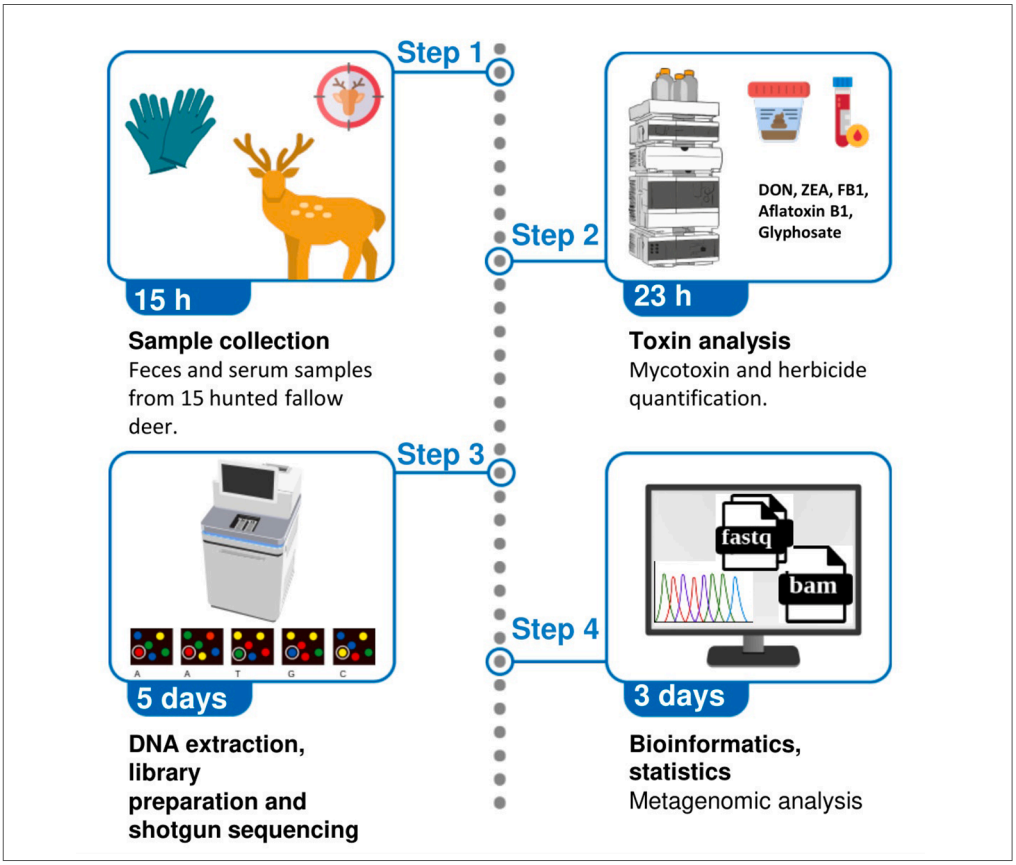
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for the assessment of the impact of mycotoxins and glyphosate residues on the gut microbiome and resistome of European fallow deer



Here, we present a protocol to describe the bacteriome of the intestinal content of toxin-exposed fallow deer. We describe steps for measuring fecal mycotoxin (deoxynivalenol, zearalenone, fumonisin B1, and aflatoxin B1) levels using liquid chromatography-mass spectrometry, as well as serum glyphosate. We then detail a short-read shotgun DNA sequencing-based bioinformatic pipeline for the toxin level-associated analysis of the bacteriome and resistome and the construction of metagenome-assembled bacterial genomes. This protocol has potential applications in further toxin level-associated metagenome studies.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Adrienn Gréta Tóth, Melinda Paholcsek, Norbert Solymosi, ..., Sára Ágnes Nagy, Szilamér Ferenczi, Zsuzsanna Szőke

tothadrienngreta@gmail.com (A.G.T.)
ferenczi.szokezsuzsi@gmail.com (Z.S.)

Highlights

Instructions for fecal mycotoxin and serum glyphosate measurements

Experimental steps for shotgun metagenomic sequencing of complex microbiological samples

Guidance on bioinformatic microbiome analysis

Tóth et al., STAR Protocols 7, 104498
June 19, 2026
https://doi.org/10.1016/j.xpro.2026.104498

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for the assessment of the impact of mycotoxins and glyphosate residues on the gut microbiome and resistome of European fallow deer

Adrienn Gréta Tóth,[1,11,12,*] Melinda Paholcsek,[2] Norbert Solymosi,[3,4] Anikó Stágel,[5] Patrik Gömbös,[6] Katalin Posta,[7] István Lakatos,[8] Sára Ágnes Nagy,[4] Szilamér Ferenczi,[7,9,10] and Zsuzsanna Szőke[8,10,*]

[1]Institute for Animal Breeding, Nutrition and Laboratory Animal Science, University of Veterinary Medicine, 1078 Budapest, Hungary

[2]Complex Systems and Microbiome-innovations Centre, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Debrecen, Hungary

[3]Centre for Bioinformatics, University of Veterinary Medicine, 1078 Budapest, Hungary

[4]Department of Physics of Complex Systems, Eötvös Loránd University, 1117 Budapest, Hungary

[5]Hungarian National Blood Transfusion Service Nucleic Acid Testing Laboratory, Budapest, Hungary

[6]Agrobiotechnology and Precision Breeding for Food Security National Laboratory, Institute of Physiology and Nutrition, Department of Physiology and Animal Health, Hungarian University of Agriculture and Life Sciences, 7400 Kaposvár, Hungary

[7]Department of Microbiology and Applied Biotechnology, Institute of Genetics and Biotechnology, Hungarian University of Agriculture and Life Sciences, Gödöllő 2100, Hungary

[8]Department of Animal Biotechnology, Agribiotechnology and Precision Breeding for Food Security National Laboratory, Institute of Genetics and Biotechnology, Hungarian University of Agriculture and Life Sciences, Gödöllő 2100, Hungary

[9]Laboratory of Molecular Neuroendocrinology, Institute of Experimental Medicine, Hungarian Research Network, Budapest, Hungary

[10]These authors contributed equally

[11]Technical contact

[12]Lead contact

*Correspondence: tothadrienngreta@gmail.com (A.G.T.), ferenczi.szokezsuzsi@gmail.com (Z.S.)
https://doi.org/10.1016/j.xpro.2026.104498

SUMMARY

Here, we present a protocol to describe the bacteriome of the intestinal content of toxin-exposed fallow deer. We describe steps for measuring fecal mycotoxin (deoxynivalenol, zearalenone, fumonisin B1, and aflatoxin B1) levels using liquid chromatography-mass spectrometry, as well as serum glyphosate. We then detail a short-read shotgun DNA sequencing-based bioinformatic pipeline for the toxin level-associated analysis of the bacteriome and resistome and the construction of metagenome-assembled bacterial genomes. This protocol has potential applications in further toxin level-associated metagenome studies.
For complete details on the use and execution of this protocol, please refer to Tóth et al.[1]

BEFORE YOU BEGIN

Climate change brings unprecedented challenges in the agricultural sector, including in the distribution and toxin-producing capacity of mycotoxin-producing fungi and the need for the widespread use of herbicides in intensive food and feed production settings. Several mycotoxins and herbicide residues from the environment can influence animal, and especially herbivore and ruminant health. These may also alter the gut microbiome composition and the microbial genomic characteristics. To investigate the effects of varying exposure levels of mycotoxins, such as zearalenone (ZEA), aflatoxin B1 (AFs), deoxynivalenol (DON), fumonisin B1 (FB1) and other toxic compounds, such as glyphosate residues (Glypho), intestinal content samples from fallow deer or other ruminants can be analyzed

using next-generation sequencing based metagenomic approaches. While traditional methods for the analysis of the gut microbiota are often targeted, prioritize faster-growing, or less fastidious bacteria and rely on the phenotypically expressed antimicrobial susceptibility, our method facilitates the identification and characterization of the broadest range of bacteria without losing the bacterial relative abundance rates.

This protocol details the steps for measuring fecal mycotoxin and serum glyphosate levels and assessing the microbiome of intestinal samples from fallow deer in light of the toxic compound levels. Thus, the protocol encompasses both wet-lab and dry-lab steps and covers the entire process from intestinal sample acquisition and toxin analysis to the shotgun sequencing and the subsequent bioinformatic analysis of the metagenomic datasets including the assessment of the bacteriome and the resistome, and the construction of metagenome-assembled genomes. The protocol can be applied in ruminants, and potentially in other animals, preferably herbivores, as well. The approach offers the possibility to focus on the set of antimicrobial resistance genes (resistome) that might contribute to the survival of gut bacteria.

Innovation

This protocol offers a comprehensive framework for analyzing fallow deer intestinal microbiomes in relation to measured levels of mycotoxins—zearalenone, aflatoxin B1, deoxynivalenol, fumonisin B1—and glyphosate in the same samples. Each experimental and bioinformatic step is built on extensively optimized procedures designed to accurately assess both toxin content and microbial composition. The novelty of this approach lies in its integrated combination of methods, enabling detailed insights from shotgun metagenomic datasets. Beyond its immediate application, this protocol can serve as a template for exploring the relationships between chemical exposures and microbiome structures in diverse biological samples.

Institutional permissions (if applicable)

According to the statement of the Institutional Review Board (NAIK MBK MÁB 004-09/2018), the study is not considered as an experiment with animals because the researchers collected samples from legally harvested fallow deer hinds; consequently, the ethical treatment rules are not applicable.

Researchers using this protocol must have permission and follow the rules of their local authorities.

Preparations for measuring fecal DON, ZEA, and FB1 mycotoxins

Timing: 3 h

1. Prepare the extraction solvent, mixing acetonitrile, water and acetic-acid (50 : 50: 0.1, v/v/v).
2. Set High-Performance Liquid Chromatography (HPLC) conditions.
   a. Set column temperature to 40°C.
   b. Set eluent flow rate to 0.3 mL/min.
   c. Set injection volume to 10 μL.
   d. Prepare mobile phases for gradient elution.

   Note: Mobile phase A: 1% (v/v) acetic acid water, mobile phase B: 1% (v/v) acetonitrile.

   e. Sonicate mobile phase B for 15 min before loading into the system.
   f. Load mobile phases into HPLC.
   g. Run mobile phase A through the degasser unit of the HPLC system (DGU-20A).

Note: These details may vary depending on the HPLC system used. By the original study, we used a Prominence-type HPLC system (Shimadzu, Kyoto, Japan) with a Kinetex XB-C18 (100 mm × 2.1 mm, 2.6 μm) HPLC column.

h. Set gradient elution program.
   i. Start with 5% mobile phase B and hold between 0–1 min.
   ii. Set a linear increase of mobile phase B from 5% to 60% between 1–3 min.
   iii. Hold mobile phase B at 60% between 3–4 min.
   iv. Set a linear increase of mobile phase B from 60% to 95% between 4–8 min.
   v. Hold mobile phase B at 95% between 8–10.9 min.
   vi. Set a linear decrease of mobile phase B from 95% to 5% between 10.9–12.5 min.
   vii. Hold mobile phase B at 5% for re-equilibration between 12.5–15 min.

3. Set Mass Spectrometry (MS) conditions.
   a. Prepare 100 mg/L standard solutions for each mycotoxin.

Note: Mycotoxin standard solvent was acetonitrile, stored in −20°C in all cases, except for FB1, which was solved in acetonitrile-water (1:1 v:v) and stored in 8°C.

   b. For calibration, prepare standard solutions of at least 3 different concentration levels (e.g., 0.1 mg/L; 0.5 mg/L; 1 mg/L). Use the extraction solvent for dilution.

Note: Weekly calibration is recommended.

   c. Set capillary voltage to 4.5 kV, interface temperature to 350 °C, ion source temperature to 250°C, heated block temperature to 200°C.
   d. Use pure nitrogen as drying (15 L/min) and nebulizing (1.5 L/min) gas.

Optional: Produce pure nitrogen with a PEAK NM32LA nitrogen generator.

   e. Set the dwell times of each m/z channels to 50 ms.
   f. Determine the parameters used for the identification of each toxin.
      i. Inject 100 mg/L standard solutions of each mycotoxin in the LC autosampler in scan mode.
      ii. Record retention times, m/z ratios and ion modes for each toxin.
   g. Prepare external calibration standards for mycotoxin quantification.
      i. Dilute the standard solutions to span the linearity range (per unit sample mass) of 0.004–4 mg/kg.
   h. Use the calibration standards to establish the limits of detection (LOD) and the limits of quantification (LOQ) for each toxin.

Preparations for measuring fecal aflatoxin B1

Timing: 3 h

4. Prepare reagents for mycotoxin extraction.
   a. Prepare the extraction solvent using 2% formic acid in MeCN/H2O (1:1 v/v).
   b. Measure and mix 0.8g MgSO4 (anhydrous salt) and 0.2 NaCl.
5. Set High Performance Liquid Chromatography (HPLC) conditions.
   a. Set column temperature to 30°C.
   b. Set eluent flow rate to 0.2 mL min−1.
   c. Set injection volume to 10 μL.
   d. Prepare mobile phases.
   e. Degas/sonicate mobile phases as described by the previous steps.
   f. Load mobile phases into HPLC.

Note: Mobile phase A: 0.1% (v/v) formic acid and 5 mmol/L ammonium-formate in LC–MS grade water. Mobile phase B: LC-MS grade methanol.

g. Set gradient elution program.
   i. Set a linear increase of mobile phase B from 10% to 60% between 0–3 min.
   ii. Set a linear increase of mobile phase B from 60% to 100% between 3–8 min.
   iii. Wash the column with 100% mobile phase B between 8–11 min.
   iv. Set a linear decrease of mobile phase B from 100% to 10% between 11–12 min.
   v. Hold mobile phase B at 10% for re-equilibration between 12–15 min.

Note: By the original study, we used a Prominence-type HPLC system (Shimadzu, Kyoto, Japan) with a Kinetex XB-C18 (100 mm × 2.1 mm, 2.6 μm) HPLC column.

6. Set Mass Spectrometry (MS) conditions.
   a. Use the same settings as described by the previous steps.

Note: By the original study, a LCMS-2020 (Shimadzu, Kyoto, Japan) single-mass spectrometer with an ESI ion source was used.

Preparations for measuring serum glyphosate

Timing: 1 h 30 min

7. Find specific instructions for derivatizing the standards, controls, and samples in the Test Preparation section of the kit's user guide.

Note: The guide can be found at here.

Preparations for DNA extraction and metagenomic library preparation

Timing: 40 min

8. Prepare the following consumables and tools.
   a. NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) or an equivalent kit to the in-house Novogene library construction workflow.
   b. Dual-index Illumina-compatible adapters (oligonucleotides).
   c. AMPure XP beads (or equivalent magnetic bead–based cleanup system).
   d. 10 mM Tris-HCl, pH 8.5 (or kit-recommended elution buffer).
   e. Nuclease-free water.
   f. High-fidelity PCR master mix (if not included in the kit).
   g. Qubit 4.0 Fluorometer + Qubit HS dsDNA Assay Kit.
   h. Agilent Bioanalyzer (or TapeStation/Fragment Analyzer) with High Sensitivity DNA kit.
   i. Fragmentation device (e.g., Covaris) or enzymatic fragmentation mix (if included in the kit).
   j. Thermal cycler.

Note: By the original study, a BIOER LifeECO Thermal Cycler 96-well PCR Instrument with a Gradient was used.

   k. Magnetic stand for 1.5/2.0 mL tubes or PCR plates.
9. Thaw all kit components on ice, mix by gentle inversion and briefly spin down.
10. Equilibrate AMPure XP beads to room temperature for at least 30 min before use.

Hardware

We used an Ubuntu-based system equipped with an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz for computation. The machine features a dual-socket architecture with 64 CPU cores (32 physical cores, 2 threads per core) on an x86 64 platform.

The system is configured with 1.0 TiB of RAM and with a 27 TB storage drive. For comparable workloads, at least 256 GB of RAM and a 5 TB hard drive are recommended. For more samples or higher sequencing depth per sample, high-performance computing (HPC) servers should be considered.

Software and datasets

11. Download and install Conda and R as per the software guidelines based on the user's computer operating system.

    Note: This protocol was run on a Linux (Ubuntu) operating system.

12. Download and install relevant software for bioinformatic analysis as listed in the key resources table (KRT). The versions of the software in the protocol are also listed in the key resources table.
13. Download the required datasets for the taxonomic classification.

    Note: These can be found at https://ftp.ncbi.nlm.nih.gov/blast/db/ or from https://benlangmead.github.io/aws-indexes/k2.

    Note: By the original publication, the nt database was used for taxonomic classification.

14. Download the Comprehensive Antibiotic Resistance Database (CARD) using the Resistance Gene identifier (RGI) for the resistome analysis.

    Note: RGI can be found here.

15. Download the Genome Taxonomy Database.

    Note: This can be found at https://ecogenomics.github.io/GTDBTk/installing/index.html, and https://gtdb.ecogenomic.org/downloads for the analysis of metagenome-assembled genomes.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
acetonitrile	Sigma-Aldrich	Cat#1.00029
water	Sigma-Aldrich	Cat#1.15333
methanol	Sigma-Aldrich	Cat#1.06035
ammonium formate	Carl Roth	Cat# 5093.2
MgSO4	Carl Roth	Cat# 0682.3
formic acid	Sigma-Aldrich	Cat#5.33002
sodium chloride	Sigma-Aldrich	Cat#71387
PBS	Sigma-Aldrich	Cat#P2272
Aflatoxin B1	Sigma-Aldrich	Cat#A6636
DON	Sigma-Aldrich	Cat#34124
ZEA	Sigma-Aldrich	Cat#34126

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
acetic acid	Sigma-Aldrich	Cat#45754
ethyl acetate	Sigma-Aldrich	Cat#319902
agarose	Sigma-Aldrich	Cat#A9539-25G
TAE	Sigma-Aldrich	Cat# 544797
Critical commercial assays		
ABRAXIS® Glyphosate Plate ELISA Kit	Gold Standard Diagnostics	Cat# PN 500205
DNeasy® PowerSoil® Pro Kit	Qiagen	Cat# 47014
High Sensitivity DNA kit	Agilent	
Qubit® HS dsDNA Assay Kit	Thermo Fisher Scientific	Cat#Q32851
Deposited data		
Raw short reads	Tóth et al.[1]	NCBI: PRJNA1091038
Oligonucleotides		
Adapter 5: 5-AGATCGGAAGAGCGTCGTGTAG GGAAAGAGTGTAGATCTCGGTG GTCGCCGTATCATT-3	Illumina	N/A
Adapter 3: 5-GATCGGAAGAGCACACGTCTGAACT CCAGTCACGGATGACTATCTCGTATG CCGTCTTCTGCTTG-3	Illumina	N/A
Software and algorithms		
Bioconda	Grüning et al.[2]	https://bioconda.github.io/
BLAST	Altschul et al.[3]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Bowtie2 (v2.5.3)	Langmead et al.[4]	https://github.com/BenLangmead/bowtie2
CheckM2 (v1.0.1)	Chklovski et al.[5]	https://github.com/chklovski/CheckM2
Comprehensive Antibiotic Resistance Database (CARD, v.3.2.9)	McArthur et al.[6]	https://card.mcmaster.ca/
DAS Tool (v1.1.6)	Sieber et al.[7]	https://github.com/cmks/DAS_Tool
DESeq2 (v1.49.4)	Love et al.[8]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
FastQC (v0.12.1)	Andrews et al.[9]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GTDB-Tk (v2.3.2)	Chaumeil et al.[5]	https://github.com/Ecogenomics/GTDBTk
Kraken2 (v2.1.3)	Wood et al.[10]	https://github.com/DerrickWood/kraken2
MaxBin2 (v2.2.7)	Wu et al.[11]	https://nf-co.re/modules/maxbin2
MEGAHIT (v1.2.9)	Li et al.[12]	https://github.com/voutcn/megahit
MetaBAT2 (v2.12.1)	Kang et al.[13]	https://bioconda.github.io/recipes/metabat2/README.html
MetaDecoder(v1.0.18)	Liu et al.[14]	https://github.com/liu-congcong/MetaDecoder
microbiome (v1.16.0)	Lahti et al.[15]	https://www.bioconductor.org/packages/release/bioc/html/microbiome.html
MultiQC (v1.14)	Ewels et al.[16]	https://github.com/MultiQC/MultiQC
PEAR (v0.9.11)	Zhang et al.[17]	https://cme.h-its.org/exelixis/web/software/pear/doc.html
PGAP (v2024-04-27.build7426)	Tatusova et al.[18]	https://github.com/ncbi/pgap
phyloseq (v1.38.0)	McMurdie et al.[19]	https://www.bioconductor.org/packages/release/bioc/html/phyloseq.html
Prodigal (v2.6.3)	Hyatt et al.[20]	https://github.com/hyattpd/Prodigal
Prokka (v1.14.6)	Seemann et al.[21]	https://github.com/tseemann/prokka
R (v4.1.2)	R Core Team[22]	https://www.r-project.org/
Resistance Gene Identifier (RGI, v6.0.3) (with Diamond)	Buchfink et al.[23]	https://github.com/arpcard/rgi
SemiBin2 (v2.0.2)	Pan et al.[24]	https://github.com/BigDataBiology/SemiBin
TrimGalore (v.0.6.7)	N/A	https://github.com/FelixKrueger/TrimGalore
vegan (v2.6-8)	Oksanen et al.[25]	https://cran.r-project.org/web/packages/vegan/index.html
VSEARCH (v2.18.0)	Rognes et al.[26]	https://github.com/torognes/vsearch

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Agilent Bioanalyzer	Agilent Technologies	https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-instrument/2100-bioanalyzer-instrument-228250
Dual-Action Shaker KL 2	Edmund Bühler	https://www.edmund-buehler.de/en/laboratory-equipment/shakers-up-to-15-kg/dual-action-shaker-kl-2
Kinetex XB-C18 (100 mm × 2.1 mm, 2.6 μm) HPLC column	Kinetex	Cat#00D-4496-AN
LCMS-2020	Shimadzu	https://www.ssi.shimadzu.com/products/liquid-chromatograph-mass-spectrometry/single-quadrupole-lc-ms/lcms-2020/index.html
LifeECO Thermal Cycler 96-well PCR	BIOER Technology	https://www.bulldog-bio.com/wp-content/uploads/2018/05/lifeEco_user_manual.pdf
MagNA Lyser Instrument	Roche Applied Sciences	Cat#03358976001
Amicon filter	Millipore	Cat#UFC9003
Genius NM32LA nitrogen generators	Peak Scirntific	https://www.peakscientific.com/products/nitrogen/genius-nm32la-nitrogen-generator/
Illumina NovaSeq 6000	Illumina	https://www.illumina.com/systems/sequencing-platforms/novaseq.html
Prominence-type HPLC system	Shimadzu	https://www.shimadzu.com/an/products/liquid-chromatography/hplcuhplc/i-series/index.html
Qubit 4.0 Fluorometer	Thermo Fisher Scientific	https://www.thermofisher.com/hu/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/models/qubit-4.html

STEP-BY-STEP METHOD DETAILS

Sample collection and handling

Timing: 1 h per sample (after hunting)

In this section, we summarize the sample collection and handling steps used by the 15 sampled animals in the original study.

1. Hunt animals for sampling and select individuals appearing clinically healthy for inclusion in the study.

   Note: In the original research project, 20 animals were hunted and 15 were found healthy and eligible for further steps.

2. Dissect, eviscerate and weigh each animal after hunting.
   a. Perform the pathological dissection of each individual focusing on organs that might show alterations due to toxin exposure (liver, kidneys, intestinal tract, reproductive organs).
   b. Determine the age of each individual based on tooth wear and body and skull size.
   c. Determine the body condition score (BCS) of each individual based on kidney-fat index.[27]

   Note: See Figure 1 for two extremes in kidney sizes. The kidney on the left is from an animal with poor BCS and on the right from an animal with excellent BCS.

   d. Examine the ovaries and uterus to determine whether the individuals pregnant or non-pregnant.
   e. Assess the body condition score based on the kidney-fat index.[27]
   f. Eviscerate the fallow deer.
   g. Weigh each torso (body without the head, neck, limbs and tail).

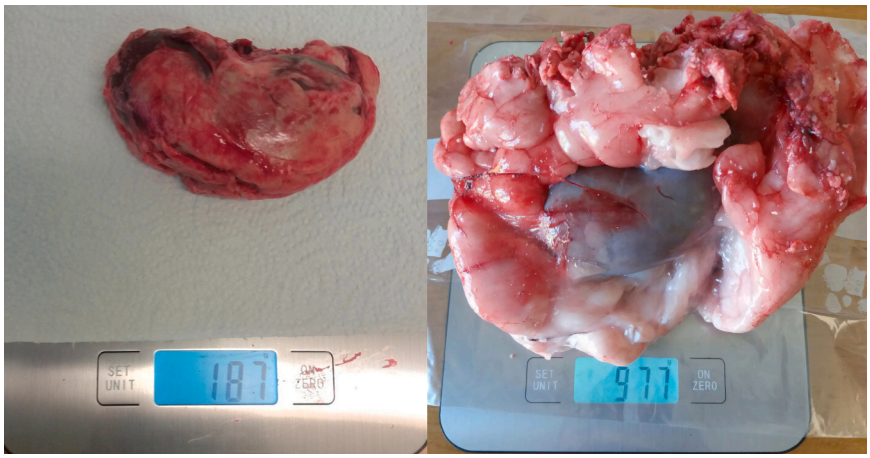


Figure 1. Kidneys from the sampled fallow deer
Body condition is associated with the amount of fat accumulated around the kidneys.[27]

3. Obtain large intestinal content samples.
   a. After hunting and evisceration, obtain feces from fallow deer cutting the large intestine open with a sterile scalpel. Use sterile gloves and sterile sample storage options.

   ⚠ CRITICAL: Collect fecal samples as cleanly as possible using only sterile tools. The risk of contamination at this step is high, which may influence the results of the metagenome analysis.

   b. Place the fecal sample in a cooler bag or on ice immediately.
   c. After transportation to the laboratory, store samples at −70°C for subsequent mycotoxin and metagenome analysis.
4. Obtain blood samples.
   a. After hunting and evisceration, draw peripheral venous blood in a sterile serum tube from fallow deer.

   Note: Blood samples were collected from the pulmonary vein using a 5 mL sterile syringe.

   b. Let the blood clot for at least 30 min at 20°C–25°C.
   c. Centrifuge the blood at 1,500–2,000 × *g* for 15 min.
   d. Transfer the serum (clear top layer) into a clean microcentrifuge tube.
   e. After transportation to the laboratory, store at −70°C until the glyphosate extraction procedure begins.

Measurement of DON, ZEA, and FB1 mycotoxins from fecal samples

Timing: 9 h (3 h per sample)

This section describes the steps for DON, ZEA, FB1 mycotoxin quantification from frozen fecal samples using HPLC (High-Performance Liquid Chromatography) and MS (Mass Spectrometry).

5. Prepare the fecal samples for mycotoxin extraction.
   a. Thaw frozen fecal samples to 20°C–25°C.
   b. Weigh 1.00 g of fecal material into polypropylene centrifuge tubes.
6. Extract the mycotoxins from the matrix.

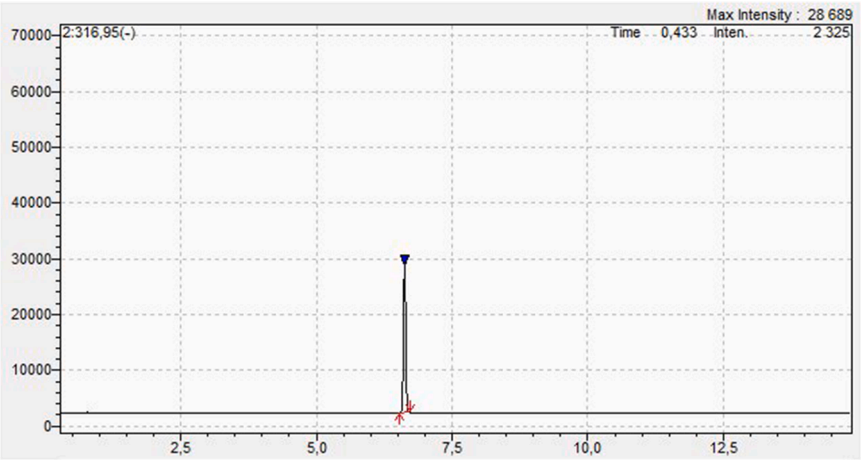


Figure 2. Chromatogram of zearalenone peak in the standard solution

a. Add 4.00 mL of extraction solvent containing the mix of acetonitrile, water, and acetic acid (50 : 50: 0.1, v/v/v).
b. Place tubes in an ultrasonic water bath for 15 min.
c. Transfer tubes to a circular shaker running at 400 rpm for 120 min.

Note: In the original study, a Dual-Action Shaker KL 2 (Edmund Bühler) was used.

d. Centrifuge the mixture at 2325 × *g* for 5 min at 20°C–25°C.
e. Transfer 1.5 mL of the supernatant to a clean polypropylene centrifuge tube.
f. Centrifuge again at 18626 × *g* for 10 min at 4°C.
g. Filter the resulting supernatant through a 0.45 μm syringe filter.
h. Transfer filtrate to vials for LC–MS analysis.

7. Perform LC-MS.
   a. Set FB1 m/z value to 722.4 (retention time: 4.62 min, positive ion mode), DON m/z value to 355.0 (retention time: 2.70 min, negative ion mode) and ZEA m/z value to 317.0 (retention time: 5.78 min, negative ion mode) for detection and identification.

Note: On Figures 2, 3, and 4 examples of chromatograms on the m/z=317 (negative) MS channel (which is used for analyzing zearalenone) are presented.

Note: On Figure 5 is an example of the total ion chromatograms (TIC, positive and negative mode) of sample extracts.

   b. Quantify the mycotoxins using the external standard calibrations with a linear range of 0.004–4 mg/kg.
   c. Set the LOD for FB1, DON, and ZEA to 0.013, 0.033, and 0.003 mg/kg, respectively.
   d. Set the LOQ for FB1, DON, and ZEA to 0.040, 0.098, and 0.009 mg/kg, respectively.

Note: LOD and LOQ were calculated by the software, using the signal-to-noise ratio (S/N) of sample chromatograms. LOD: S/N=3.3; LOQ: S/N=10.

Note: The values for m/z, LOD and LOQ might be slightly different in other experiments using different instruments, softwares, adducts, ionization modes or cone voltages or due to differences in calibration.

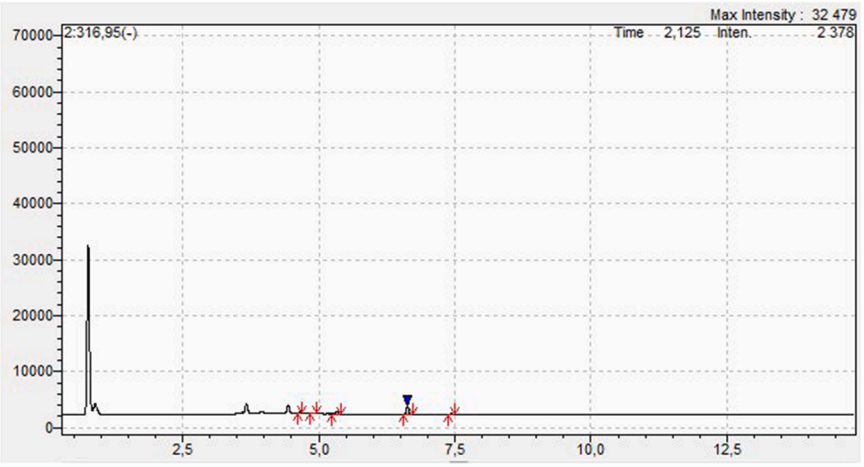


Figure 3. Chromatogram of zearalenone peak in the sample extract

Note: Although batch processing reduces the average preparation time per sample compared to single-sample analysis, the total processing time naturally increases with the number of samples. Individual handling steps, such as weighing, adding extraction solvents, syringe filtration, and sequential LC-MS analysis, inevitably add extra time to the overall process.

Measurement of aflatoxin B1 from fecal samples

Timing: 4 h 15 min (45 min per sample)

This section describes the steps for aflatoxin B1 quantification from frozen fecal samples using HPLC (High-Performance Liquid Chromatography) and MS (Mass Spectrometry).

8. Prepare the fecal samples for mycotoxin extraction.
   a. Thaw frozen fecal samples to 20°C–25°C.
   b. Weigh 1.00 g of fecal material into polypropylene centrifuge tubes.
9. Extract the mycotoxins from the matrix using a modified QuEChERS based approach.
   a. Add 4 mL of 2% formic acid in MeCN/H2O (1/1 V/V) to the sample.
   b. Mix with a horizontal vortex at 320 rpm for 15 min.

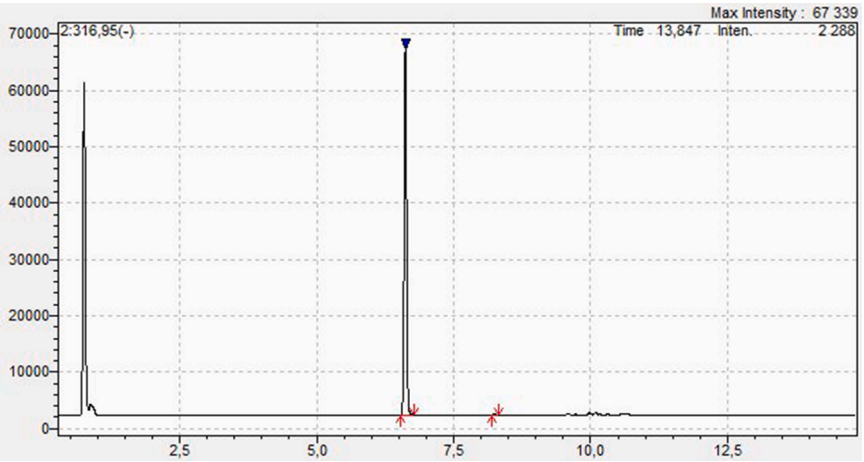


Figure 4. Chromatogram of zearalenone peak in spiked sample extract

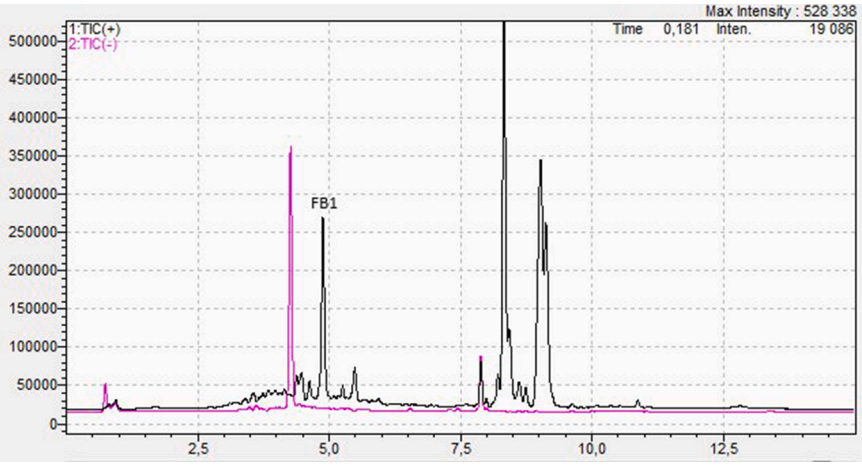


Figure 5. Total ion chromatograms (TIC, positive and negative mode) of sample extracts

Because of many interfering peaks, we had to work with the unique m/z channels to provide the proper selectivity of the method. (Although, the peak of FB1 is still apparent on the TIC chromatogram at 4.9 min.).

Note: In the original study, a Dual-Action Shaker KL 2 (Edmund Bühler) was used.

c. Add the mixture of 0.8 g MgSO4 anhydrous salt and 0.2 g NaCl.
d. Mix immediately vortexing for 60 s.
e. Centrifuge the mixture at 2325 × *g* for 5 min at 20°C–25°C.
f. Dilute 500 μL of the supernatant (containing MeCN) to 1 mL with purified water.

Note: The purified water was produced by an Adrona Crystal 7 instrument, 0.07 MΩ.

g. Filter the diluted supernatant with 0.22 μm membrane filter.

Note: The filter was produced by PVDF, La-Pha-Pack GmbH).

h. Centrifuge again at 18626 × *g* for 10 min at 4°C.
i. Filter the resulting supernatant through a 0.45 μm syringe filter.
j. Transfer an aliquot for LC–MS analysis.

10. Perform HPLC.
11. Perform MS.
    a. Set the target ion m/z value to 313.0 (retention time: 7.30 min, positive ion mode) for quantification.
    b. Set confirmatory ion m/z values to 350.9 and 663.1 for verification of identity.

Note: The values for m/z might be slightly different in other experiments using different instruments, adducts, ionization modes or cone voltages or due to differences in calibration.

Note: Although batch processing reduces the average preparation time per sample compared to single-sample analysis, the total processing time naturally increases with the number of samples. Individual handling steps, such as weighing, adding extraction solvents, syringe filtration, and sequential LC-MS analysis, inevitably add extra time to the overall process.

Measurement of serum glyphosate levels

Timing: 2 h 30 min (2 h per sample)

This section describes the steps for glyphosate quantification from serum samples using a commercial kit.

12. Prepare the serum samples for glyphosate analysis.
    a. Filter 500 μL of each sample using a centrifugal filter unit.

    Note: In the original study, a Millipore Amicon centrifugal filter unit was used.

    b. Centrifuge the samples at 8,000 × *g* for 15 min to separate the supernatant from any solid particles or debris.
    c. Transfer 300 μL of the supernatant to labelled microcentrifuge tubes.
    d. Add 200 μL ethyl acetate to the tubes.
    e. Vortex the tubes for 30 s to ensure thorough mixing of components.
    f. Centrifuge the tubes for 3 min at 8,000 × *g* to separate the different phases in the samples.
    g. Transfer the bottom aqueous phase containing glyphosate to a new labelled microcentrifuge tube.
    h. Use the extracted aqueous phase for the glyphosate analysis.
13. Use the ABRAXIS Glyphosate Plate ELISA Kit (PN 500205, Gold Standard Diagnostics, Warminster, US) according to the manufacturer's instructions.

    Note: Manufacturer's instructions can be found here.

    Note: The glyphosate-spiked fallow deer serum sample recovery rate was between 71.98%–86.96% in the original study.

    Note: Although batch processing reduces the average preparation time per sample compared to single-sample analysis, the total processing time naturally increases with the number of samples.

DNA extraction, library preparation, and sequencing

Timing: 4 days 10 h (4 days 4 h per sample)

This section describes the steps for DNA extraction, library preparation and shotgun metagenomic sequencing from frozen fecal samples using an Illumina sequencer for paired-end short reads.

14. Extract DNA from each individual sample (without pooling) using the DNeasy PowerSoil Pro Kit (Qiagen, Germany) following the manufacturer's instructions with minor modifications.
    IMPORTANT: If the fecal sample is dry, pre-hydrate and resuspend.

    Note: Alternative DNA extractions methods and kits can be used provided that they yield high-quality, inhibitor-free DNA compatible with Illumina library preparation.

    a. Add 1 mL of 1× PBS to 0.3 g of fecal sample in a 2-mL microcentrifuge tube.
    b. Vortex vigorously for 3 min or until the sample fully suspends.
    c. Centrifuge the mixture at 100 × *g* for 30 s to sediment coarse particles.
    d. Transfer 750 μL of the supernatant to a new tube.
    e. Centrifuge the supernatant at 21,000 × *g* for 5 min to pellet microbial cells.
    f. Discard the supernatant.
    g. Resuspend the pellet in 800 μL of Solution CD1.
    h. Transfer the suspension into PowerBead Pro Tubes (Qiagen, Germany).
    i. Incubate the mixture at 65°C for 10 min to facilitate chemical lysis.
    j. Perform two lysis cycles at 6,000 rpm for 30 s each.

Optional: Include brief cooling on ice between cycles if needed.

Note: Use a MagNA Lyser Instrument (Roche Applied Sciences, Germany) for mechanical lysis.

k. Add 70 μL of Solution C6.
l. Incubate at 20°C–25°C for 5 min, and centrifuge according to the manufacturer's instructions.

15. Quantify DNA concentrations using the Qubit HS dsDNA Assay Kit on a Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA).

Note: Alternative fluorometric or spectrophotometric quantification methods can be used, but Qubit-based assays are recommended for accurate low-concentration measurements.

Note: If needed, see troubleshooting problem 1.

16. Assess DNA integrity and purity by agarose gel electrophoresis (0.8%–1.0% agarose) and/or spectrophotometric ratios (OD260/280, OD260/230).

Note: Alternative integrity and purity assessment methods can be used (e.g., TapeStation, Fragment Analyzer).

⚠ CRITICAL: Prepare High-quality genomic DNA (OD260/280 = 1.8–2.0; 10 ng/μL; 100–500 ng per sample).

Note: Ensure that each DNA sample has correct purity (OD260/280 = 1.8–2.0), sufficient concentration (10 ng/μL), at least 100–500 ng total DNA available per library.

Note: If needed, see troubleshooting problem 2, 3, and 4.

17. Perform the sequencing library preparation.
    a. Fragment the DNA.
        i. Aliquot 100–500 ng of genomic DNA per sample into a low-bind tube.
        ii. Adjust the volume to the kit-recommended value (e.g., 25–50 μL) with nuclease-free water or fragmentation buffer.
        iii. Fragment the DNA to an average insert size of 350 bp by enzymatic fragmentation.

Note: This includes NEBNext Ultra II FS or a similar kits.

        iv. Add Fragmentation Buffer and Fragmentation Enzyme Mix according to the manufacturer's instructions.
        v. Incubate at 37°C for 10–15 min to obtain 350 bp fragments.
        vi. Inactivate the enzymes at 65°C for 30 min, or as recommended by the kit.

Optional: Apply mechanical (e.g., Covaris) fragmentation as follows: Transfer the DNA to Covaris microtubes in the appropriate buffer and volume. Fragment using instrument settings optimized for 350 bp inserts, according to the manufacturer. Briefly spin down the tubes and keep on ice.

Note: The exact temperature and timing should follow the chosen kit's protocol. Do not over-fragment the DNA.

Note: If needed, see troubleshooting problem 5.

    b. Perform end repair and A-tailing.

i. Combine the fragmented DNA with the End Repair/A-Tailing reagents from the library prep kit.

Note: An example is: fragmented DNA: up to 60 μL, end repair / A-Tailing Reaction Buffer, end repair / A-Tailing Enzyme Mix, nuclease-free water to the recommended final volume (e.g., 60–70 μL).

ii. Mix gently by pipetting.
iii. Briefly spin down.
iv. Incubate in a thermal cycler using the conditions specified by the kit.

Note: E.g., 20°C for 30 min, then 65°C for 30 min.

v. Briefly spin down and keep the reactions on ice until adapter ligation.

c. Perform the adapter ligation.
   i. Prepare the Illumina-compatible adapter mix (dual-index adapters) at the concentration recommended by the kit.
   ii. Set up the adapter ligation reaction.
      (1) Add the following to each end-prepped DNA sample: end-prepped DNA (from step 17/b), ligation Master Mix (from the kit), indexed adapter mix, nuclease-free water to the final reaction volume.
   iii. Mix thoroughly, briefly spin, and incubate at 20°C for 15–30 min (or as specified by the kit).

Note: In the original study, Illumina-compatible adapters equivalent to the following sequences were used: Forward adapter: 5-AATGATACGGCGACCACCGAGATCTACAC[i5*] ACACTCTTTCCCTAC ACGACGCTCTTCCGATCT-3, Reverse adapter: 5-GATCGGAAGAGC ACACGTCTGAACTCCAGTCAC[i7*]ATCTCGTATGCCGTCTTCTGCTTG-3.

Note: If using adapters that require USER treatment, add USER Enzyme as recommended and incubate further (e.g., 37°C for 15 min).

d. Perform size selection and cleanup (using AMPure XP).
   i. Use AMPure XP beads according to the kit's guidelines.

Note: Use a bead-to-sample ratio that enriches fragments in the 350–450 bp insert range (e.g., 0.6×–0.8× stepwise selection or a single 0.8×–0.9× cleanup).

ii. Perform a single-step bead cleanup.
iii. Add 0.8× volume of AMPure XP beads to the ligation reaction.
iv. Mix thoroughly and incubate at 20°C–25°C for 5–10 min.
v. Place the tube on a magnetic stand for 5 min until the beads have fully collected.
vi. Carefully remove and discard the supernatant without disturbing the bead pellet.
vii. Wash the beads twice with 70% ethanol while on the magnet, without resuspending the pellet.
viii. Remove all residual ethanol and air-dry the beads for 5–10 min (do not overdry).
ix. Elute the DNA from the beads with 20–30 μL of 10 mM Tris-HCl, pH 8.5 (or elution buffer).
x. Gently mix the eluate.
xi. Incubate the eluate for 2–5 min.
xii. Place the eluate on the magnet.
xiii. Transfer the eluate to a clean tube.

e. Perform library amplification (index PCR).
   i. Set up the PCR amplification of adapter-ligated fragments.
   ii. Use a high-fidelity PCR master mix and index primers, according to the kit recommendations.

Note: E.g., per reaction: adapter-ligated DNA: 5–10 μL, 2× High-Fidelity PCR Master Mix, i7 and i5 index primers, nuclease-free water to the final volume (e.g., 25 μL).

iii. Amplify the libraries in a thermal cycler using a limited number of cycles.

Note: This typically requires 6–10 cycles.

Note: An example for this is: 98°C for 30 s, 6–10 cycles of: 98°C for 10 s, 60–65°C for 30 s, 72°C for 30 s, 72°C for 5 min.

iv. Hold at 4°C.

⚠ CRITICAL: Use the minimum number of cycles sufficient to obtain adequate yield and avoid over-amplification.

v. Clean up the PCR products using AMPure XP beads (e.g., 0.8–1.0× ratio) as described in step 17.
vi. Elute in 20–30 μL of elution buffer.

f. Perform library quality control and pooling.
   i. Quantify each amplified library using the Qubit HS dsDNA Assay Kit.
   ii. Assess the fragment size distribution of each library using an Agilent Bioanalyzer (High Sensitivity DNA kit).

Note: Equivalent platforms can be used.

⚠ CRITICAL: Verify that the main peak corresponds to 500–600 bp total size (insert 350–450 bp plus adapters) and that there is minimal adapter/primer dimer signal.

iii. Calculate the effective molar concentration (nM) of each library from the Qubit concentration.
iv. Average fragment size.
v. Pool libraries equimolarly according to their effective concentrations to achieve the desired sequencing depth.

Note: E.g., 20 million paired-end reads per sample.

vi. Adjust the final pooled library to the concentration and volume required by the sequencing platform.

Note: For this, follow the facility's or manufacturer's denaturation and loading protocol.

Note: If needed, see troubleshooting problem 6.

18. Perform shotgun sequencing.
    a. Use an Illumina NovaSeq 6000 (Illumina, USA) or equivalent platform with a 150-bp paired-end sequencing strategy.

Note: The NovaSeq 6000 recommendation is to use Unique Dual Index (UDI) adapters to mitigate index hopping.

⚠ CRITICAL: Achieve minimum yields of 20 million reads per sample to construct metagenome-assembled genomes.

Note: Alternative library preparation kits, quantification and fragment size assessment methods, and sequencing platforms can be used.

Note: If needed, see troubleshooting problem 7.

Note: Although batch processing reduces the average preparation time per sample compared to single-sample analysis, the total processing time naturally increases with the number of samples.

Bioinformatic analysis

Timing: 2 days (10 h per sample)

This section describes the bioinformatic steps that ar required for the analysis of the bacteriome and the resistome, including the metagenome assembly of bacterial genomes.

Note: Tools that can be installed via Bioconda[2] are stored in separate directories within our pipeline. The activation and deactivation scripts of these environments are not included in the bioinformatic analysis steps.

Optional: Alternate settings for thread numbers can be used throughout the pipeline.

19. Perform steps for the taxonomic classification of reads.
    a. Perform the initial quality check of the reads using FastQC[9] and merge the results using MultiQC.[16]

```
mkdir QC
for f in *_1.gq.gz
do
fastqc -o QC -t 38 $f
done
cd QC
multiqc.
```

    b. Merge forward and reverse reads were merged with PEAR (v0.9.11).[17]

```
for f in *_1.fq.gz
do
r=${f/'_1.fq.gz'/'_2.fq.gz'}
o=${f/'_1.fq.gz'/''}
pear --threads 38 -f $f -r $r -o $o
done
```

Note: Create a merged FASTQ file by concatenating the assembled reads and the unassembled forward and reverse FASTQ files.

c. Perform the quality-based filtering and trimming of the merged reads with TrimGalore (v.0.6.7, https://github.com/FelixKrueger/TrimGalore), setting 20 as a quality threshold and a minimal length of 50 bp.

```
for f in *_merged.fastq
do
trim_galore --cores 8 \
--output_dir trimmed \
--quality 20 \
--length 50 \
--dont_gzip \
$f
done
```

d. Repeat quality check to validate the results of quality-based trimming and filtering.
e. Dereplicate the trimmed reads with VSEARCH (v2.18.0).[26]

```
for f in *.fq
do
de=${f/'.fq'/'_derep.fa'}
vsearch --threads 38 \
--derep_fulllength $f \
--strand plus \
--output $de \
--sizeout \
--uc $de.uc \
--fasta_width 0
done
```

f. Perform the taxonomic classification using Kraken2 (v2.1.3)[10] using the nt Database inclusive of GenBank,[28] RefSeq[29] TPA[30] and PDB.[31]

```
for f in *_derep.fa
do
rpt=${f/'.fa'/'_nt.rpt'} out=${rpt/'.rpt'/'.kraken'}
kraken2 --threads 40 --confidence 0.5 -db $db --report $rpt $f > $out
done
kraken-biom *.rpt --fmt json -o fdeer_read.biom
```

⚠ CRITICAL: Use the –confidence 0.5 parameter for taxon assignment to obtain more precise species-level hits.

g. Manage the taxon classification data in R (v4.1.2)[22] using functions of the packages phyloseq (v1.38.0)[19] and microbiome (v1.16.0).[15]

20. Perform steps for contig assembly and the taxonomic classification of contigs.
    a. Perform the quality-based filtering and trimming of the merged reads with TrimGalore (v.0.6.7, https://github.com/FelixKrueger/TrimGalore), setting 20 as a quality threshold and a minimal length of 50 bp.
    b. Assemble the trimmed and filtered reads to contigs by MEGAHIT (v1.2.9)[12] using default settings.

```
for f in *_val_1.fq

do

r=${f/'1_val_1'/'2_val_2'}

o=${f/'_val_1.fq'/''}

d="megahit_"$o

megahit -t 40 -1 $f -2 $r -o $d

done
```

Note: Create a directory for the assembled contigs called 'contigs'.

    c. Perform the taxonomic classification using Kraken2 (v2.1.3)[10] using the nt Database inclusive of GenBank,[28] RefSeq,[29] TPA[30] and PDB.[31]

```
for f in *_final.contigs.fa

do

rpt=${f/'.fa'/'_nt.rpt'}

out=${rpt/'.rpt'/'.kraken'}

kraken2 --threads 40 --confidence 0.5 -db $db --report $rpt $f > $out

done

kraken-biom *.rpt --fmt json -o fdeer_contig.biom
```

⚠ CRITICAL: Use the –confidence 0.5 parameter for taxon assignment to obtain more precise species-level hits.

Note: If needed, see troubleshooting problem 8 and 9.

21. Perform further steps for antimicrobial resistance gene (ARG) analysis.
    a. Gather all possible open reading frames (ORFs) with Prodigal (v2.6.3)[20] from the contigs.

```
find . -name '*.fa' | parallel -j 15 prodigal -q -m -n -p meta -i {} -a {.}'_prt.fsa' -d {.}'_
nuc.fsa' -o {.}'.draft' -s {.}'.tab'
```

⚠ CRITICAL: Remove asterisks (*) from all prt.fsa files that are stop codons inserted by Prodigal.

b. Align the protein-translated ORFs to the sequences of the Comprehensive Antibiotic Resistance Database (CARD, v.3.2.9)[6,32] by Resistance Gene Identifier (RGI, v6.0.3) with Diamond.[23]

```
rgi load \
--card_json /data/dbs/broadstreet-v3.2.9/card.json \
--local
mkdir arg_res
for f in *_prt.fsa
do
rgi main --input_sequence $f \
--input_type protein \
--output_file 'arg_res/'${f/'_prt.fsa'/'_res'} \
--local \
--include_nudge \
--clean \
-n 20
echo $f
done
```

Optional: Filter the ORFs classified as perfect or strict with 90% identity and 90% coverage.

Optional: Based on the methods of Hendriksen et al.,[33] express ARG abundance as fragments per kilobase per million fragments (FPKM) of contigs containing ARGs.

Optional: Include or exclude nudged hits according to the study objectives.

i. Calculate as follows: for the ith contig, FPKMi = qi/(li × Q) × 106, where qi is the number of reads that mapped to the contig, li is the length of contig and Q is the total number of mapped reads.
ii. Calculate q values by aligning all trimmed and filtered reads to the contigs by Bowtie (v2.5.3) with the parameter of –very-sensitive-local.[4]

```
for f in *_1_val_1.fq
do
u=$(cut -d '_' -f 1 <<< $f)
d=megahit_$u
bowtie2-build --threads 40 $d/final.contigs.fa $d/final.contigs
bowtie2 --very-sensitive-local -p 40 -x $d/final.contigs -1 $f -2 ${f/'_1_val_1.fq'/'_2_val_
2.fq'} | samtools view --threads 40 -Sb $d/final.contigs.sam | samtools sort --threads 40 > $d/
final.contigs.bam
```

```
samtools index $d/final.contigs.bam

done
```

Note: After this step, all essential files are deposited in the corresponding megahit_* repository.

c. All data management procedures, analyses and plotting were performed in R environment (v4.1.2).[22]

22. Construct metagenome-assembled bacterial genomes (MAGs).

Note: Before binning, set the directory where your megahit_* repositories.

```
r=/data/fdeer/analysis
```

a. Perform binning using the following binning tools: SemiBin2 (v2.0.2),[24] MetaBAT2 (v2.12.1),[13] MaxBin2 (v2.2.7)[11] and MetaDecoder(v1.0.18).[14]

```
cd $r

for f in megahit_*/final.contigs.fa

do

d=$r/${f/'/final.contigs.fa'/''}

cd $d

SemiBin2 single_easy_bin -i final.contigs.fa -b final.contigs.bam -o semibin

done

cd $r

for f in megahit_*/final.contigs.fa

do

d=$r/${f/'/final.contigs.fa'/''}

cd $d

docker run -u `id -u`:`id -g` \

-w /home/data \

-v $d:/home/data metabat \

runMetaBat.sh -t 20 \

/home/data/final.contigs.fa \

/home/data/final.contigs.bam

mv final.contigs.fa.metabat-* metabat

for b in metabat/*.fa

do

mv $b ${b/'bin'/'metabat'}

done

done
```

```
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/’/final.contigs.fa’/’’}
cd $d
cut -f1,3 final.contigs.fa.depth.txt > final.contigs.abund
mkdir maxbin2
/./../MaxBin-2.2.7/run_MaxBin.pl -contig final.contigs.fa -abund final.contigs.abund -out
maxbin2/maxbin
done
```

Note: /../../../ refers to the installation directory of MaxBin2.

```
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/’/final.contigs.fa’/’’}
cd $d
samtools view -h --threads 40 final.contigs.bam > final.contigs.sam
mkdir metadecoder
metadecoder coverage -s final.contigs.sam -o metadecoder/metadecoder.COVERAGE
metadecoder seed --threads 50 -f final.contigs.fa -o metadecoder/metadecoder.SEED
metadecoder cluster -f final.contigs.fa -c metadecoder/metadecoder.COVERAGE
-s metadecoder/metadecoder.SEED -o metadecoder/metadecoder
rm final.contigs.sam
done
```

Note: Metabat refers to the Docker image name, ’/home/data’ to the working directory inside container, and metabat to the output directory.

Note: Alternative and/or further binning tools can be used.

b. Optimize the bins by DAS Tool (v1.1.6).[7]

```
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/’/final.contigs.fa’/’’}
cd $d
```

```
/../../../bin/Fasta_to_Contig2Bin.sh -i ./maxbin2 -e fasta > bins_maxbin2.tsv
/../../../bin/Fasta_to_Contig2Bin.sh -i ./metabat -e fa > bins_metabat.tsv
/../../../bin/Fasta_to_Contig2Bin.sh -i ./metadecoder -e fasta > bins_metadecoder.tsv
gunzip semibin/output_bins/*.gz
/../../../bin/Fasta_to_Contig2Bin.sh -i ./semibin/output_bins -e fa >
bins_SemiBin2.tsv
DAS_Tool -i bins_maxbin2.tsv,bins_metabat.tsv,bins_metadecoder.tsv,
bins_SemiBin2.tsv -l maxbin2,metabat,metadecoder,SemiBin2
-c final.contigs.fa -o das_res/run1 --write_bins --write_bin_evals
--threads 20
done
```

Note: /../../../ refers to the Conda environment directory where the script Fasta_to_Contig2-Bin.sh is installed.

c. Taxonomically classify the resulting bins using GTDB-Tk (v2.3.2)[34] with the Genome Taxonomy Database (GTDB)[35] and Kraken2 (v2.1.3),[10] run on a database of complete bacterial genomes from National Center for Biotechnology Information (NCBI) RefSeq.[29]

```
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/’/final.contigs.fa’/’’}
cd $d
rm das_res/run1_DASTool_bins/unbinned.fa
gtdbtk identify --genome_dir das_res/run1_DASTool_bins/
--out_dir das_res/identify --extension fa --cpus 25
gtdbtk align --identify_dir das_res/identify --out_dir das_res/align --cpus 25
gtdbtk classify --genome_dir das_res/run1_DASTool_bins/
--align_dir das_res/align --out_dir das_res/classify --mash_db das_res/mashdb
-x fa --cpus 25
done
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/’/final.contigs.fa’/’’}
cd $d/das_res/run1_DASTool_bins
for fa in *.fa
do
```

```
kraken2 --threads 30 --db $db --report ${fa/'.fa'/'.rpt'}
--output ${fa/'.fa'/'.kraken'} $fa
done
done
```

Note: 'db' refers to the path of the directory where the database used for Kraken2 classification is stored.

Note: Alternative taxon classification tools and databases can be used.

d. Assess the quality of the genome bins using CheckM2 (v1.0.1).[28]

```
cd $r
for f in megahit_*/final.contigs.fa
do
d=$r/${f/'/final.contigs.fa'/''}
cd $d/das_res/run1_DASTool_bins
checkm2 predict --threads 20 --input ./ --output-directory bins_checkm2 -x .fa
done
```

Note: Before running this code, set the path of the directory where the database used for CheckM2 is stored.

Optional: Filter bins with less than 90% completeness.

Note: If needed, see troubleshooting problem 10.

e. Use PGAP (v2024-04-27.build7426)29 (including taxcheck-only mode) for genome annotation and further characterization of the binned, pre-classified contigs.

```
$PGAP_INPUT_DIR/pgap.py -v --cpus 36 -r --taxcheck-only -o SemiBin_61_check -g SemiBin_61.fa -s
'Escherichia coli'
```

Note: The script below is an example that should be run in the directory where the bins are stored. Set PGAP INPUT DIR, the directory where files necessary for running PGAP are stored before running the command. The –taxcheck-only flag can be disabled for genome annotation.

Note: Alternative genome annotation tools, such as Prokka[21] can also be used.

Note: Although batch processing reduces the average preparation time per sample compared to single-sample analysis, the total processing time naturally increases with the number of samples.

EXPECTED OUTCOMES

In this protocol, we have refined and assembled the results of existing toxin quantification and metagenome analysis steps for our application in the study of potential toxin-induced changes in the intestinal microbiome of herbivores. This enhancement allows for a more precise understanding of

how different levels of toxins that are often consumed by herbivores, such as various mycotoxins or herbicides affect the intestinal bacterial community composition and resistome.

In our case, higher ZEA levels could have been associated with decreased alpha diversity, whereas higher aflatoxin levels had the opposite effect. Furthermore, a potential link between mycotoxin exposure and antimicrobial resistance can be suggested if changes in the abundance of antibiotic resistance genes (ARGs) are observed in different toxin level groups. Furthermore, starting from 20 million paired end reads generated per sample, five complete bacterial genomes could have been assembled from the metagenomic data in our study. These findings highlight the complex interplay between environmental toxins, gut microbiota, and animal health. Understanding these interactions can be crucial for developing strategies to mitigate the negative effects of toxin exposure on wildlife populations.

QUANTIFICATION AND STATISTICAL ANALYSIS

Three study groups, each consisting of five fallow deer, were formed for the statistical analyses: one group with low ZEA levels (<10 ng/g), one with medium levels (10–30 ng/g), and one with high levels (>30 ng/g). ARG abundance (FPKM of contigs containing ARGs and the number of ARGs) was compared across conditions (mycotoxins, other chemical compounds, weight) using linear models.

The statistical analyses below were performed in R[11] environment, using the functions of phyloseq,[19] microbiome[15] and vegan[25] packages. Within-subject (α) diversity was assessed using the number of observed genera (richness) and the Inverse Simpson's Index (evenness). These indices were calculated from 1,000 iterations of rarefied operational taxonomic unit (OTU) tables at a sequencing depth of 1,448,710 reads, and the average value across iterations was used for each sample. α-diversity expressed by the Inverse Simpson's Index was compared between ZEA-level groups and across conditions using linear models.

```
library(phyloseq)

dat = import_biom('fdeer_read.biom', parseFunction=parse_taxonomy_greengenes)

colnames(otu_table(dat)) = gsub("-", "_", unlist(lapply(strsplit(colnames(otu_table(dat)),
"_"), "[[", 1)))

sd = data.frame(SampleID=colnames(otu_table(dat)))

rownames(sd) = sd$SampleID

sample_data(dat) = sd

taxa = 'Genus'

pseq = subset_taxa(dat, Kingdom=='Bacteria')

pseq.glom = tax_glom(pseq, taxrank=taxa, NArm=TRUE)

alpha = pseq.glom

depth = c()

i = 1

pseq.rarified = rarefy_even_depth(alpha, rngseed=i)

depth = c(depth, as.numeric(colSums(otu_table(pseq.rarified))[1]))

Observed = estimate_richness(pseq.rarified, measures='Observed')

InvSimpson = estimate_richness(pseq.rarified, measures='InvSimpson')

for(i in 2:1000){

pseq.rarified = rarefy_even_depth(alpha, rngseed=i)
```

```
depth = c(depth, as.numeric(colSums(otu_table(pseq.rarified))[1]))
Observed = cbind(Observed, estimate_richness(pseq.rarified, measures='Observed'))
InvSimpson = cbind(InvSimpson, estimate_richness(pseq.rarified, measures='InvSimpson'))
}
```

Between-subject (β) diversity was quantified using Bray–Curtis distances[36] based on the relative abundances of bacterial species. Principal coordinate analysis (PCoA) was applied to these distances to visualize variance among samples.

```
library(vegan)
library(microbiome)
ss = pseq.glom
tmp = inner_join(data.frame(sample_data(ss)), smpl %>% dplyr::rename(SampleID=id))
rownames(tmp) = tmp$SampleID
sample_data(ss) = tmp
opseq = transform_sample_counts(ss, function(x){x / sum(x)})
pseq.pcoa = ordinate(opseq, 'PCoA', 'bray')
txids = transform_sample_counts(pseq.glom, function(x){x / sum(x)}) |> psmelt() |> tibble()
|> filter(Abundance>=0.01) |> pull(OTU) |> unique()
pseq.core = pseq.glom
tax_table(pseq.core) = as(tax_table(pseq.glom), "matrix")[txids,]
```

Differences in core bacteriome abundance between groups were analyzed using a negative binomial generalized linear model implemented in the DESeq2 package[8] in R[11], following the recommendations of Weiss et al.[37]

```
library(DESeq2)
ss = pseq.core
tmp = inner_join(data.frame(sample_data(ss)), smpl %>% dplyr::rename(SampleID=id))
rownames(tmp) = tmp$SampleID
tmp$zea = substr(tmp$zea,1,1)
sample_data(ss) = tmp
ds = phyloseq_to_deseq2(ss, ~ zea_cc)
dds = DESeq(ds)
```

For multiple comparisons, an FDR-adjusted p-value below 0.05 was considered statistically significant. All statistical tests were two-sided.

LIMITATIONS

The protocol described above is well established and typically yields reliable results. However, some challenges remain, as it is designed for the evaluation of natural samples with versatile host behavior, consistency and content.

The results reflect only a transient snapshot of toxicological indicators and the associated metagenomic composition in the intestinal content of fallow deer, and therefore do not capture the long-term effects of different toxin exposure levels. This limitation could be addressed by sampling at multiple time points; however, obtaining repeated samples from the same individuals is challenging when working with wild animals. Given that the animals were hunted before samples were obtained, longitudinal sampling from the same hosts was inherently not possible in the protocol. However, repeated sampling is feasible in domestic herbivores after rectal sampling.

Furthermore, the generalizability of our findings may be limited by several factors. The original study was conducted within a single geographic region and during a specific time period. The sample size was relatively small - 15 deer divided into three groups of five, based on the ZEA levels - and all sampled animals were females of a similar age class (young to middle-aged).

Consequently, the samples originated from the same location, collected around the same date, from same-sex individuals with similar age, health status, and management background. While this homogeneity strengthens the internal consistency of the results for this specific group, it provides little insight into how these factors might influence the outcomes. Increasing the diversity of sampled individuals could address this limitation, although doing so would likely reduce the clarity with which toxin-induced effects can be distinguished.

TROUBLESHOOTING

Problem 1

Based on the quantification methods employed, the extraction steps produces low DNA yields. For example, Qubit HS typically measured less than 1–5 ng/μL of DNA, or highly variable technical replicates (Step 15).

Potential solution

A potential root cause can be pellet loss during microbial cell pelleting (21,000×g, 5 min), since the pellet can be tiny/transparent and easily aspirated. Furthermore, over-aggressive debris removal (100×g, 30 s) combined with transferring only 750 μL supernatant fewer microbial cells carried forward, suboptimal bead-beating (too short/weak), overheating during mechanical lysis, inefficient elution (too small volume) or over-drying of the membrane/beads, reducing recovery can also contribute to achieving such results. A potential solution can be to mark tube orientation and aspirate cautiously after 21,000×g, do not aspirate "to dryness" leave a minimal supernatant to protect the pellet. If loss is suspected: carry forward up to 1.0 mL supernatant (if compatible with the workflow) and/or repeat a short pelleting step and combine pellets. As for bead-beating, 2×30 s is adequate, but cooling on ice between cycles is essential to prevent heat-driven DNA damage and yield loss. For tough samples, add a short third cycle only with cooling. Furthermore, use room-temperature (or slightly warmed) elution buffer, incubate 2–5 min, and consider two-step elution (e.g., 2× μL) to improve total recovery by the elution.

Problem 2

Qubit quantification indicated acceptable DNA concentrations; however, NanoDrop measurements showed low 260/230 ratios (e.g., 0.3–1.2), and the DNA exhibited a brown coloration and increased viscosity. Alternatively, qPCR analysis yielded high Ct values, accompanied by low library yields (Step 16).

Potential solution

A possible cause can be the residual humic substances, bile salts, polysaccharides, and related stool-derived inhibitors remaining in the samples. Fixes include adding a post-extraction cleanup step, namely, one extra magnetic bead cleanup (AMPure-type) or a dedicated inhibitor removal step/kit or using qPCR-based library QC/quantification (not Qubit alone) to detect non-amplifiable libraries early.

Problem 3

Bioanalyzer/TapeStation profiles revealed dominant adapter dimer peaks (approximately 120–170 bp), characterized by a strong small-fragment signal and resulting in a low proportion of usable library molecules (Step 16).

Potential solution

Too high adapter concentration, too low input DNA or suboptimal bead cleanup/size selection might cause such anomalies. Apply stricter bead cleanup/size selection (e.g., 0.8× or two-step selection per kit guidance) or minimize PCR cycles to reduce enrichment of short artifacts to overcome them.

Problem 4

DNA was highly fragmented, indicating over-processed genomic DNA. This was evident from gel or TapeStation analysis, which showed a smear with no clear high–molecular-weight component (Step 16).

Potential solution

The root causes included overly aggressive bead-beating and/or overheating during lysis, as well as multiple freeze–thaw cycles. To prevent this, bead-beating intensity and total processing time should be reduced, with increased cooling between cycles, and samples should be thawed only once, aliquoted, and protected from repeated freeze–thaw events.

Problem 5

The insert size was incorrect, being either too short or too long. Instead of the expected 350 bp insert, the library showed peak shifts to 150–200 bp or, in some cases, 600–800 bp (Step 17/a).

Potential solution

The root causes can be traced to either enzymatic fragmentation, due to suboptimal timing, temperature, or enzyme ratio, or to incorrect Covaris settings. To address enzymatic fragmentation issues, the incubation time should be optimized and if inhibitors are suspected, mechanical shearing with Covaris is preferred. For Covaris-based fragmentation, the instrument settings should be validated on two to three representative samples using a TapeStation or Bioanalyzer before scaling up the workflow.

Problem 6

Low library yield was indicated after qPCR despite an adequate Qubit measurement (Step 17/f).

Potential solution

This might appear by qPCR results suggesting a low amount of amplifiable library and by poor clustering on the sequencer. The underlying causes often include inhibitor carry-over or inefficient adapter ligation, as well as PCR amplification issues, such as under-cycling or, conversely, over-cycling that introduces artifacts. To resolve this, post-ligation cleanup steps should be optimized to remove inhibitors and excess adapters. In addition, PCR conditions should be carefully tuned: a typical library amplification requires approximately 6–10 cycles, and if yields remain extremely low, the focus should be on improving template cleanliness rather than simply increasing the number of PCR cycles.

Problem 7

The target of 20 M reads per sample is missed due to uneven per-sample sequencing depth (Step 18).

Potential solution

This typically arises when libraries are pooled based on mass (Qubit ng/μL) without accounting for differences in fragment size, leading to molarity mismatches between samples. In addition, when qPCR is not used, the concentration of amplifiable library molecules is often misestimated. To mitigate this, libraries should be pooled by molarity (nM), calculated using both concentration and mean fragment size. For critical sequencing runs, qPCR-based library quantification (for example, KAPA or Illumina-compatible assays) should be performed, and pooling should be based on these measurements to ensure more uniform read distribution across samples.

Problem 8

The computational resources available on the system used are inadequate for downloading and processing the NCBI nt or core_nt database required for Kraken2 taxonomic classification (Step 20/c).

Potential solution

An alternative for using the NCBI nt database for taxon classification is as follows. Host reads (fallow deer deriving reads) can filtered by alignment to the *Dama dama* genome sequences with Bowtie2[4] using the very-sensitive-local setting to minimize the false positive match level.[38] Afterwards, the remaining reads can be taxonomically classified using smaller databases, such as NCBI RefSeq Bacteria.[29] Nevertheless, the use of such databases may be a source of bias and should we treated with suspicion. As a precaution, additional steps, such as setting high confidence scores or the use of BLAST[3] to reassure the results are recommended.

Problem 9

Upon examining the Kraken2 classification results, the presence of taxa that appear improbable for this sample, that would constitute an extraordinary scientific observation or that I suspect to derive from contamination was indicated (Step 20/c).

Potential solution

Several well-known reasons my underlie why implausible or impossible taxa appear in a metagenomic study. These results are normally classification artifacts and do not necessarily mean contamination for the following reasons. Firstly, Kraken2 relies on k-mer–based classification. Since k-mers are short, some of them can be shared by multiple organisms. When a k-mer appears in several organisms, Kraken assigns it to the lowest common ancestor (LCA), which may result in an unexpected genus or species being reported. It is even more common to observe such hits due to database contamination. Kraken2's official databases, including NCBI databases such as nt and even RefSeq, can contain mislabeled entries—such as misannotated genomes, contaminant sequences, or plasmids assigned to incorrect taxa. Furthermore, low read counts can sometimes be misinterpreted as real taxa, and plasmids are particularly prone to misclassification. If trimming and filtering are not performed properly, sequencing artifacts may also appear in the results.

Overall, based on the evaluation of the Kraken2 classification results, it is recommended to set a cut-off of <0.1% of total reads or <100 reads when treating hits as noise. In such cases, genome coverage for the assigned species should also be examined. If the sequencing depth is inconsistent or below 1×, and only a few reads map across the genome, the hit is likely misclassified. In these instances, the use of a secondary classifier such as MetaPhlAn,[39] Centrifuge,[40] or Bracken (for real abundance estimations)[41] is recommended, or the reads in question can be manually checked using BLAST.[3] If contamination is strongly suspected, comparisons with negative control samples should be considered. Additionally, as indicated in the protocol, the use of well-adjusted confidence values is crucial for Kraken2 taxonomic classification and can, by itself, mitigate some of the issues mentioned above.[42]

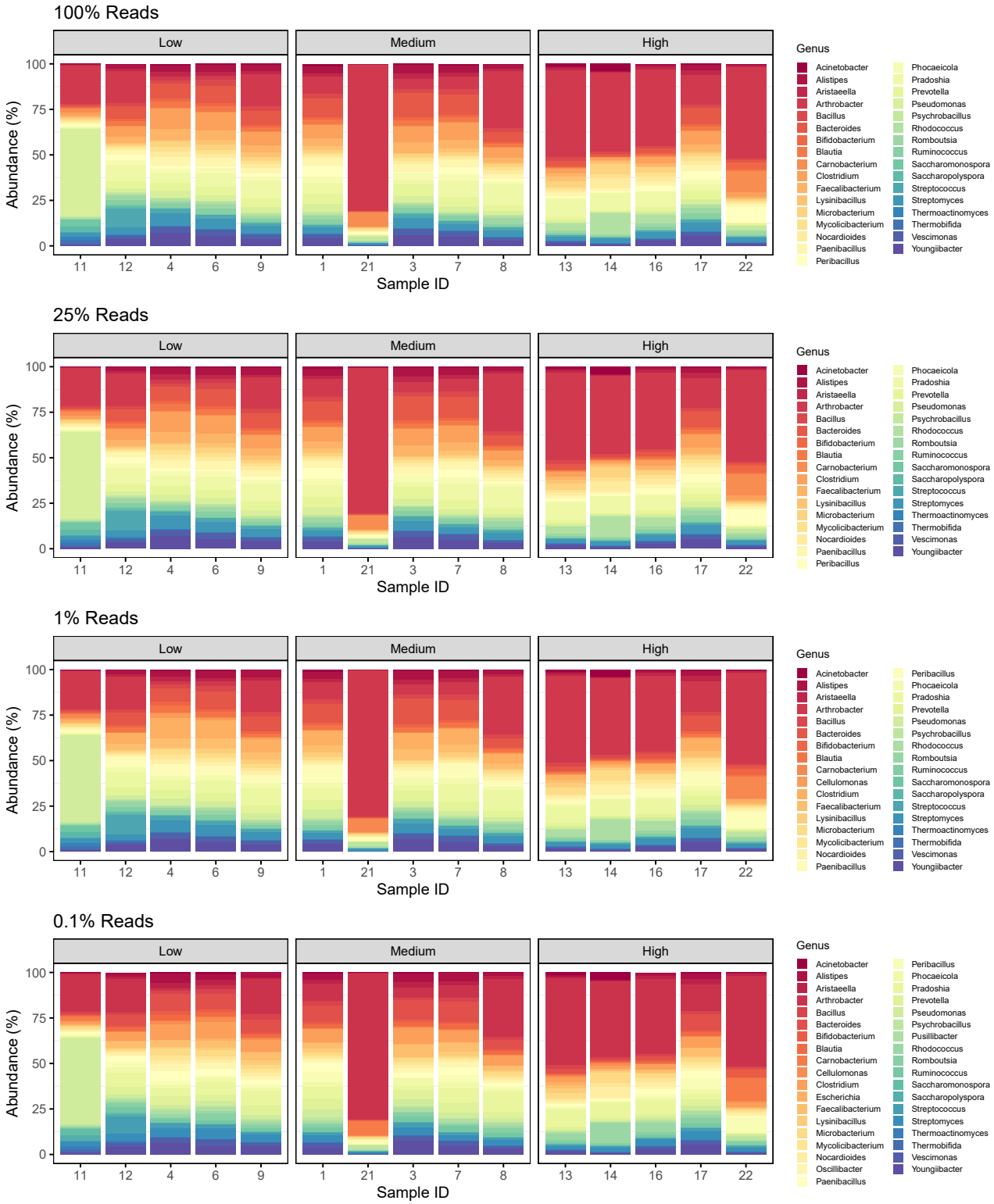


Figure 6. Effect of sequencing depth on taxonomic classification of reads

Subsampled reads represent 25% of the full read count.

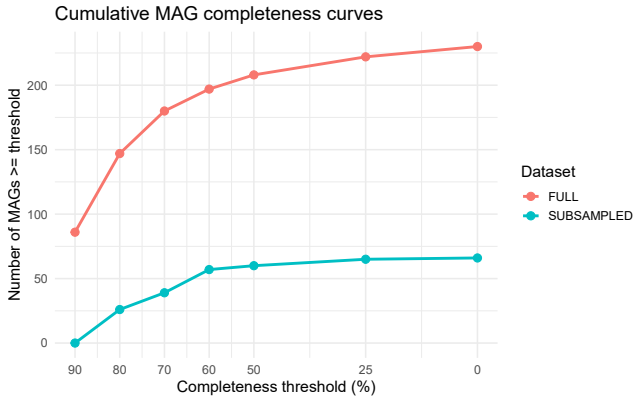


Figure 7. Effect of sequencing depth on the number and completeness of MAGs

Problem 10

While several metagenome-assembled genomes were generated, none met the required completeness thresholds (Step 22/d).

Potential solution

As indicated on Figure 6, sequencing depth does visibly affect the taxonomic classification of reads from a metagenomic sample, this if the construction of MAGs in not among the objectives, depth can be decreased to levels lower than recommended in the protocol.

However, as shown in Figure 7, higher sequencing depth has the potential to supply additional reads that would improve MAG completeness. Consequently, this study strongly recommends preserving the sequencing libraries for possible supplementary sequencing.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Adrienn Gréta Tóth (tothadrienngreta@gmail.com).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Adrienn Gréta Tóth (tothadrienngreta@gmail.com).

Materials availability

This study did not generate new materials.

Data and code availability

The raw short-read data are publicly available and accessible through PRJNA1091038 from the NCBI Sequence Read Archive (SRA). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was funded by the Hungarian National Laboratory Project (grant no. RRF-2.3.1-21-2022-00007), the Agribiotechnology and Precision Breeding for Food Security National Laboratory, and the Flagship Research Groups Programme of the Hungarian University of Agriculture and Life Sciences. Further support was received from the Flagship Research Groups Programme of the Hungarian University of Agriculture and Life Science (Flagship Research Groups 2026). The study was also supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the MILAB Artificial Intelligence National Laboratory and by the strategic research fund of the University of Veterinary Medicine Budapest (grant no. SRF-003).

AUTHOR CONTRIBUTIONS

A.G.T., N.S., S.F., and Z.S. took responsibility for the integrity of the data and the accuracy of the data analysis. A.G.T., N.S., S.F., and Z.S. conceived the study. I.L. collected the biological samples. A.S., I.L., M.P., and Z.S. performed the

laboratory processes. A.G.T. and N.S. participated in the bioinformatic analysis. A.G.T., N.S., S.Á.N., S.F., and Z.S. participated in the drafting of the manuscript. A.G.T., A.S., I.L., K.P., M.P., N.S., S.Á.N., S.F., and Z.S. critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Tóth, A.G., Nagy, S.Á., Lakatos, I., Solymosi, N., Stágel, A., Paholcsek, M., Posta, K., Gömbös, P., Ferenczi, S., and Szőke, Z. (2025). Impact of mycotoxins and glyphosate residue on the gut microbiome and resistome of European fallow deer. iScience *28*, 113539.

2. Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., and Köster, J.; Bioconda Team (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat. Methods *15*, 475–476.

3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

4. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

5. Chklovski, A., Parks, D.H., Woodcroft, B.J., and Tyson, G.W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat. Methods *20*, 1203–1212.

6. McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., et al. (2013). The Comprehensive Antibiotic Resistance Database. Antimicrob. Agents Chemother. *57*, 3348–3357.

7. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. *3*, 836–843.

8. Love, M.I., Huber, W., and Anders, S. (2014). Differential analysis of count data—the DESeq2 package. Genome Biol. *15*, 550.

9. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

10. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257.

11. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605–607.

12. Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

13. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359.

14. Liu, C.-C., Dong, S.-S., Chen, J.-B., Wang, C., Ning, P., Guo, Y., and Yang, T.-L. (2022). MetaDecoder: a novel method for clustering metagenomic contigs. Microbiome *10*, 46.

15. Lahti, L., and Shetty, S. (2018). Introduction to the microbiome R package. https://microbiome.github.io/tutorials.

16. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

17. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina paired-end read merger. Bioinformatics *30*, 614–620.

18. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., and Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. *44*, 6614–6624.

19. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217.

20. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. *11*, 119.

21. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

22. R Core Team (2025). R: A language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing). https://www.R-project.org/.

23. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

24. Pan, S., Zhao, X.-M., and Coelho, L.P. (2023). SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. Bioinformatics *39*, i21–i29.

25. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2024). vegan: Community Ecology Package, R package version 2.6.8.

26. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.

27. Flesch, J., Mulley, R., and Asher, G. (2002). Development of a body condition scoring system for farmed fallow deer (Dama dama), M. Crête, ed., pp. 20–26.

28. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). Nucleic Acids Res. *41*, D36–D42.

29. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

30. Cochrane, G., Bates, K., Apweiler, R., Tateno, Y., Mashima, J., Kosuge, T., Mizrachi, I.K., Schäfer, S., and Fetchko, M. (2006). Evidence standards in experimental and inferential INSDC third-party annotation data. OMICS *10*, 105–113.

31. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

32. Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., et al. (2017). CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. *45*, D566–D573.

33. Hendriksen, R.S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., et al. (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat. Commun. *10*, 1124.

34. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. Bioinformatics *38*, 5315–5316.

35. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. *50*, D785–D794.

36. Bray, J.R., and Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. *27*, 325–349.

37. Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome *5*, 27.

38. Czajkowski, M.D., Vance, D.P., Frese, S.A., and Casaburi, G. (2019). GenCoF: a graphical user interface to rapidly remove human genome contaminants from metagenomic datasets. Bioinformatics *35*, 2318–2319.

39. Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A.M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat. Biotechnol. *41*, 1633–1644.

40. Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.

41. Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L. (2017). Bracken: estimating species abundance in metagenomics data. PeerJ. Comput. Sci. *3*, e104.

42. Liu, Y., Ghaffari, M.H., Ma, T., and Tu, Y. (2024). Impact of database choice and confidence score on the performance of taxonomic classification using Kraken 2. aBIOTECH *5*, 465–475.