

Species Identification Report



Nick Famoso
National Park Service

Invoice number	20250702_1
Project ID	NPSNF
Email address	nicholas_famoso@nps.gov
Sequencing date	Aug 2025
Report date	Aug 2025
Technician	Ellen Liebig
Bioinformatician	Daniel Sanchez
Sequencing kit	MiniSeq Mid-output (300 cycle)
QIIME version	QIIME2 v2024.10 (amplicon)

Bat Ecology & Genetics Lab, School of Forestry, NAU, P.O. Box 15018, Flagstaff, AZ 86011
nau.edu/sff

<https://linktr.ee/speciesfromfeces>

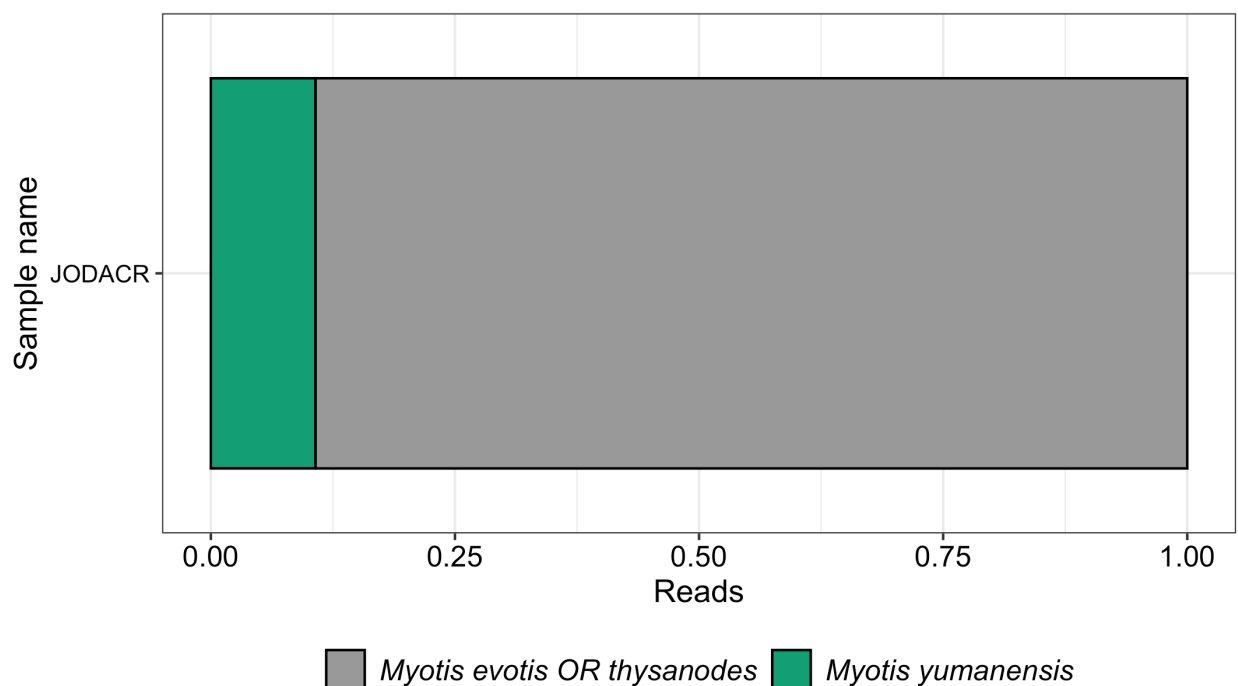
Questions? Faith.Walker@nau.edu; Daniel.Sanchez@nau.edu

Results:

We received and processed one fecal sample with the goal of detecting one or more bat species. We successfully amplified the target region allowing for detection of two bat species. More information can be found in the figures below and the associated results spreadsheet. See the key considerations section below for additional insight into interpreting the figures and this metabarcoding dataset.

None of the DNA extraction blanks nor PCR negative template controls prepared with your samples yielded taxonomic information. Our positive control consisting of eastern hemisphere bat species yielded 5/5 expected sequences and no unexpected sequences. If you suspect that a detection is surprising or could have significant management or conservation implications, we recommend re-amplification and re-sequencing to gather multiple lines of evidence. If this is the case, please feel free to contact us for further recommendations.

Per-sample bat detections



Accompanying files:

Along with a PDF of the detection barplots, we included an Excel (xlsx) spreadsheet that includes all figures, all taxonomic data, and sequencing pass and read summaries.

NPSNF_20250702_1_NFamoso_BEGLresults.xlsx
NFamoso_detection_plot.pdf

Key considerations:

Bar plot (per-sample bat detections) - The stacked bar plots included in your results package only depict samples where bats were detected, unless only taxa other than bats were detected among all samples. Samples that failed or only amplified non-bat taxa are not shown in these plots but can be found in the associated results spreadsheet. In cases where certain bat taxa in a sample have low signal (i.e., few reads) compared to taxa with much higher read numbers, we apply a log transformation to enhance the visibility of these detections. Note that *the number of sequencing reads for a species may not reflect its abundance at a sampling location*.

Heat tree plot (all detections at-a-glance) - We provide a heat tree for datasets with more than two species detections, which may include detections of other non-bat wildlife. The **heat tree plot** (Foster et al. 2017) summarizes taxonomic detections across your entire dataset, showing full lineage hierarchies, average read abundance of each detection (color fill), and the number of samples (node size) sharing that detection (i.e., frequency of occurrence). Note that because our analysis focuses on bats, we recommend approaching non-bat classifications with caution. We generally suggest cross-referencing them against relevant wildlife inventories. We are happy to take a second look at any non-bat taxa you may find interesting, however.

Taxonomic nomenclature - The taxonomic nomenclature used for classification is directly sourced from either the Barcode of Life Database or NCBI GenBank. It is important to note that nomenclature may differ among databases and could possibly contain outdated species names. Therefore, we advise considering updated or synonymous taxonomic nomenclature during your analysis of the samples.

Geographic accuracy - It is important to acknowledge that we may not always have precise geographic information for bat species and largely rely on species range maps, state/provincial records, or the available literature for inspecting the taxa detected in your samples. If we detect a bat species outside of its known range, we always double-check the sequences against other databases. If the range boundaries are in close proximity to where the sample was collected, we consider it a plausible detection. If you suspect that a detection is erroneous (e.g., contamination) or could have significant management or conservation implications, we recommend re-amplification and re-sequencing to gather multiple lines of evidence. If this is the case, please feel free to contact us for further recommendations. It is also important to

Bat Ecology & Genetics Lab, School of Forestry, NAU, P.O. Box 15018, Flagstaff, AZ 86011

3

nau.edu/sff

<https://linktr.ee/speciesfromfeces>

Questions? Faith.Walker@nau.edu; Daniel.Sanchez@nau.edu

consider whether the detection in question had amplified in more than one sample or replicate, which can lend strength to its credibility.

Sample Pass/Fail - A sample “Pass” indicates that the sample successfully yielded DNA sequences leading to a taxonomic classification. A sample “Fail” indicates that the sample failed to PCR amplify or that the sample failed to produce taxonomically informative DNA sequences. Sample dropout typically occurs when DNA contained in the samples are overly-degraded, do not contain target DNA or contain exceedingly low concentrations of target DNA, or if they contain high levels of PCR inhibitors. The state of non-tissue substrates collected from the environment may be a product of a number of variables (e.g., time, humidity, heat, preservation...) and these samples can introduce substances that inhibit PCR amplification success. We cannot control for degradation nor low/absent concentration target DNA as this is very often a feature of the sample itself. In cases of procedural errors, we always inspect samples for batch (sample-specific) or systematic errors (errors occurring across different sample batches within the same preparation). In some cases, this may only be apparent when sequence data has already been obtained and analyzed. We maintain detailed records of our experiments to evaluate whether procedural errors occurred and we will always note in the results summary whether we believe this to be the case.

OTU dataset - An operational taxonomic unit (OTU, i.e., molecular feature) in this dataset refers a distinct DNA sequence (i.e., amplicon sequence variant) and serves as the principal biodiversity unit on which our taxonomic classifications were based. Depending on the marker region, a single OTU is typically a molecular proxy for a single species or a single genus. Sometimes different OTUs can be attributed to the same species or the same OTU can be attributed to different species. We provide this dataset in the XLSX data package, along with their corresponding DNA sequences and raw taxonomic classifications. OTU data can be useful for re-aggregating your taxonomic data or for evaluating diversity in cases where certain species lack DNA references. The DNA sequences can also be used to cross-verify against different reference libraries or update classifications as new references are added in the future.

Each OTU contains two elements: a unique DNA sequence and its unique hash MD5 label (Feature ID)—an identifier encoded from the properties of the sequence itself. So every time we obtain that same exact DNA sequence among different sequencing runs or even different samples, it will always return the same 32 character Feature ID. In that sense, this allows the feature ID to be treated as a distinct taxon. The OTU dataset in your report package is referenced to your samples and the number of reads (i.e., copies) an OTU was detected among.

Methodology:

We extracted genomic DNA and amplified a short-section of cytochrome oxidase subunit I (COI) from the samples using our standard methodology (Walker et al. 2016; Walker et al. 2022). Amplified product was sequenced to obtain DNA sequences (reads) of one or more taxa per sample. Sequencing reads were computationally processed using QIIME2 (Bolyen et al. 2019). Priming regions were removed using cutadapt (Martin 2011) to isolate the 202 base pair fragment of interest. We removed low quality reads, and filtered out PCR artifacts (chimeric reads) using DADA2 (Callahan et al. 2016). To avoid low abundance variants with sequencing/PCR errors, the unique sequences were post-clustered using LULU curation (Frøslev et al. 2017). We also screened reading frames to alleviate presence of any remaining pseudogenes or other artifacts. Sequences were then classified using a naïve-Bayes machine learning classifier (Bokulich et al. 2018) that we trained against our custom reference database. References were derived from all available chiropteran COI references in the Barcode of Life Database (Ratnasingham and Hebert 2007). We retained species classifications only if they were classified with at least 90% bootstrap support. Any variants not classified using the machine learning algorithm to species were cross-referenced against the National Center for Biotechnology Information's (NCBI) GenBank database (Benson et al. 2009) using BLAST (Altschul et al. 1990). Taxa were classified using Least Common Ancestor (LCA) analysis in MEGAN v6 (Huson et al. 2007). This cross-referencing step helps to alleviate any false negative bat classifications in the naïve-Bayes's model or identify non-bat taxa that may have co-amplified.

References:

Foster ZSL, Sharpton TJ, Grünwald NJ. 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput Biol.* 13(2):e1005404. <https://doi.org/10.1371/journal.pcbi.1005404>

Walker FM et al. 2016. Species From Feces: Order-Wide Identification of Chiroptera From Guano and Other Non-Invasive Genetic Samples Russo D, editor. *PLOS ONE.* 11(9):e0162342. <https://doi.org/10.1371/journal.pone.0162342>

Walker FM et al. 2022. Endangered Nectar-Feeding Bat Detected by Environmental DNA on Flowers. *Animals.* 12(22):3075. <https://doi.org/10.3390/ani12223075>

Bolyen E et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 37(8):852–857

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>

Bat Ecology & Genetics Lab, School of Forestry, NAU, P.O. Box 15018, Flagstaff, AZ 86011
nau.edu/sff

<https://linktr.ee/speciesfromfeces>

Questions? Faith.Walker@nau.edu; Daniel.Sanchez@nau.edu

Callahan BJ et al. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 13(7):581–583. <https://doi.org/10.1038/nmeth.3869>

Frøslev TG et al. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat Commun*. 8(1):1188. <https://doi.org/10.1038/s41467-017-01312-x>

Bokulich NA et al. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*. 6(1):1–17. <https://doi.org/10.1186/s40168-018-0470-z>

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes*. 7(3):355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>

Benson DA et al. 2009. GenBank. *Nucleic Acids Res*. 37(suppl_1):D26–D31. <https://doi.org/10.1093/nar/gkn723>

Altschul SF et al. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Res*. 17(3):377–386. <https://doi.org/10.1101/gr.5969107>