheaper one, or mixing in meat from a non-declared species, in order to increase profit. In the top 10 list of food categories with potential suspicion of fraud, meat and meat products (other than poultry) are number 3, and poultry meat and poultry meat products are number 7, together with fats and oils in the EU (European Commission 2021).

Next-generation sequencing (NGS) methods are progressively replacing or complementing traditional protein and DNA-based analytical methods for the authentication of food products regarding the authenticity of the declared species as they compensate for some of their shortcomings (Haynes et al. 2019). A variety of NGS methods for species identification in meat products have been described. For example, DNA metabarcoding is based on the PCR amplification of genetic markers, which is followed by massively parallel sequencing of the amplicons (Dobrovolny et al. 2019; Liu et al. 2021; Mahama et al. 2022; Pan et al. 2020; Xing et al. 2019). In addition, PCR-free, shotgun-sequencing methods have been developed for the analysis of food products (Akbar et al. 2021; Hellmann et al. 2020; Jiang et al. 2022; Liu et al. 2017; Ripp et al. 2014). Despite the promising approach of NGS methods for meat product authentication and their application by individual competent food control laboratories, these methods still need to be sophisticatedly validated and standardised to become reliable and easily implementable tools for a wider range of competent food control authorities.

Preckel et al. (2021) were first to validate a published DNA metabarcoding method based on short 16S rRNA gene (16S rDNA) targets (Dobrovolny et al. 2019) for the identification of species in mammal and poultry meat products. The applicability of the method was demonstrated by analysing 25 reference samples, 56 food products and 23 pet food products and by comparing the results with those obtained with commercial DNA chips and/or multiplex real-time PCR. Thus, it could be shown that the method is both robust and reproducible. Only recently, the method was validated through an interlaboratory ring trial with 15 participating

laboratories (Dobrovolny et al. 2022), demonstrating that the method can be applied to different sequencing platforms and is suitable for implementation in routine food control analyses. However, for standardization by national bodies or European or International standardisation committees such as CEN or ISO, more information on the consistency of the results with the results of the standard analytical methods used and the robustness of the different analytical steps is crucial.

Here, the cooperation of the German National Reference Centre for Authentic Food (NRZ-Authent) with the German competent food control authorities regarding the analysis of meat products and declared species is described, in the framework of a laboratory cooperation within operation OPSON X. Sixty-three meat product samples collected by the authorities, analysed by using various standard analytical methods, and 5 commercially available proficiency test samples, 9 known mock mixtures, and 6 negative controls of 5 different types were subjected to the 16S rDNA metabarcoding protocol published by Dobrovolny et al. (2019), and analysed by 3 different bioinformatics data processing strategies.

The aim of this cooperation was not only to provide additional information regarding the species of the meat product samples for the assessment of authenticity, but also to further validate the 16S rDNA metabarcoding method regarding the comparability with diverse standard analytical methods used by the competent authorities. Moreover, the study aimed to gain information about the source of unspecific reads commonly observed in negative controls and mock mixtures, as well as about the robustness of the bioinformatic evaluation of the generated DNA sequences.

2 Materials and methods

2.1 Sample material

Monitoring samples were collected from the market or at producing companies within operation OPSON X, and analysed for authenticity of the declared species using various standard analytical methods. Aliquots of extracted DNA from 61 samples collected by 7 agencies from 4 Federal States were transferred to the NRZ-Authent for 16S rDNA metabarcoding. DNA of 2 additional meat product monitoring samples were provided by an 8th agency. The meat product samples comprised various product types from minced meat to goulash soup (Fig. 1). Furthermore, 8 DNA extracts from 5 samples produced for an interlaboratory comparison study in 2019 (LVU Lippold, Herbolzheim, Germany) and in 2021 (LVU Lippold) were provided by 2 agencies (DNA extracts from 3 of these samples were individually prepared in both laboratories). Table 1 and Online Resource

² https://www.bvl.bund.de/DE/Arbeitsbereiche/01_Lebensmittel/03_Verbraucher/16_Food_Fraud/06_OPSON_Operationen/OPSON-X/OPSON_Operationen_node.html Accessed 4.4.2023.

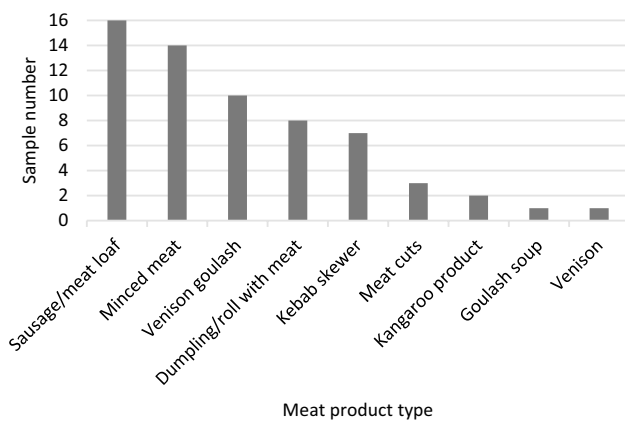


Fig. 1 Number of meat product samples of different product types taken by food control authorities within OPSON X (n=60) and 2 official control samples taken outside OPSON X (one kebab skewer sample and the only goulash soup sample) that were analysed with the 16S rDNA metabarcoding method in this study, in addition to conventional reference methods

1 (Supplementary Information) show the standard analytical methods used by the food control authorities for the analysis of the samples. DNA concentrations of all provided DNA extracts were measured fluorometrically at the NRZ-Authent with the DeNovix dsDNA High Sensitivity Kit (DeNovix, Wilmington, DE, USA) on a DS-11FX µl volume spectrophotometer/fluorometer (DeNovix, Wilmington, DE, USA). Aliquots of DNA extracts were prepared with concentrations of 10 ng/µl, samples with lower concentrations were left unchanged.

Additionally, 3 types of mock mixtures with defined species shares were prepared:

- 2 mock mixtures with animal species shares based on mitochondrial genome copies were provided by the Austrian competent food control authority
- 6 further mock mixtures were prepared at the NRZ-Authent as mixtures of total DNA of individual species
- 1 additional mock mixture sample was prepared as a mixture (volume/volume) from commercial goat's and cow's milk previously authenticated by real-time PCR

Aliquots of all mock mixtures were prepared with concentrations from 5 to 10 ng/µl.

2.2 PCR and sequencing reaction

16S rDNA metabarcoding analysis was performed according to Dobrovolsky et al. (2019) with minor changes. Briefly, an approximately 120 bp region of the 16S rDNA was amplified in all samples (with 13 samples analysed in duplicates) by PCR ('amplicon PCR'). Primers for mammal and poultry 16S rDNA were synthesized with adapters for subsequent

index PCRs and sequencing (in italics) in "nGS scale" by metabion international AG (Planegg, Germany), and sequences were as follows:

Fwd_Ma_{OH} 5'-TCGTCCGGCAGCGTCAGATGTGTATAAGAGACAGGACGAGAAGACCCTATGGAGC-3', Rev_Ma_{OH} 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCGAGGTCACCCCAACC-3', Fwd_P_{OH} 5'-TCGTCCGAGCGTCAGATGTGTATAAGAGACAGGACGAGAAGACCCTGTGGAAC-3', Rev_P_{OH} 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAAGGTCGCCCAACC-3'.

The amplicon PCR reaction mix (25 µl) contained 12.5 µl 2× HotStarTaq Master Mix (Qiagen, Hilden, Germany), 3 µl 25 mM MgCl₂, 1.75 µl of each primer (2.14 µM) and 2.5 µl sample DNA. In addition, 2 amplicon PCR negative controls were prepared with molecular biology grade water instead of sample DNA. PCR was conducted in a GeneExplorer 96 thermocycler (Bioer Technology, Hangzhou, China) with the following cycling protocol: 15 min initial denaturation at 95 °C, 25 cycles of 30 s denaturation at 95 °C, 30 s annealing at 62 °C and 30 s elongation at 72 °C, and final elongation of 10 min at 72 °C. Aliquots of 3 µl of each PCR were checked for successful amplification with agarose gel electrophoresis on 2% agarose gels. 20 µl of each PCR product were purified using 90 µl Mag-Bind TotalPure NGS magnetic beads (Omega Bio-tek, Norcross, GA, USA). The amplicons were incubated with magnetic beads in deepwell plates for 2 min with shaking and additional 5 min without shaking and the supernatant was discarded. The amplicons were washed twice with 200 µl ethanol (80%) and the ethanol was removed carefully. The amplicons were resuspended in 53 µl 10 mM Tris-HCl (pH 8.5) and incubated for 2 min with shaking and 2 min without shaking. Finally, the plate was incubated for 2 min on the magnetic adapter and the purified DNA was transferred to a new plate.

The purified amplicons were used for a second PCR ('index PCR') to index the amplicons from the different samples and to introduce adapters for binding the DNA to the sequencing flow cell. 5 µl of purified amplicon DNA were mixed with 5 µl each of two indices from the Nextera XT Index Kit v2 Set B (Illumina, San Diego, CA, USA), 25 µl 2× HotStarTaq Master Mix (Qiagen, Hilden, Germany) and 10 µl molecular biology grade water (total volume of 50 µl). Two index PCR negative controls were prepared with molecular grade water instead of amplicon DNA. The index PCR cycling protocol started with an initial denaturation step (15 min at 95 °C) that was followed by 8 cycles of denaturation (30 s at 95 °C), primer annealing (30 s at 55 °C) and elongation (30 s at 72 °C) and the protocol ended with a final elongation step (10 min at 72 °C). The index amplicons were purified with magnetic beads as described above, but with the following changes: 50 µl of PCR reactions were used, the washed PCR products were resuspended in 28 µl Tris buffer

Table 1 Analytical methods applied by the competent food control authorities for the authentication of food control meat products and proficiency test samples

Type of method	Product/literature	Target species
PCR-RFLP	BVL L.06.26/27-2 (2007–12) (Federal Office of Consumer Protection and Food Safety 2007)	Horse
PCR-RFLP	Meyer et al. (1995)	
PCR-RFLP	Wolf et al. (1999)	
Singleplex real-time PCR	In-house method	Cattle
Singleplex real-time PCR	In-house method	Chicken
Singleplex real-time PCR	SureFood ANIMAL ID Beef IAAC, CONGEN Biotechnologie, Berlin, Germany	Cattle
Singleplex real-time PCR	SureFood ANIMAL ID Beef IAAC, CONGEN Biotechnologie, Berlin, Germany	Pig
Multiplex real-time PCR	BVL L08.00-61 (2016–03) (Federal Office of Consumer Protection and Food Safety 2016a, b)	Cattle, pig, turkey, chicken
Multiplex real-time PCR	BVL L08.00-62 (2016–03) (Federal Office of Consumer Protection and Food Safety 2016a, b)	Cattle, pig, sheep, equids
Multiplex real-time PCR	First-Animal Tetra I, GEN-IAL GmbH, Troisdorf, Germany	Pig, cattle, chicken, turkey
Multiplex real-time PCR	SureFood ANIMAL ID 4plex Beef/Sheep/Goat + IAAC, CONGEN Biotechnologie, Berlin, Germany	Cattle, sheep, goat
DNA chip	MEAT 4.0 LCD-Array Kit, Chipron GmbH, Berlin, Germany	Cattle, bison, pig, sheep, goat, buffalo, horse, donkey, hare, rabbit, kangaroo, roe deer, red deer, axis deer, fallow deer, reindeer, springbok, chicken, turkey, goose, Mallard Duck, Muscovy duck, pheasant, ostrich
DNA chip	MEAT 5.0 LCD-Array Kit, Chipron GmbH, Berlin, Germany	Cattle (<i>Bos taurus/Bos bison</i>), sheep, equine, goat, camels, buffalo, pig, kangaroo, hare, rabbit, reindeer, roe deer, red deer, fallow deer, springbok, canine/dog, cat, chicken, turkey, goose, ostrich, Mallard duck, Muscovy duck, pheasant
DNA chip	MEAT Plus 3.0 LCD-Array Kit, Chipron GmbH, Berlin, Germany	Red fox, racoon dog, racoon, muskrat, bear, brown rat, black rat, mouse, wapiti, European elk, kudu, llama, guinea fowl, Japanese quail, pigeon
Sanger sequencing	DIN CEN/TS 17.303:2019-06 (DIN 2019)	Originally described for fish species in general
ELISA	ELISA-TEK Cooked 4 Species Kit, ELISA Technologies, Gainesville, FL, USA	Cattle, pig, poultry (chicken and turkey), sheep
ELISA	ELISA-TEK Cooked Meat Beef Kit, ELISA Technologies, Gainesville, FL, USA	Cattle
ELISA	ELISA-TEK Cooked Meat Horse Kit, ELISA Technologies, Gainesville, FL, USA	Horse
ELISA	ELISA-TEK Cooked Meat Pork Kit, ELISA Technologies, Gainesville, FL, USA	Pig
ELISA	ELISA-TEK Cooked Meat Poultry Kit, ELISA Technologies, Gainesville, FL, USA	Poultry (chicken and turkey)
ELISA	ELISA-TEK Cooked Meat Sheep Kit, ELISA Technologies, Gainesville, FL, USA	Sheep

and 25 µl of supernatants were transferred to a new plate. The success of the index PCR amplification was checked with electrophoresis on 2% agarose gels, and amplicons were quantified using the Qubit dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA, USA) and a Spark multimode microplate reader (Tecan Group, Männedorf, Switzerland).

Each amplicon was diluted with 10 mM Tris–HCl (pH 8.5) to a concentration of 4 nM and then equimolarly pooled. One of the 2 amplicon PCR negative controls and 1 of the 2 index PCR negative controls were diluted at a ratio of 1:46, corresponding to the mean dilution factor for all samples. The other 2 PCR negative controls were used undiluted. 5 µl of the pooled library was denatured for 5 min with 5 µl 0.2 M NaOH (pH 8.5) and subsequently diluted with buffer HT1 (from the MiSeq Reagent Kit v2 (300-cycles), Illumina, San Diego, CA, USA) to a concentration of 8 pM. 270 µl of the diluted library were spiked with 30 µl PhiX solution (12.5 pM; Illumina). Finally, the library pool was loaded on a MiSeq Reagent Kit v2 flow cell and 126 bp were sequenced on a MiSeq sequencing instrument (Illumina) in both directions.

2.3 Sequence data processing

Before further data processing, sequencing adapters were removed from the reads. Demultiplexing was achieved with the software on the MiSeq instrument with default settings allowing one mismatch in the index sequences. Further data processing was performed on a locally installed Galaxy platform (Jalili et al. 2020). Three data processing workflows were tested in parallel.

For the ‘OTU clustering workflow’, the sequence reads were processed according to Dobrovolny et al. (2019) with minor changes. In brief, the qualities of the reads were checked with FastQC³ and forward and reverse primer sequences were removed with Cutadapt (Martin 2011). Reads with lengths of less than 50 bp were deleted and bases with quality scores below 15 were trimmed with Trimmomatic (Bolger et al. 2014). Forward and reverse reads were joined with fastq-join (Aronesty 2013) with a maximum of 9% difference between matching segments, and joined sequences were converted to FASTA format with FASTQ to FASTA converter (Blankenberg et al. 2010). Reads were dereplicated and sorted with VSearch (Rognes et al. 2015), and sequences were clustered into de novo Operational Taxonomic Unit (OTU) sequences using a 97% threshold with the USEARCH-Tool Suite (Edgar 2013). OTU representative sequences were used to search a locally installed copy of the NCBI BLAST nucleotide database (version 2019-12-09) using blastn (Camacho et al. 2009; Cock et al. 2015) with

a 97% identity cut-off and the top hit in each case was used for assigning taxonomy.

For setting up the ‘dereplication workflow’, the OTU clustering step was removed from the workflow above. Thus, the dereplicated sequences could be directly checked against the nucleotide database. All other steps were left unchanged.

A ‘dada2 workflow’ was set up using the software package DADA2 according to Callahan et al. (2016). To remove the primers, 21 and 18 bases were truncated at the start of each forward and reverse read, respectively, and all sequences were truncated after 80 bases. After determining the error model of the sequencing data, the denoised forward and reverse reads were merged with a minimum of 12 bases and without allowing mismatches in the overlapping segments. The resulting amplicon sequence variants (ASV) were purified from chimeric sequences and were queried against the same locally installed copy of the NCBI nucleotide database as in the two other workflows.

After the execution of each workflow, the read percentages per species were summed up in all samples. Species with less than 0.5% of total sample reads were considered negative and the remaining read percentages per species were re-adjusted to a total of 100%.

The demultiplexing of the sequencing raw data was repeated with the BCL software (Illumina) installed on a separate computer and with allowing zero mismatches in the index sequences as this was not possible with the software on the instrument. This time, only the reads in the negative control samples and in the mock mixtures were recorded.

3 Results

One of the 63 meat product samples from the competent food control authorities did not yield a PCR product after the amplicon PCR step, despite repeated reactions and therefore, could not be analysed. For the remaining 62 monitoring samples (8 samples measured in duplicates) as well as the 5 boiled sausage proficiency test samples (3 samples measured in duplicates) and the 9 mock mixtures (2 samples measured in duplicates), PCR amplicons were produced. Thus, together with 6 negative controls, a total of 95 samples were analysed in parallel on a flow cell of the MiSeq Reagent Kit v2, and data analysis was performed using 3 different workflows. The compositions of the mock mixtures were known before carrying out the experiment, whereas the compositions of the proficiency test samples were only disclosed after data evaluation. The actual compositions of the monitoring samples remained unknown, but the results of the metabarcoding analysis of these samples, as well as those of the proficiency test samples, were compared with the results of the standard analytical methods used by the competent food control authorities.

³ <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

Table 2 Total reads per sample after sequencing and recovery of reads and total number of different sequences, OTUs or ASVs per sample after different bioinformatic workflows

	Total reads per sample	Read recovery (total number of different sequences, OTUs or ASVs per sample)		
		Dereplication (sequences)	OTU clustering (OTUs)	dada2 (ASVs)
Minimum	82,918	83.7% (1984)	81.1% (270)	92.0% (7)
Maximum	176,736	93.1% (6275)	91.8% (1000)	99.1% (28)
Mean	133,462	90.7% (3428)	88.7% (520)	96.1% (15)
Standard deviation	18,845	2.0% (992.5)	2.3% (143.8)	1.9% (4.8)

The values were calculated for all samples except the negative controls

Table 3 Total reads in negative control samples for demultiplexing on the MiSeq device with max. one mismatch in index sequences and outside the MiSeq device with zero mismatches in index sequences

Negative control type	Sample ID	Demultiplexing on MiSeq device with max. 1 MM in indexes	Demultiplexing outside MiSeq device with 0 MM in indexes (reduction in %)
Amplicon PCR (undiluted)	OPSONX_099	248	55 (77.8)
Amplicon PCR (diluted)	OPSONX_100	171	51 (70.2)
Index PCR (undiluted)	OPSONX_101	191	69 (63.9)
Index PCR (diluted)	OPSONX_102	162	51 (68.5)
Index sequencing (combination 1)	OPSONX_103	199	62 (68.8)
Index sequencing (combination 2)	OPSONX_104	264	59 (77.7)
Mean		205.8	57.8 (71.2)

Table 4 Number of identified false-positive species in mock mixtures and read percentages per false-positive species

Bioinformatic workflow	Number of false-positive species in all nine mock mixtures	False-specific reads/sample	Reads per false-positive species (%)		
			Sum	Mean (%)	Maximum
Dereplication (0 MM)	173	0.37	0.35	0.026	0.0037
Dereplication (1 MM)	196	0.54	0.34	0.032	0.0042
OTU clustering (0 MM)	127	0.36	0.39	0.035	0.0046
OTU clustering (1 MM)	149	0.53	0.37	0.040	0.0051
dada2 (0 MM)	61	0.27	0.33	0.052	0.0241
dada2 (1 MM)	82	0.45	0.32	0.064	0.0233

All false-positive species results were counted without considering a read percentage threshold

3.1 NGS read performance

In total, 16.16 m reads were produced with a Q30 quality score of 90.9%, a mean of 133 462 reads per sample (excluding negative controls) and a mean of 90.7% (dereplication), 88.7% (OTU clustering) and 96.1% (dada2) evaluated reads after executing the different bioinformatic workflows (Table 2). The negative controls yielded 171–264 reads per sample (205.8 reads on average), which could be reduced to 51–69 reads (57.8 reads on average) per sample by allowing zero mismatches in the index sequences during demultiplexing (Table 3). No trend was observed with regard to the type of negative control (undiluted or diluted amplicon PCR, undiluted or diluted index PCR, or simulated index

combinations) and the resulting read numbers. The false-positive results of the mock mixtures with the known compositions were used for determining the read percentage threshold for species detection. The highest read percentages for particular false-positive species were 0.32–0.39% depending on the bioinformatic workflow used (Table 4). Although the total percentage of false-positive reads per sample decreased by demultiplexing with no mismatches in the indices, the maximum percentage of reads per given false-positive species in the mock mixture samples was not reduced. Thus, for the further analysis of the metabarcoding results, a threshold of 0.5% reads was used to rate a species as positive. Duplicate measurements carried out for 13 samples gave very similar results (Fig. 2). However, the

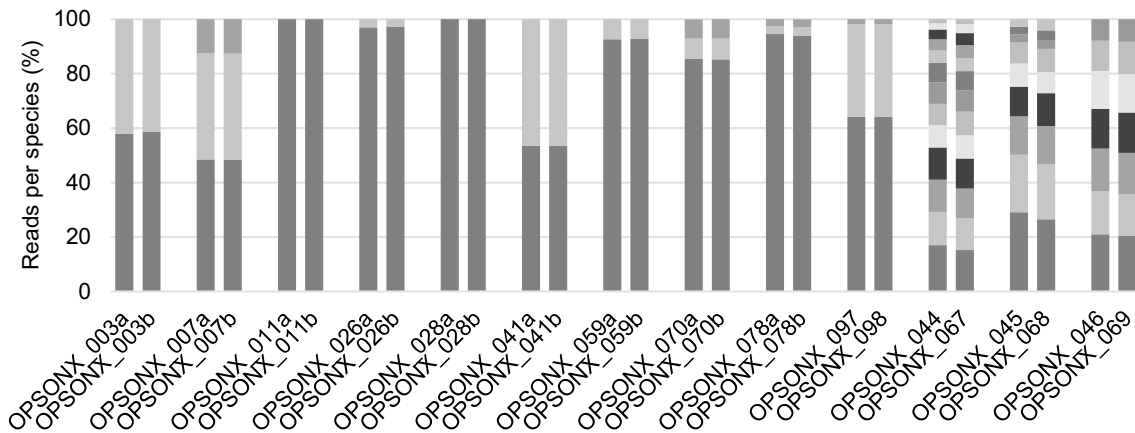


Fig. 2 Read percentages per species in samples measured in duplicates. The first 8 duplicates (pairs of bars) show the results of OPSON X meat samples and one official control sample taken outside of OPSON X (OPSONX_070a/b). Sample OPSONX_078a was a mock mixture with DNA content-based shares of species, whereas samples OPSONX_097/OPSONX_098 were aliquots with species

shares based on mitochondrial genome target copies. For these 8 samples, duplication started with performing 2 separate amplicon PCRs. The last 3 duplicates (boiled sausage samples) were from a commercial interlaboratory ring trial and DNA extractions were performed in 2 different laboratories. The different species are indicated by different greyscales

measured species proportions in the duplicate proficiency test samples, for which the DNA extraction had been carried out in different laboratories and which were much more diverse in species composition, differed somewhat more than the duplicate measurements of the mock mixtures, for which only one DNA extract was used and which differed less in species variety.

3.2 Mock mixtures and proficiency test samples

Nine mock samples were analysed containing 20 different species in total (Table 5) of which 2, Atlantic salmon and desert locust, were not target species for the metabarcoding approach. With applying a threshold of 0.5% reads, no false-positive species were determined. Also, almost all target species could be detected except for fallow deer (*Dama dama*), a finding which has been reported by Preckel et al. (2021) and is attributed to 2 mismatches between the reverse primer for mammals and the 16S rRNA gene of fallow deer (Fig. 3). In addition, read percentages for species that were present in low concentrations in the mock mixtures (0.5% and 0.1%, respectively) were below the chosen read percentage threshold, and the respective species were therefore also considered as negative. Although the 16S rDNA metabarcoding method was originally developed for mammal and poultry species, reads for salmon were also recovered. The forward primer for mammals matches Atlantic salmon DNA without mismatches and the two reverse primers match salmon DNA with 3 mismatches each (Fig. 4). This appears to be sufficient for at least partial amplification of salmon DNA, probably due to the positions of the mismatches in the primer Rev_P_{OH} at and near the primer start. Read percentages

for desert locust, the other non-target species, were below the threshold. This was expected, as the primer sequences could not be found within desert locust DNA sequences from GenBank (data not shown). When the demultiplexing was repeated with allowing zero mismatches in the index sequences, no differences in recovered species above the threshold of 0.5% were found (Online Resource 1, Supplementary Information). However, the recovered percent reads per species do not necessarily reflect the actual share of species in the samples but differ in part considerably (Table 5).

When applying the determined threshold of 0.5% reads to the proficiency test samples, all species were identified in the 5 meat mixtures (boiled sausages; three samples measured as duplicates) from the proficiency tests, with the exception of fallow deer which was present in 4 of the 5 samples. No false-positive species were detected in the samples, assuming possible contaminations with cattle in two proficiency test samples, as was described in the respective interlaboratory study reports (LVU 2019, 2021) (Table 5). However, it should be noted that the applied NGS method could not distinguish between wild boar (*Sus scrofa*) and domesticated pig (*Sus scrofa domesticus/Sus domesticus/Sus domestica*) (which was present in 2 samples).

3.3 Comparison of 16S rDNA metabarcoding with standard methods: food control samples and proficiency test samples

As most of the results provided by the food control laboratories were of qualitative and not quantitative nature, only the qualitative metabarcoding results for the 62 successfully investigated food control samples

Table 5 Composition and read percentages per species in mock mixtures and proficiency test samples

Sample type	Sample ID	Species	Species recovered by NGS	Sample composition (%)	Reads per species after dereplication workflow (%)	Reads per species after OTU clustering workflow (%)	Reads per species after dada2 workflow (%)	
Mock mixture	DNA mixture (proportions of target gene)	OPSONX_096	Horse	Yes	66.1	59.0	58.8	61.4
			Goat	Yes	30	36.9	36.9	34.6
			Chicken	Yes	1.9	2.1	2.2	2.0
			Turkey	Yes	1.9	2.0	2.1	2.0
			Sheep	No	0.1	–	–	–
Mock mixture	DNA mixture (proportions of target gene)	OPSONX_097/ OPSONX_098	Sheep	Yes	67.5	64.3/64.0	64.2/64.1	65.8/65.6
			Cattle	Yes	30	34.0/34.1	34.0/34.0	32.5/32.7
			Goat	Yes	1.9	1.8/1.9	1.8/1.9	1.7/1.7
			Horse	No	0.5	–/–	–/–	–/–
			Pig	No	0.1	–/–	–/–	–/–
Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_078a/ OPSONX_078b	Cattle	Yes	88	94.6/94.1	94.5/94.5	94.8/94.4
			Sheep	Yes	10	2.8/3.0	2.9/2.9	2.8/2.9
			Goat	Yes	2	2.6/2.9	2.6/2.6	2.5/2.7
Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_086	Goat	Yes	20	42.1	42.0	43.5
			Cattle	Yes	20	35.1	35.2	34.0
			Sheep	Yes	20	19.0	19.0	18.7
			Salmon (<i>Salmo salar</i>) ^a	Yes	20	3.8	3.9	3.8
			Desert locust ^b	No	20	–	–	–
Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_087	Springbok	Yes	8.96	23.4	23.4	24.9
			European roe	Yes	8.96	15.7	15.7	15.4
			Common wallaroo	Yes	8.96	12.1	12.1	11.9
			Chital	Yes	8.96	9.8	9.6	9.5
			Red deer	Yes	8.96	9.5	9.6	9.1
			Horse	Yes	8.96	8.4	8.4	8.4
			Siberian roe	Yes	8.96	6.7	6.5	6.4
			Bennet's kangaroo	Yes	8.96	6.3	6.4	6.3
			Red kangaroo	Yes	8.96	2.5	2.6	2.5
			European rabbit	Yes	8.96	2.2	2.2	2.2
			Fallow deer	No	8.96	–	–	–
			Eastern grey kangaroo	Yes	1.49	3.4	3.5	3.4
			Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_088	European roe	Yes	40
European rabbit	Yes	40				8.8	8.9	8.7
springbok	Yes	18				37.6	37.5	36.4
Horse	Yes	2				1.1	1.1	1.1

Table 5 (continued)

Sample type	Sample ID	Species	Species recovered by NGS	Sample composition (%)	Reads per species after dereplication workflow (%)	Reads per species after OTU clustering workflow (%)	Reads per species after dada2 workflow (%)	
Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_089	Red kangaroo	Yes	85	48.8	48.8	51.1
			Eastern grey kangaroo	Yes	5	35.9	35.8	34.3
			Common wallaroo	Yes	5	10.5	10.5	10.0
			Bennet's kangaroo	Yes	2	2.2	2.3	2.1
			European rabbit	Yes	2	0.9	0.9	0.9
Mock mixture	DNA mixture (proportions of DNA content)	OPSONX_090	European hare	Yes	1	1.7	1.7	1.6
			Red deer	Yes	77	84.1	84.1	84.9
			Red kangaroo	Yes	10	4.3	4.4	4.2
			Siberian roe	Yes	5	4.8	4.7	4.4
			Fallow deer	No	5	—	—	—
			Eastern grey kangaroo	Yes	1	3.9	4.0	3.9
			European roe	Yes	1	1.9	1.9	1.7
Mock mixture	Milk mixture	OPSONX_095	Chital	Yes	1	0.9	0.9	0.9
			Goat	Yes	99	96.3	96.0	96.3
			Cattle	Yes	1	2.6	2.6	2.6
Proficiency test sample	Meat mixture (boiled sausage) (w/w)	OPSONX_044/ OPSONX_067	Wild goat ^b			1.1	1.4	1.1
			Chicken	Yes	9.8	4.1/4.7	4.1/4.8	4.1/4.6
			Cattle	Yes	8.5	8.2/8.5	8.2/8.6	8.1/8.3
			Sheep	Yes	7.8	4.4/4.7	4.5/4.8	4.3/4.6
			Red deer	Yes	7.8	17.1/15.4	17.0/15.2	18.5/14.9
			Ostrich	Yes	7.8	7.8/8.5	7.8/8.6	7.6/8.3
			Kangaroo	Yes	7.4	7.6/7.7	7.8/7.9	7.5/7.6
			Rabbit	Yes	7.1	1.5/2.0	1.5/2.0	1.5/2.0
			Wild boar ^c	Yes	6.3	12.0/11.5	12.2/11.7	11.7/13.3
			Horse	Yes	6.3	3.5/4.3	3.6/4.4	3.6/4.3
			Fallow deer	No	6.3	—/—	—/—	—/—
			Roe	Yes	6.3	12.6/11.8	11.8/11.0	12.4/11.6
			Elk	Yes	6.3	7.2/6.9	7.3/7.0	7.0/6.7
			Turkey	Yes	6.3	2.3/3.1	2.3/3.2	2.3/3.0
			Reindeer	Yes	6.1	11.8/10.9	11.9/10.9	11.4/10.6
Proficiency test sample	Meat mixture (boiled sausage) (w/w)	OPSONX_045/ OPSONX_068	Pig	Yes	32.6	10.5/11.7	10.9/12.1	10.2/11.2
			Roe	Yes	14.4	30.1/27.1	29.1/26.5	31.8/29.8
			Kangaroo	Yes	13.9	10.5/11.9	10.7/11.7	10.5/11.7
			Pheasant	Yes	13.1	13.9/13.7	14.1/13.9	13.5/13.3
			Hare	Yes	10.9	8.3/7.7	8.5/7.8	8.4/7.6
			Elk	Yes	10.9	21.4/20.4	21.1/20.3	20.6/19.6
			Horse	Yes	4.3	2.7/4.0	2.8/4.1	2.3/3.3
			Cattle	Yes	Possible contamination ^d	2.7/3.5	2.8/3.6	2.7/3.5
			Fallow deer	No	Possible contamination ^e	—/—	—/—	—/—

Table 5 (continued)

Sample type	Sample ID	Species	Species recovered by NGS	Sample composition (%)	Reads per species after dereplication workflow (%)	Reads per species after OTU clustering workflow (%)	Reads per species after dada2 workflow (%)	
Proficiency test sample	Meat mixture (boiled sausage) (w/w)	OPSONX_046/ OPSON_069	Cattle	Yes	25.0	15.7/15.2	15.9/15.3	15.2/14.6
			Goat	Yes	16.4	7.9/8.3	8.1/8.3	7.8/8.0
			Sheep	Yes	14.3	11.1/11.8	11.1/11.9	10.9/11.5
			Wild boar ^c	Yes	13.9	15.5/15.1	15.6/15.1	15.2/14.6
			Red deer	Yes	12.1	20.8/20.4	21.0/20.5	22.8/23.0
			Muscovy duck	Yes	10.4	14.4/14.5	14.0/14.1	14.3/14.1
			Goose	Yes	7.8	14.4/14.7	14.5/14.8	13.9/14.1
			Fallow deer	No	Possible contamination ^f	–/–	–/–	–/–
Proficiency test sample	Meat mixture (boiled sausage) (w/w)	OPSONX_048	Red deer	Yes	16.0	23.1	23.3	25.5
			Cattle	Yes	15.6	9.2	9.4	8.9
			Ostrich	Yes	15.3	19.5	19.2	18.5
			Hare	Yes	14.4	10.1	10.1	9.9
			Kangaroo	Yes	14.2	13.7	13.3	13.5
			Sheep	Yes	12.6	12.5	12.6	12.2
			Pheasant	Yes	12.0	11.9	12.0	11.4
Proficiency test sample	Meat mixture (boiled sausage) (w/w)	OPSONX_049	Pig	Yes	28.9	10.7	11.0	10.4
			Horse	Yes	17.8	19.6	19.5	18.0
			Muscovy duck	Yes	16.4	21.0	20.6	20.2
			Reindeer	Yes	13.8	41.2	41.4	44.0
			Goat	Yes	12.0	6.3	6.4	6.2
			Fallow deer	No	11.1	–	–	–
			Cattle	Yes	Possible contamination ^g	1.1	1.1	1.1

Species with less than 0.5% reads were considered as negative (“–”). All positive reads of each sample were adjusted to 100%. DNA mixture and milk mixture samples were produced as mock mixtures with defined species compositions and were used for threshold determination. Meat mixtures were boiled sausages produced for commercial proficiency tests that were analysed in this study without prior knowledge of species composition. Two meat mixtures were analysed in duplicates starting from two amplicon PCRs. Two proficiency test samples were analysed in duplicates with DNA extractions performed in different laboratories

^aSalmon and desert locust were introduced as non-target species in mock mixture. Salmon was nevertheless recovered with read percentages above the threshold (hit = salmon environmental sample) whereas desert locust was negative as expected

^bFirst hit of manually performed BLAST search in GenBank showed 100% identity to both wild goat (*Capra aegagrus*) and domestic goat (*Capra hircus*)

^cNGS did not discriminate between wild boar (*Sus scrofa*) and domestic pig (*Sus scrofa domesticus*)

^dContamination of cattle in proficiency test sample is likely as 97% of participants of interlaboratory comparison test identified cattle as positive

^ePossibly contamination of fallow deer in proficiency test sample as 40% of participants of interlaboratory comparison test identified fallow deer as positive and another 20% judged the presence of fallow deer as dubious

^fContamination of fallow deer in proficiency test sample is likely as 83% of participants of interlaboratory comparison test identified fallow deer as positive

^gContamination of cattle in proficiency test sample is likely as 86% of participants of interlaboratory comparison test identified cattle as positive

and the 5 boiled sausage samples from the interlaboratory comparison tests were compared to the results obtained by the food control agencies using standard analytical methods. As 3 proficiency test samples had been independently analysed by 2 food control agencies

using different standard methods, these results were compared separately, resulting in a total of 70 comparisons. In 71.4% of the samples, exactly the same species (highlighted in green font in the Online Resource 1, Supplementary Information) were identified by the

		*	20	*	40	*		
Fwd_MaOH	:	GACGAGAAGACCCTATGGAGC-----					:	21
Fwd_POH	:	GACGAGAAGACCCTGTGGAAC-----					:	21
fallow deer	:	GACGAGAAGACCCTATGGAGCTTTAACTACTTAGCCCAAAGAATCAAATTTTATCAC					:	57
Rev_MaOH	:	-----					:	-
Rev_POH	:	-----					:	-
		60	*	80	*	100	*	
Fwd_MaOH	:	-----					:	-
Fwd_POH	:	-----					:	-
fallow deer	:	CAAGGAAACAACAACACTCTTTATGGGTTAACAGCTTTGATTGGGGTGATCTCGGA					:	113
Rev_MaOH	:	-----GGTTGGGGTGACCTCGGA					:	18
Rev_POH	:	-----GGTTGGGGCGACCTTGA					:	18

Fig. 3 Comparison of amplicon primer sequences to fallow deer (*Dama dama*) DNA sequence. Fwd_MaOH: sequence of forward mammal primer; Fwd_POH: sequence of forward poultry primer; fallow deer: part of fallow deer 16S rDNA extracted from GenBank sequences KJ870163.1 and JN632629.1; Rev_MaOH: reverse-com-

plementary sequence of reverse mammal primer; Rev_POH: reverse-complementary sequence of reverse poultry primer. Primer sequences are indicated without NGS adapters and positions with mismatches to fallow deer DNA are highlighted in white

metabarcoding approach and the standard methods, regardless of the bioinformatic workflow used. The discrepancies between NGS and the standard analytical methods regarding the identified species in the other samples are described in the following sub-sections.

3.3.1 NGS detected species that were not target species for the used standard methods

See Online Resource 1 (Supplementary Information): species highlighted in purple font. In 5 control samples, particular species were discovered by metabarcoding, that were not target species for the standard analytical methods:

- In a kangaroo meat sample (OPSONX_021), about 11% reads for red deer were measured using the metabarcoding method, that were missed when applying Sanger sequencing of 2 distinct meat pieces of the sample, and could also not be detected using a multiplex real-time PCR for cattle, pig, sheep and equids as well as with ELISA kits for beef, poultry, pork, horse and sheep.
- In a red deer sausage with declared pork and red deer content (OPSONX_042), chamois (*Rupicapra rupicapra*) was detected by metabarcoding, which was not a target species for the DNA chip used (MEAT 5.0 LCD-Array).
- In a red deer salami sample (OPSONX_047), sika deer (*Cervus nippon*) was detected with about 14% of reads by all three bioinformatic workflows, that was not a target species of the DNA chip used (also MEAT 5.0 LCD-Array).

- In another red deer sausage (OPSONX_043), sika deer was identified with a much smaller read percentage (about 1%) only by the dereplication and the OTU clustering workflow but not by the dada2 workflow. Here, it is unclear, whether the reads for sika deer are due to sika deer traces in the sample or whether they resulted from amplification or sequencing errors that were removed by the denoising strategy implemented in the dada2 software package.
- In a liver sausage sample with declared wild boar/pork and deer content (OPSONX_040), the DNA chip used detected only “roe deer (*Capreolus capreolus*)” which is the European roe, whereas the metabarcoding also found Siberian roe (*Capreolus pygargus*) (4.1% of reads) in addition to European roe.

Moreover, 2 proficiency test samples contained elk (*Alces alces*), which was correctly detected by the metabarcoding method. These samples were independently analysed in 2 different food control laboratories using the MEAT 5.0 LCD-Array, but one laboratory did not detect elk (in OPSONX_067 and OPSONX_068), while the other laboratory identified elk (in OPSONX_044 and OPSONX_045) because the DNA chip MEAT Plus 3.0, which contains elk as a target species, was additionally used (Online Resource 1, Supplementary Information). In a third proficiency test sample (OPSONX_048), the metabarcoding method detected common wallaroo (*Osphranter/Macropus robustus*) in addition to red kangaroo and Eastern grey kangaroo that was not detected with the DNA chip, which only targets “Kangaroo (*M. rufus*, *M. giganteus*)” (the red kangaroo and the Eastern grey kangaroo).

		*	20	*	40	*	6		
Fwd_MaOH	:	GACGAGAAGACCCTATGGAGC-----						:	21
Fwd_POH	:	GACGAGAAGACCCTGTGGAAC-----						:	21
salmon	:	GACGAGAAGACCCTATGGAGCTTTAGACACCAGGCAGATCACGTCAAGTAACCTTGAAT						:	59
Rev_MaOH	:	-----						:	-
Rev_POH	:	-----						:	-
		0	*	80	*	100	*		
Fwd_MaOH	:	-----						:	-
Fwd_POH	:	-----						:	-
salmon	:	TAACAAGTAAAAACGCAGTGACCCCTAGCCCATATGTCCTTTGGTTGGGGCGACCGCGGG						:	118
Rev_MaOH	:	-----GGTTGGGGTGCCTCGGA						:	18
Rev_POH	:	-----GGTTGGGGCGACCTTGA						:	18

Fig. 4 Comparison of primer sequences to Atlantic salmon (*Salmo salar*) DNA sequence. Fwd_MaOH: sequence of forward mammal primer; Fwd_POH: sequence of forward poultry primer; salmon: part of Atlantic salmon 16S rDNA extracted from GenBank sequence KR476892.1; Rev_MaOH: reverse-complementary sequence of

reverse mammal primer; Rev_POH: reverse-complementary sequence of reverse poultry primer. Primer sequences are indicated without NGS adapters and positions with mismatches to salmon DNA are highlighted in white

Furthermore, for one sample (OPSONX_046=_069), the comparative method (DNA chip) reported the white-fronted goose (“Goose”, *Ansa albifrons*), whereas metabarcoding found the greylag goose (*Anser anser*) (Online Resource 1, Supplementary Information, highlighted in orange). However, a manually performed browser-based BLAST search in GenBank revealed identical sequence identities of the NGS sequence with the white-fronted goose and the greylag goose.

3.3.2 False-positive or false-negative metabarcoding results

See Online Resource 1, Supplementary Information: species highlighted in red font. As in the mock mixtures, fallow deer could not be recovered by metabarcoding in all 7 samples (OPSONX_012, _043, _044=_067, _045=_068, _046=_069, _047 and _049) that were judged positive for fallow deer with standard methods, presumably due to the mismatching NGS primer sequence (see above). In one sample (OPSONX_038) with declared egg content, chicken DNA was only determined with PCR–RFLP, real-time PCR and DNA chip, but not with ELISA or the metabarcoding method.

Moreover, in 4 samples (OPSONX_037, _051, _052 and _063) with high read percentages for cattle (*Bos taurus*), wild yak (*Bos mutus*) was also recovered in small proportions (max. 0.7% reads). Upon closer inspection of the BLAST hit description (Accession number CP027084.1), it was obvious that this GenBank sequence did not come from the mitogenome but was detected on a chromosomal region within an NGS whole genome

assembly study. Furthermore, when submitting the identified NGS sequence from this study manually to a GenBank BLAST search, a hit for a chromosomal region of cattle was also found with 100% sequence identity (accession number LR962746.1) (Fig. 5). Thus, it is very likely that the wild yak reads derive from sequencing a cattle 16S rDNA pseudogene [nuclear mitochondrial DNA, NUMT; see (Gaziev and Shaikhaev (2010))] present not only in cattle but also in wild yak.

3.3.3 Unresolved cases of discrepancy

See Online Resource 1, Supplementary Information: species highlighted in blue font. In 4 samples, particular species (OPSONX_001 and OPSONX_054: turkey; OPSONX_025: cattle; OPSONX_042: cattle, goat and sheep) were detected with the standard methods, but not with the metabarcoding method. However, as only qualitative but not quantitative information of the identified species was provided by the food control laboratories for most samples, it remains largely unclear, whether these findings were due to traces or substantial ingredients of the meat products. It seems likely that some species were missed by NGS due to very low concentrations of the DNA of the respective species, on the other hand false-positive findings by using the standard methods may also be a possible reason.

In 3 other samples, individual species (OPSONX_047 and _049: cattle; OPSONX_053: pig) were identified by metabarcoding in small percentages (0.9–1.5%), that were not reported by the food control laboratories, although the species were target species for the methods used. The reasons for these discrepancies (either false-positive metabarcoding results or false-negative results by standard methods)

	*	20	*				
ASV1	:	TTTA	ACTAACCA	ACCCAAAGAGAATAGATTTAACCAT	:	37	
B. <u>taurus</u> <u>mitoch.</u>	:	:	37	
B. <u>mutus</u> <u>mitoch.</u>	:	A...A.....	:	37	
ASV29	:	G.....	AG.C.A..CA.....	:	37	
B. <u>taurus</u> <u>chrom.</u>	:	G.....	AG.C.A..CA.....	:	37	
B. <u>mutus</u> <u>chrom.</u>	:	G.....	AG.C.A..CA.....	:	37	
		40	*	60	*		
ASV1	:	TAAGGAATAACAACAATCTCCATGAGTTGGTAGTTTC	:		:	74	
B. <u>taurus</u> <u>mitoch.</u>	:	:	74	
B. <u>mutus</u> <u>mitoch.</u>	:	.G.....	C.....	:	74	
ASV29	:	C.....	C.....	C.....	C...C.....	:	74
B. <u>taurus</u> <u>chrom.</u>	:	C.....	C.....	C.....	C...C.....	:	74
B. <u>mutus</u> <u>chrom.</u>	:	C.....	C.....	C.....	C...C.....	:	74

Fig. 5 Comparison of cattle (*Bos taurus*) and wild yak (*Bos mutus*) 16S rDNA sequences and probable pseudogene sequences of the two species. ASV1 and ASV29 were amplicon sequence variants from sample OPSONX_037 analysed with the dada2 bioinformatic pipeline. Whereas the first hit of ASV1 was cattle mitochondrial DNA, the first hit of ASV29 was DNA from wild yak chromosome 16.

However, the sequence of the chromosome 16 part of wild yak is identical to a corresponding chromosome 16 part of cattle. Sequences were extracted from GenBank sequences MN714195.1 (*Bos taurus* mitochondrion), MK033130.1 (*Bos mutus* mitochondrion), LR962746.1 (*Bos taurus* chromosome 16), CP027084.1 (*Bos mutus* chromosome 16)

could not be clarified because the actual composition of the samples was not known.

3.4 Comparison of the different bioinformatic workflows

Overall, the dereplication workflow, the OTU clustering workflow and the dada2 workflow produced very similar results. In only 1 sample (OPSONX_043), the identified species differed between the workflows: Low percentages (0.8% and 1.0%, respectively) of sika deer were recovered by the dereplication and the OTU clustering workflow, but not with the dada2 workflow (see above and Online Resource 1, Supplementary Information). However, since the actual composition was not known and sika deer was not a target species of the standard method used, it cannot be determined which workflow was correct. Otherwise, only minor differences between the workflows were observed that did not affect the outcome of the identified species in the samples. For example, the recovery of reads after performing the data processing was higher with the dada2 workflow compared to the other workflows (Table 2), and the dada2 workflow produced less false-positive species in the mock mixtures compared to the other workflows (Table 4). However, since these species were below the threshold, this is considered as negligible. On the other hand, the dereplication workflow and the OTU clustering workflow produced substantially more different sequences per sample compared to the dada2 workflow. While for the dereplication workflow an

average of 3,428 different sequences per sample had to be queried against the sequence database, the average number of sequences per sample decreased to 520 OTUs (15.2% of dereplicated sequences) and 15 ASVs (0.4% of dereplicated sequences), respectively (Table 2).

4 Discussion

63 meat product samples, 5 boiled sausage proficiency test samples, 9 mock mixtures with defined species compositions and 6 negative controls of 5 different types were analysed using a published 16S rDNA metabarcoding method. Using the resulting sequence data, the source of unspecific reads commonly observed in samples, the robustness of the bioinformatic data processing and the comparability with diverse standard analytical methods were investigated.

4.1 Source of unspecific sequences

No significant differences in read numbers were observed in the different types of negative controls (undiluted or diluted amplicon PCR, undiluted or diluted index PCR, or simulated index combinations). Therefore, it seems likely that the main proportion of reads in the negative controls resulted either from incorrect index molecules or from errors during amplification or sequencing of the indices and thus belong to other samples on the flow cell rather than originating from (cross-) contamination with DNA from the environment,

other samples or other amplicons. Otherwise, much higher read numbers would be expected in the undiluted negative controls compared to the diluted controls. The same applies to the undiluted and diluted negative controls in comparison to the simulated negative controls where the index combinations were only used for the sequencing sample sheet and for which no PCR reaction had been prepared. Thus, assuming an average of about 133 000 reads per sample, of which about 206 reads were due to erroneous indices or sequencing errors in index sequences from other samples (see Sect. 3.1), one would expect about 0.16% of reads from other samples of the same run in the respective sample. This could be reduced to about 0.04% reads when allowing no mismatches in the index sequences. Unfortunately, the software on board the MiSeq instrument does not allow restriction to zero mismatches, so the demultiplexing with allowing zero mismatches has to be carried out independent of the machine. The use of longer indices (> 8 nucleotides), either from an alternative indexing kit or as internal parts of the amplicon primers in a one-step amplicon protocol (Kozich et al. 2013), could also reduce the number of reads from other samples.

However, as the maximum read percentages for false-positive species in the mock mixtures noticeably exceeded 0.16% or 0.04%, respectively, additional sources for unspecific reads besides false index sequences or cross-contaminations must be considered. It was observed that many false-positive reads with higher percentages were assigned to species closely related to those actually present in the samples, like urial (*Ovis vignei*, e.g. 0.37% in OTU workflow) and argali (*Ovis ammon*, 0.26% in OTU workflow) in samples containing sheep, or sambar deer (*Rusa unicornis*, e.g. 0.30% in OTU workflow) and Eld's deer (*Rucervus eldii*, 0.12% in OTU workflow) in samples containing deer species. This suggests that single nucleotide sequence changes caused by the amplification or sequencing steps or incorrect sequences in the NCBI nucleotide database could be the cause for these non-specific reads. Unexpectedly, these related but non-specific species were also identified using the dada2 workflow to some extent, which has an implemented denoising step to derive error rates from the sequences and remove incorrect sequences from the dataset. Thus, these particular sequences could therefore also represent natural variations in the genomes of the animals present in the samples. Moreover, the probable sequencing of a pseudogene in cattle (see Sect. 3.4) was another reason for false-specific sequences in cattle containing samples. The same pseudogene appears to be present in wild yak, which was identified as first BLAST hit but also with the same sequence identity (100%) in cattle chromosomal DNA (see above). Therefore, the use of a customized database with

only reliable mitogenome sequences, as described in Dobrovolny et al. (2019) and Preckel et al. (2021), is preferable, provided that it contains all species potentially present in meat products with sufficient sequence variation. In addition, the bioinformatic workflows should be adapted to consider not only the first BLAST hit but all hits within an appropriate identity range, so that species with identical sequences are reported only on genus level. Otherwise, the first hit could be wrongly interpreted as an undeclared ingredient, while the declared species would not be found. An appropriate bioinformatics pipeline, 'FoodMe', for the analysis of metabarcoding of food and feed samples can be downloaded from Zenodo (Denay 2022).

4.2 Robustness of the bioinformatic data processing

Although the three bioinformatic workflows pursued different strategies, the results were almost always identical. Thus, the data evaluation seems to be quite robust and different algorithms can be applied. Even the denoising step, the core of the dada2 tool suite, also produced almost no different results and could not remove the erroneous sequences that lead to the identification of small percentages of related species such as urial or sambar deer. However, as the dereplication and the OTU clustering workflow produced considerably more sequences per sample, these workflows are also much more demanding than the dada2 workflow in terms of computing power and computing time. With regard to this, the dada2 workflow seems more favourable, especially for the processing of large sequencing datasets.

4.3 Comparability of the metabarcoding approach with standard analytical methods

For most of the samples and especially for samples with less complex species compositions, the 16S rDNA metabarcoding approach identified exactly the same species. However, the main advantage of this metabarcoding approach compared to the species-specific methods such as for example real-time PCR, DNA chip, or ELISA is certainly the discovery of unexpected species. Although the DNA chips extensively used in this study cover a fairly wide range of target species, they could not detect elk, sika deer, Siberian roe, chamois or common wallaroo in contrast to the NGS method. Moreover, a combination of Sanger sequencing and multiplex real-time PCR was not able to identify an unexpected admixture of red deer in a kangaroo meat product. On the other hand, fallow deer cannot be identified by metabarcoding using the described primers from Dobrovolny et al. (2019). However, an additional reverse primer for the detection of fallow deer is described in the method specification for the detection of mammals and birds

by DNA metabarcoding, which will be published by the Federal Office of Consumer Protection and Food Safety (BVL, Germany) through an update of the Official Collection of Methods of Analysis in 2023. As shown in this study and also by others (e.g. Preckel et al. 2021), the read percentages do not necessarily reflect the actual species shares of the samples accurately. It should therefore be borne in mind that the metabarcoding method is more of a screening method for compositional analysis in terms of species in complex food samples, but not for quantifying the exact proportion of the particular species (Defra 2020). As a screening method, it is a useful alternative to the DNA MEAT LCD-Arrays (Chipron, Berlin, Germany) that were used by many laboratories in this study but have recently (in May 2022) been withdrawn from the market (according to a customer letter from manufacturer).

For best results, samples should be screened using 16S rDNA metabarcoding and results confirmed with orthogonal species-specific methods, especially when quantitative results are required.

5 Conclusion

This study showed the successful cooperation of the competent German food control authorities with the German NRZ-Authent within the framework of operation OPSON X in the application and validation of a published 16S rDNA metabarcoding method. The NGS method proved to be a suitable method for the identification of mammal and poultry species, especially for food samples with complex species compositions, and was in some cases able to identify particular species that had been overlooked when using standard analytical methods like DNA chips, Sanger sequencing, or multiplex real-time PCR. As expected, fallow deer was the only species that could not be detected with the applied NGS method. False-positive hits which occurred with small read percentages were likely caused by indexing errors, introduced by PCR or sequencing error, or reflect natural variability in animal's mitochondrial genome. Consequently, an appropriate read percentage threshold must be applied, but this also means that small admixtures of species can lead to read percentages that are below the threshold and thus give false-negative results. The bioinformatic data evaluation was shown to be very robust, as different strategies and algorithms led to nearly identical results. The dada2 workflow may be favourable when analysing large datasets as it is less demanding in computing capacities due to the output of significantly fewer sequences per sample to be matched with the database. In order to exclude pseudogenes or other non-meaningful hits from the results, the use of an audited sequence reference database (Neto et al. 2021) containing

only validated mitogenome or 16S rDNA sequences is recommendable. Furthermore, it would make sense to include more BLAST hits than only the first for a valid identification of the species (or genus), as sometimes, two or more species share the same 16S rDNA sequences and a specification of the genus may be more valid.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00003-023-01437-w>.

Acknowledgements The authors kindly thank Sandra Bornkessel of the Federal Office of Consumer Protection and Food Safety (Germany) for relaying the laboratory cooperation between the German competent food control agencies and the NRZ-Authent. We also kindly thank Kai-Uwe Scholibo and Adrian Prager for the technical assistance in the laboratory, Andreas Dötsch for providing computer facilities and Regina Klapper for giving advice when needed.

Author contributions KK und IH contributed to the study conception and design. Material preparation, data collection and analysis were performed by KK, AG, GD, LG, AG, MH, MH, IH, GN, MP, KP, BS, IV, SW and AW. The first draft of the manuscript was written by KK and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Availability of data and material The raw sequencing data, i.e. the fastq files after demultiplexing on the MiSeq instrument with default settings (allowing one mismatch in index sequences), have been deposited in the NCBI Sequence read Archive (SRA) (BioProject accession number PRJNA926813).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References


- Akbar A, Shakeel M, Al-Amad S, Akbar A, Ali AK, Rahmeh R, Alo-taibi M, Al-Muqatea S, Areeba S, Arif A, Fayyaz M, Ahmad Khan I, Ahmed S, Hussain A, Ghulam Musharraf S (2021) A simple and sensitive NGS-based method for pork detection in complex food samples. *Arab J Chem* 14(5):103124. <https://doi.org/10.1016/j.arabjc.2021.103124>

- Aronesty E (2013) Comparison of sequencing utility programs. *Open Bioinform J* 7(1):1–8. <https://doi.org/10.2174/1875036201307010001>
- Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, the Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26(14):1783–1785. <https://doi.org/10.1093/bioinformatics/btq281>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13(7):581–583. <https://doi.org/10.1038/nmeth.3869>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinform*. <https://doi.org/10.1186/1471-2105-10-421>
- Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N (2015) NCBI BLAST+ integrated into Galaxy. *GigaScience*. <https://doi.org/10.1186/s13742-015-0080-7>
- European Commission (2021) The EU agri-food fraud network and the administrative assistance and cooperation system: 2020 annual report. Publications Office, Luxembourg. <https://doi.org/10.2875/20163>
- Defra (2020) Authenticity methodology working group view on use of next generation sequencing for food authenticity testing. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1016866/Authenticity_Methodology_Working_Group_view_on_use_of_Next_Generation_Sequencing_for_food_authenticity_testing.pdf. Accessed 15 Jan 2023
- Denay G (2022) CVUA-RRW/FoodDMe: Foodme v1.6.3 (1.6.3). Zenodo. <https://doi.org/10.5281/zenodo.7078595>
- DIN (2019) DIN CEN/TS 17303:2019-06. Foodstuffs - DNA barcoding of fish and fish products using defined mitochondrial cytochrome b and cytochrome c oxidase I gene segments. Beuth Verlag, Berlin. <https://doi.org/10.31030/2856520>
- Dobrovolny S, Blaschitz M, Weinmaier T, Pechatschek J, Cichna-Markl M, Indra A, Hufnagl P, Hochegger R (2019) Development of a DNA metabarcoding method for the identification of fifteen mammalian and six poultry species in food. *Food Chem* 272:354–361. <https://doi.org/10.1016/j.foodchem.2018.08.032>
- Dobrovolny S, Uhlig S, Frost K, Schlierf A, Nichani K, Simon K, Cichna-Markl M, Hochegger R (2022) Interlaboratory validation of a DNA metabarcoding assay for mammalian and poultry species to detect food adulteration. *Foods* 11(8):1108. <https://doi.org/10.3390/foods11081108>
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996–998. <https://doi.org/10.1038/nmeth.2604>
- Federal Office of Consumer Protection and Food Safety (2007) L 06.26/27-2:2007-12 Untersuchung von Lebensmitteln—Nachweis Pferd-spezifischer DNA-Sequenzen in Fleisch-Vollkonserven mit der PCR und Bestätigung durch Restriktionsanalyse. Amtliche Sammlung von Untersuchungsverfahren nach § 64 LFGB, Beuth Verlag, Berlin
- Federal Office of Consumer Protection and Food Safety (2016a) L 08.00-61:2016a-03 Untersuchung von Lebensmitteln - Nachweis der Tierarten Rind, Schwein, Pute und Huhn in Wurstwaren durch Multiplex-real-time PCR. Amtliche Sammlung von Untersuchungsverfahren nach § 64 LFGB, Beuth Verlag, Berlin
- Federal Office of Consumer Protection and Food Safety (2016b) L 08.00-62:2016b-03 Untersuchung von Lebensmitteln—Nachweis der Tierarten Rind, Schwein, Schaf und Equiden in Wurstwaren durch Multiplex-real-time PCR. Amtliche Sammlung von Untersuchungsverfahren nach § 64 LFGB, Beuth Verlag, Berlin
- Gaziev AI, Shaikhaev GO (2010) Nuclear mitochondrial pseudogenes. *Mol Biol* 44(3):358–368. <https://doi.org/10.1134/S002689310030027>
- Haynes E, Jimenez E, Pardo MA, Helyar SJ (2019) The future of NGS (next generation sequencing) analysis in testing food authenticity. *Food Control* 101:134–143. <https://doi.org/10.1016/j.foodcont.2019.02.010>
- Hellmann SL, Ripp F, Bikar S-E, Schmidt B, Köppel R, Hankeln T (2020) Identification and quantification of meat product ingredients by whole-genome metagenomics (All-Food-Seq). *Eur Food Res Technol* 246:193–200. <https://doi.org/10.1007/s00217-019-03404-y>
- Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, Taylor J, Nekrutenko A (2020) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res* 48(W1):W395–W402. <https://doi.org/10.1093/nar/gkaa434>
- Jiang M, Xu S-F, Tang T-S, Miao L, Luo B-Z, Ni Y, Kong F-D, Liu C (2022) Development and evaluation of a meat mitochondrial metagenomic (3MG) method for composition determination of meat from fifteen mammalian and avian species. *BMC Genom* 23(1):36. <https://doi.org/10.1186/s12864-021-08263-0>
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79(17):5112–5120. <https://doi.org/10.1128/AEM.01043-13>
- Liu YC, Ripp F, Koeppel R, Schmidt H, Hellmann SL, Weber M, Krombholz CF, Schmidt B, Hankeln T (2017) AFS: identification and quantification of species composition by metagenomic sequencing. *Bioinformatics* 33(9):1396–1398. <https://doi.org/10.1093/bioinformatics/btw822>
- Liu X, Liu Z, Cheng Y, Wu H, Shen W, Liu Y, Feng Q, Yang J (2021) Application of next-generation sequencing technology based on single gene locus in species identification of mixed meat products. *J Food Qual*. <https://doi.org/10.1155/2021/4512536>
- LVU (2019) Ermittlung der Tierart (2019)—Laborvergleichsuntersuchung „Tierarten (2019)“, report Herbolzheim, Germany
- LVU (2021) Ermittlung der Tierart (2021)—Laborvergleichsuntersuchung „Tierarten (2021)“, report Herbolzheim, Germany
- Mahama S, Chebako, H, Sirikwanpong, S, Mahamad, P, Santiworakul, NY, Suksuwan, A, Dahlan, W, Nopponpunth, V (2022) Simultaneous identification of four meat species (cattle, chicken, fish, and pig) using next generation sequencing (NGS). In: Proceedings of the international halal science and technology conference 14(1):182–193. <https://doi.org/10.31098/ihsatec.v14i1.500>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10–12. <https://doi.org/10.14806/ej.17.1.200>
- Meyer R, Höfelein C, Lüthy J, Candrian U (1995) Polymerase chain reaction–restriction fragment length polymorphism analysis:

- a simple method for species identification in food. *J AOAC Int* 78(6):1542–1551. <https://doi.org/10.1093/jaoac/78.6.1542>
- Neto L, Pinto N, Proença A, Amorim A, Conde-Sousa E (2021) 4SpecID: reference DNA libraries auditing and annotation system for forensic applications. *Genes* 12(1):61. <https://doi.org/10.3390/genes12010061>
- Pan Y, Qiu D, Chen J, Yue Q (2020) Combining a COI mini-barcode with next-generation sequencing for animal origin ingredients identification in processed meat product. *J Food Qual.* <https://doi.org/10.1155/2020/2907670>
- Preckel L, Brünen-Nieweler, C, Denay, G, Petersen, H, Cichna-Markl, M, Dobrovoly, S, Hochegger, R (2021) Identification of mammalian and poultry species in food and pet food samples using 16S rDNA metabarcoding. *Foods* 10(11):2875. <https://doi.org/10.3390/foods10112875>
- Ripp F, Kromholz CF, Liu Y, Weber M, Schäfer A, Schmidt B, Köp-pel R, Hankeln T (2014) All-food-Seq (AFS): a quantifiable screen for species in biological samples by deep DNA sequencing. *BMC Genom* 15:639. <https://doi.org/10.1186/1471-2164-15-639>
- Rognes T, Mahé, F, xflouris (2015) vsearch: VSEARCH version 1.0.16. <https://doi.org/10.5281/zenodo.15524>
- Wolf C, Rentsch J, Hübner P (1999) PCR–RFLP analysis of mitochondrial DNA: a reliable method for species identification. *J Agric Food Chem* 47(4):1350–1355. <https://doi.org/10.1021/jf9808426>
- Xing R-R, Nan W, Hu R-R, Zhang J-K, Han J-X, Chen Y (2019) Application of next generation sequencing for species identification in meat and poultry products: a DNA metabarcoding approach. *Food Control* 101:173–179. <https://doi.org/10.1016/j.foodcont.2019.02.034>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kristina Kappel¹  · Andreas Gadelmeier² · Grégoire Denay³ · Lars Gerdes⁴ · Andrea Graff⁵ · Margit Hagen⁶ · Melanie Hassel⁷ · Ingrid Huber⁴ · Gabriele Näumann⁵ · Melanie Pavlovic⁴ · Klaus Pietsch⁸ · Barbara Stumme⁵ · Inger Völkel⁹ · Simone Westerdorf¹⁰ · Anne Wöhlke¹¹ · Rupert Hochegger¹² · Erik Brinks¹³ · Charles Franz¹³ · Ilka Haase²

¹ National Reference Centre for Authentic Food, Max Rubner-Institut, Hermann-Weigmann-Str. 1, 24103 Kiel, Germany

² National Reference Centre for Authentic Food, Max Rubner-Institut, E.-C.-Baumann-Straße 20, 95326 Kulmbach, Germany

³ Chemical and Veterinary Analytical Institute Rhein-Ruhr-Wupper (CVUA-RRW), Deutscher Ring 100, 47798 Krefeld, Germany

⁴ Bavarian Health and Food Safety Authority (LGL), Veterinärstr. 2, 85764 Oberschleißheim, Germany

⁵ Institute for Hygiene and Environment, Hamburg Public Laboratories for Food Safety, Health Protection and Environmental Assessment, Marckmannstraße 129 a/b, 20539 Hamburg, Germany

⁶ State Institute of Chemical and Veterinarian Analysis Eastwestphalia-Lippe, Westerfeldstr. 1, 32758 Detmold, Germany

⁷ Institute for Food of Animal Origin, LUA RLP, Blücherstr. 38, 56073 Koblenz, Germany

⁸ Chemical and Veterinary Analytical Institute Freiburg (CVUA Freiburg), Bissierstrasse 5, 79114 Freiburg, Germany

⁹ Chemical and Veterinary Analytical Institute Westphalia (CVUA Westfalen), Zur Taubeneiche 10-12, 59821 Arnsberg, Germany

¹⁰ Chemical and Veterinary Analytical Institute Rheinland (CVUA Rheinland), Winterstr. 19, 50354 Hürth, Germany

¹¹ Food and Veterinary Institute Braunschweig/Hannover, Lower Saxony State Office for Consumer Protection and Food Safety, Dresdenstr. 2 und 6, 38124 Brunswick, Germany

¹² Department for Molecular Biology and Microbiology, Institute for Food Safety Vienna, Austrian Agency for Health and Food Safety (AGES), Spargelfeldstrasse 191, 1220 Vienna, Austria

¹³ Department of Microbiology and Biotechnology, Max Rubner-Institut, Hermann-Weigmann-Str. 1, 24103 Kiel, Germany