WHICH VOLE IS WHICH:

DNA-BASED SPECIES IDENTIFICATION FOR WISCONSIN'S THREE MICROTUS SPECIES

by

Madeline Opie

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Biological Sciences

at

The University of Wisconsin-Milwaukee

May 2024

ABSTRACT

WHICH VOLE IS WHICH:
DNA-BASED SPECIES IDENTIFICATION FOR WISCONSIN'S THREE *MICROTUS* SPECIES

by

Madeline Opie

The University of Wisconsin-Milwaukee, 2024
Under the Supervision of Professor Emily Latch

Accurate species identification is necessary to implement conservation strategies in the wild. When traditional morphology-based species identification is challenging due to phenotypic plasticity, overlapping characteristics, or the species are otherwise cryptic, DNA-based species identification may be more suitable. Of the three species of *Microtus* in Wisconsin, two are listed as threatened at the state level. Both *M. ochrogaster* and *M. pinetorum* have stable population levels at the national level but are along the northern edge of their ranges in Wisconsin. Small and vulnerable populations of *M. ochrogaster* and *M. pinetorum* are limited to isolated patches in the southwestern portion of the state. A primary challenge in conservation efforts is the overlap in distribution and habitat with the third and more common species, the meadow vole (*M. pennsylvanicus*). The three species have similar morphology, but *M. ochrogaster* and *M. pennsylvanicus* are nearly identical morphologically. To distinguish between the three species in the field, biologists rely on overall appearance, however, small variations in an individual organism's morphology can result in misidentification in the field. We evaluated two approaches for genetic species identification of the three species: a length polymorphism in the *avrpr1a* gene and novel, system-specific DNA barcoding in the *COI* gene. We evaluated

methods using tissue and less invasive samples (fecal, whisker, and hair) that will minimize the need for handling animals in the field. A robust genetic species identification method will allow the Wisconsin Department of Natural Resources to gather accurate monitoring data and improve conservation efforts for *M. ochrogaster* and *M. pinetorum.* In addition, our approach for the development of system-specific DNA barcodes can be applied to similar systems that encounter challenges in morphology-based species identification.

To my dad, who gave me his copy of *A Sand County Almanac* when I was sixteen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BOLD	Barcode of Life Data
HY	Henterly et al. 2011
IUCN	International Union for Conservation of Nature
MIOC	*Microtus ochrogaster*
MIPI	*Microtus pinetorum*
MIPE	*Microtus pennsylvanicus*
PCR	Polymerase Chain Reaction
SGCN	Species of Greatest Conservation Need
WDNR	Wisconsin Department of Natural Resources
UWM FS	University of Milwaukee – Field Station

ACKNOWLEDGEMENTS

I would like to extend thanks and gratitude to my advisor, Dr. Emily Latch. It has been a privilege learning from you over the last three years, about much more than just wildlife genetics. I've learned to be a better scientist, teacher, writer, and human being from my time as your student. You're a rock star, simply put. You're incredible at what you do and even more so because you're able to support your students as they succeed in ways that they might not have thought were possible. I will be forever grateful that I was able to lean on you and share with you during difficult times. And although I wish it wasn't so, it was so impactful that you truly understood what my family was going through. Thank you.

I want to also thank my committee members, Dr. Jeffrey Karron and Dr. Gretchen Meyer. I was fortunate to have both of you as professors and mentors during my time at UWM. Jeff, thank you for posing the "big picture" questions, while also taking the time to assure the details of my work were in order. Your love of ecology is eminent in the way you teach and speak about the natural world. I look up to you as a scientist and a teacher. Gretchen, I'm so fortunate to have met you through my work at the field station and to have been your student in the Field Methods course, where you let me sample voles during one of our classes! It was such a fun surprise to find voles in the pit traps. I appreciated your willingness to let me pester you with questions about biology, natural history, or just life in general, as we hiked during class. Thank you, Jeff and Gretchen, for participating on my committee, sharing your knowledge from your fields of expertise, and making me a better science communicator.

To all our collaborators that helped with sample collection for this project—thank you! Thanks to Paul White and Richard Staffen of the Wisconsin Department of Natural Resources

for providing us with our initial wild Wisconsin samples; it was a pleasure working on this project with you. Thank you to Dr. Annaliese Beery and her staff at UC Berkeley for collecting samples and sending them to us. Lastly, thank you, to our museum collaborators, Dianna Krejsa, Mammal Curator at the UW Zoological Museum and Heath Garner, Curator of Collections at the Museum of Texas Tech University for your help in coordinating and sending us museum specimen samples. The samples from each of our collaborators contributed to crucial representation of the three species and ensured we had a well-rounded dataset; without all of you this work would not have been possible.

To my Latch lab mates, past and present, thank you so much for your guidance and friendship. Special thanks to Dr. Xueling Yi for always answering my questions during my first year; even though you were preparing for your defense, you always had time to work with me. Thank you to Dr. Vera De Ferran for helping me shape the story that I wanted to tell with my research. To Billie Harrison, thank you for bringing your humor and counsel, and supplying me with Grenadian chocolate: it was crucial to my success in this endeavor. Lastly, Chandika RG, I'm so proud to work with you and call you my friend. You were always there with unwavering support be it lab work, coding, writing, etc.—literally everything and anything. I know you will probably be modest about it, but I just wanted you to know that I'm so, so grateful for you and all the ways you've helped me over the past three years.

To my mom and dad, Charlene and Richard, and my brother and sister, Abe and Meredith, I can't express in words how grateful I am to have our family. We're rock solid and I feel so lucky to be loved and supported by each of you. Teetee, you're my go to, you're always there to listen and offer advice. I am grateful to have you as a godmother and friend. To my

friends in Wisconsin and beyond (Adrianne, Alex, Amber, Caterina, and Rachael) thank you so much for your support, I love you all dearly. Finally, Elly girl, you walked into my life right as I was starting my journey as a graduate student and now that I've finished, we're moving on to the next chapter together. You deserve an honorary degree for dating someone in graduate school; you were there every step of the way, and it wouldn't have been possible without you. I love you.

Funding from the Wisconsin Department of Natural Resources and the James and Dorothea Levenson Ecology & Field Biology Fellowship supported this project.

Introduction

Accurate species identification is necessary when implementing conservation strategies in the wild. Traditional morphology-based species identification can provide efficient and accurate results in many cases. However, problems can arise in morphology-based species identification when species portray phenotypic plasticity among individuals or life stages, exhibit overlapping characteristics between species, or are otherwise cryptic (Clare, et al., 2007; Tyagi, et al., 2019; Hubert, et al., 2008). Traditional methods of species identification in these situations can result in species misidentification, leading to inaccurate monitoring data. As the field has advanced, DNA-based species identification has become a reliable, cost-effective, and accessible tool. Unique sets of base-pair mutations within the sequences of one or more genes can act as a DNA barcode for species identification (Hebert, et al., 2003; Hebert and Gregory, 2005). Sequencing DNA barcodes requires only a sample of genetic material from the animal to be completed, minimizing, or even eliminating the need to handle animals in the field (Waits and Paetkau, 2005).

In 2003, Hebert, et al. proposed that the 648bp fragment of the 5' terminus of the cytochrome c oxidase subunit 1 (COI) gene be used as a standard DNA-based identification and species discovery system. The use of the mitochondrial genome is advantageous over the nuclear genome because of its lack of introns, limited exposure to recombination, and haploid inheritance in animals (DeSalle, et al., 2017; Ladoukakis and Zouros, 2016). *COI* was chosen specifically for its existing universal primers, the lack of indels, and high rate of molecular evolution relative to other mitochondrial genes (Hebert, et al., 2003). The Barcode of Life Data

(BOLD) Systems database was created to increase accessibility and availability of *COI* gene sequences and other standardized DNA barcodes (Hebert and Gregory 2005). The database is a public reference library that users can access to compare unknown specimen sequences to voucher specimen sequences (Ratnasingham and Hebert, 2007; Meiklejohn, et al., 2019).

Accuracy of DNA barcoding for species identification relies on a "barcode gap" between intra- and interspecific genetic differences, where the range of genetic variation within a species does not overlap with the range of interspecific differences between species (Meyer and Paulay, 2005). A primary concern of using DNA barcoding is that while the *COI* gene may be easily sequenced regardless of the condition of the specimen, the presence and/or magnitude of a barcode gap is inconsistent across taxa (Phillips, et al., 2022). Another critical consideration in evaluating DNA barcoding for reliable species identification is the number and representation of voucher sequences in the BOLD reference library. Lack of reference data is the most common reason for species identification errors in DNA barcoding (Becker, et al., 2011). Species without any representation cannot be matched or excluded from query sequences; in addition, species with inadequate or uneven representation misrepresent species intra- and inter-specific genetic differences (Phillips, et al., 2019). For these reasons, DNA barcoding must be evaluated for its feasibility case-by-case, or rather species-by-species, to attain accurate species identification results.

Wisconsin's three *Microtus* species represent a unique opportunity to address some of the concerns regarding DNA barcoding and determine the feasibility of genetic methods for species identification in under-sampled, sympatric, and cryptic species. Effective conservation strategies for these species rely on accurate identification that could be improved by using

genetic methods. The prairie vole (*Microtus ochrogaster*) and woodland vole (*M. pinetorum*) are listed as Species of Greatest Conservation Need (SGCN) in Wisconsin (Wisconsin Department of Natural Resources, 2015). Both *M. ochrogaster* and *M. pinetorum* have stable national population levels (Cassola, 2016a; Cassola, 2016b), but are along the northern edge of their ranges in Wisconsin (Figure 1). Small and vulnerable populations of *M. ochrogaster* and *M. pinetorum* are currently limited to isolated patches in the southwestern corner of the state, despite the larger potential ranges depicted by the IUCN (Wisconsin Department of Natural Resources, 2015; Figure 2). A primary challenge in conservation efforts for *M. ochrogaster* in particular, is the overlap in distribution and habitat with the third and more common species of *Microtus* in Wisconsin, the meadow vole (*M. pennsylvanicus*).

The three *Microtus* species in Wisconsin have similar morphology. *M. pinetorum* has a significantly shorter tail than the other two species (Wisconsin Department of Natural Resources, 2013; Table 1), though this trait can vary by sex, life stage, and environment in rodents (Kingsley, et al., 2017; Jackson and van Aarde, 2003; Miller, et al., 2010). Species identification between *M. ochrogaster* and *M. pennsylvanicus* can be more difficult as the two are nearly identical morphologically (Everson, et al., 2023). Number of plantar tubercles is primarily used to distinguish between *M. pennsylvanicus* and *M. ochrogaster* in the field. Everson, et al. (2023) aimed to disentangle the morphology of *M. pennsylvanicus* and *M. ochrogaster* in sympatric Kentucky populations by evaluating standard external measurements (head-body length, tail length, hind-foot length, ear length from notch, and body mass) and the reflectance spectra of fur. Even though they found statistical differences between species' measurements, there was too great an overlap in the range of external morphology

measurements to unambiguously distinguish these two species. Dental characteristics, specifically molar cusp patterns, were 94% accurate in identifying species, indicating that even the most reliable method may not be a perfect diagnostic character (Everson, et al., 2023). Dental morphology also requires inspection of skulls, so is not useful for field-based species identification of live animals. Mammae count may also be an accurate method of species identification. *M. pennsylvanicus* has eight, *M. ochrogaster* has six, and *M. pinetorum* has four, which correlates with each species' litter size (Hall, 1981). However, to date, no formal study has evaluated mammae count as a method for species identification. The current practice of species identification in the field relies on overall appearance, including size, dorsal and ventral pelage, plantar tubercles, and tail length (Table 1). Despite these carefully collected taxonomic characteristics, overlaps in an organism's morphology can result in misidentification in the field, especially for juvenile individuals (Nekrutenko, et al., 2000).

To determine the best method of DNA-based species identification for *Microtus* voles, we explore two distinct approaches: a length polymorphism in the arginine vasopressin receptor 1a (*avpr1a*) gene (henceforth, HY approach) and DNA-barcoding of the *COI* gene. In both approaches, we use tissue and a variety of less invasively collected samples (fecal pellets, whiskers, and fur). Henterly, et al. (2011) reported a species-specific difference within the *avpr1a* gene, which is involved in social and reproductive behavior (Hammock and Young, 2005). *M. ochrogaster*, which has a monogamous mating system, has a microsatellite (~430bp) in the regulatory region of the *avpr1a* gene that is absent in the promiscuous *M. pennsylvanicus*. Whereas this large size difference in the *avpr1a* gene was readily observable, producing accurate and repeatable species identification, the Henterly, et al. (2011) study was

conducted in Kansas and Indiana and only tested males. Due to the reported regional differences between populations of *Microtus* (Jackson and Cook, 2019), our study aimed to determine if this variant was diagnostic in Wisconsin populations, whether the reported differences were consistent in female individuals and, for the first time, test the method on *M. pinetorum*. The HY approach is inexpensive and only requires gel electrophoresis, suggesting a possible advantage over a DNA barcoding approach. However, because it is based on nuclear DNA and amplifies a relatively large fragment, it may not be robust for non-invasively collected samples with small amounts of potentially degraded DNA.

We set out to compare our *Microtus* species identification using the *avpr1a* gene length polymorphism to a DNA barcoding approach that we developed specifically for species identification between the three *Microtus* species in Wisconsin. The DNA barcoding approach has two theoretical advantages over the length polymorphism approach for non-invasively collected samples. First, the higher copy number of mtDNA per cell allows for a greater chance of recovering genetic information from non-invasively collected samples that often contain DNA of low quantity and/or low quality. Second, the shorter DNA fragments in the *COI* gene are expected to amplify robustly even in degraded samples, potentially permitting less invasive sampling techniques in the field. To further test the capability of DNA barcoding, we aim to test its efficacy when there are insufficient reference sequences in the database. BOLD holds 15,756,430 reference sequences as of publication, representing 351,225 species. The 38 species of *Microtus* have differing representation in the database. *Microtus* pennsylvanicus has 149 sequences, *M. ochrogaster* has 19 sequences, and *M. pinetorum* has only 2 sequences. In this study, we evaluate two sequence lengths of the *COI* gene. The first region is ~445bp in length,

the second is ~122bp and is contained within the first region. A longer barcode will survey more SNPs used for species identification, but we anticipated a shorter barcode might have a higher chance of amplifying in our degraded and low-quality DNA samples (Meusnier, et al., 2008). By comparing two distinct DNA-based species identification approaches, the HY approach and a DNA-barcoding approach, we aim to determine which is most useful for reliable identification of Wisconsin's three *Microtus* species. Accurate species identification will facilitate the development of targeted conservation strategies and robust population monitoring, improving conservation efforts for *M. ochrogaster* and *M. pinetorum.*

Table 1. Comparison of Wisconsin's three *Microtus* species' physical characteristics most often used for species identification (Everson, et al. 2023; Ford, et al., 2007; Laerm and Ford, 2007; Schwartz and Schwartz, 2001; Smolen, 1981)

Characteristic	***M. ochrogaster***	***M. pennsylvanicus***	***M. pinetorum***
Weight	37-73 g	25-65g	19-37g
Ventral pelage	Yellow or rusty	Gray or silver	Silver or white
Dorsal pelage	Grayish brown; grizzled appearance	Reddish brown; short, smooth fur	Reddish brown; smooth, silky fur
Plantar tubercles	5	6 (sometimes 5)	5
Tail length	<2x length of hind foot	>2x length of hind foot	~1x length of hind foot

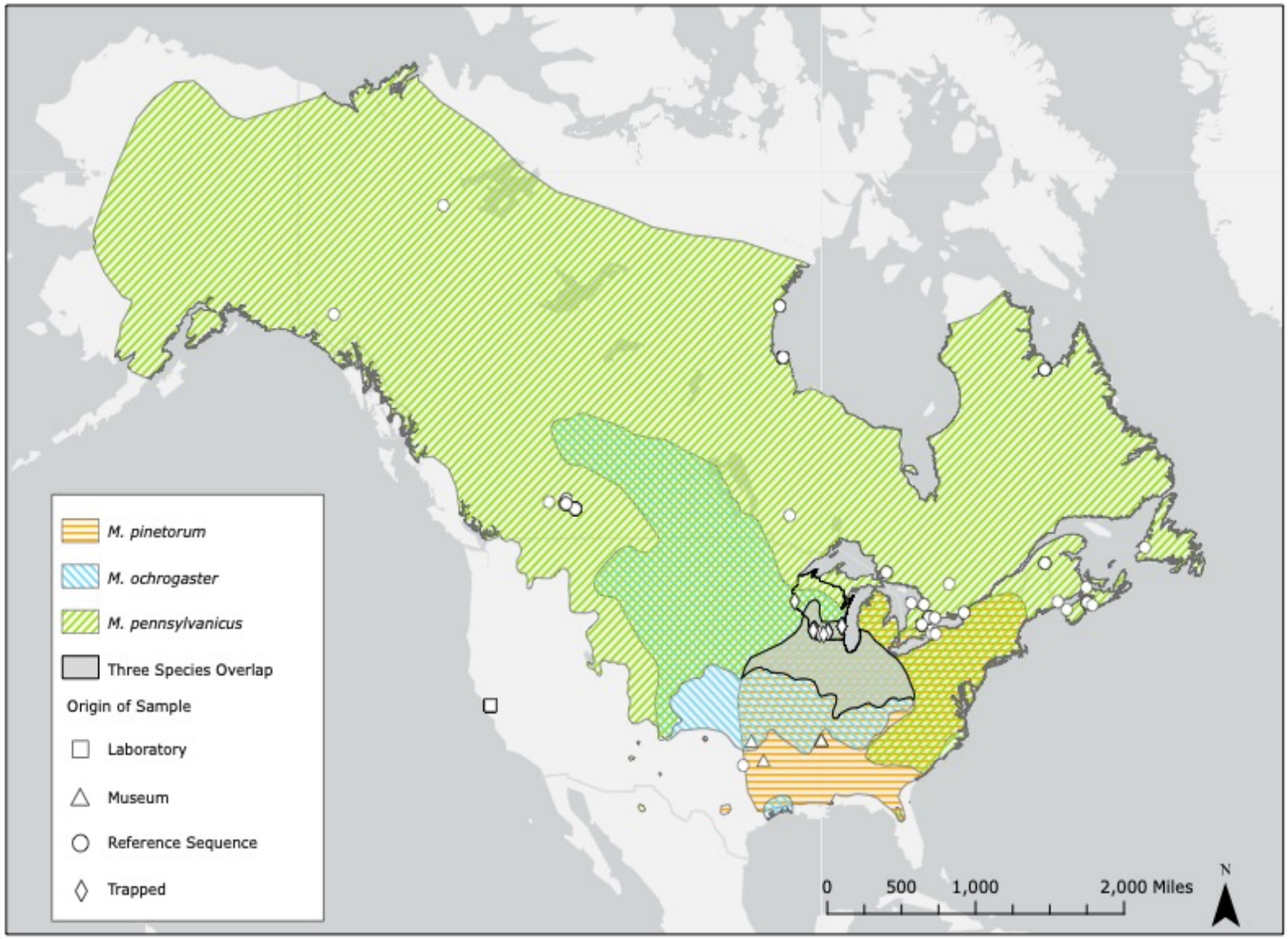

Figure 1. IUCN Suitable habitat ranges for all three Wisconsin *Microtus* species and origin of samples used in this study. Wisconsin, the focal state of this study and the range of overlap between all three species is highlighted (Cassola, 2016; Cassola,

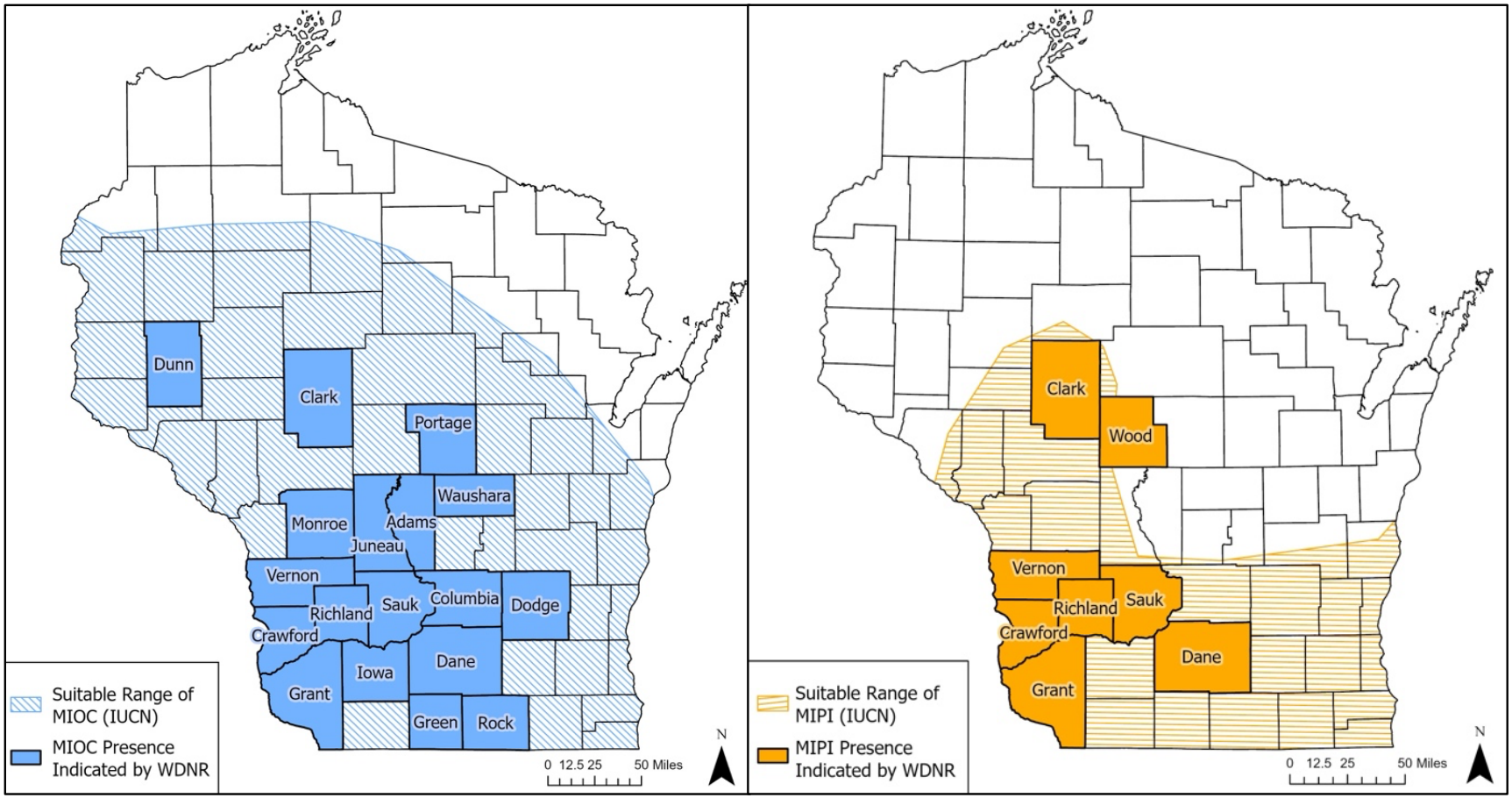

Figure 2. IUCN suitable habitat ranges for *M. ochrogaster* (MIOC; left) and *M. pinetorum* (MIPI; right) and the Wisconsin counties that host current populations of each species (Cassola, 2016; Cassola, 2016; Wisconsin Department of Natural Resources, 2015). *M. pennsylvanicus* occurs throughout Wisconsin.

Methods

Sample Collection

We collected 85 samples from 65 voles representing all three species (Figure 1; Table 2). When possible, multiple samples were collected from individual voles. We sampled tissue (n = 58) fecal pellet (n = 19) whisker (n = 6), and fur (n = 2). We designated the specimen status of samples from each vole as either "known" or "unknown." Samples collected in sites where two or more *Microtus* species overlap in potential range were considered "unknown" and were identified to species using traditional morphological methods (primarily ventral pelage and number of tubercles). In this group, 25 fresh tissue samples were collected from voles trapped in 2021 during the annual small mammal trapping efforts completed for the state's biotic inventory. Tissue was collected from a 2 mm ear biopsy punch, placed in 1 ml of ethanol in 1.5 ml tubes, and stored at 4°C.

Samples collected from allopatric sites where only one *Microtus* species occurs, lab-bred populations, or verified museum collections were considered "known." *M. pennsylvanicus* and *M. ochrogaster* samples were collected through collaboration with Dr. Annaliese Beery of the University of California – Berkeley. The Beery lab studies neurobiological mechanisms of sociality in peer relationships with *M. pennsylvanicus* and *M. ochrogaster* as model species (e.g., Lee and Beery, 2022). We received 1-2 mm of tissue from the tail of eight individuals, fecal pellets from six individuals and paired samples (fecal and tissue) from six individuals. Liver tissue was sampled from five *M. pinetorum* specimens housed at the Museum of Texas Tech University (TTU). In addition, we sampled six *M. pennsylvanicus* tissue (one liver, five muscle)

specimens and one *M. ochrogaster* tissue (liver) specimen from the University of Wisconsin Zoological Museum (UWZM). Tissue and fecal pellets were transported on ice and stored at -20°C.

To obtain "known" samples from *M. pennsylvanicus* in Wisconsin, we collected tissue, fecal pellets, whiskers, and fur from individuals trapped at the University of Wisconsin - Milwaukee (UWM) Field Station, allopatric habitat where only *M. pennsylvanicus* is found (Figure 2). Trapping was conducted October through November of 2021 and May through October of 2022, in accordance with UWM IACUC 21-22 #19 and a UWM Field Station Use Permit. Three Havahart X-Small traps and 11 Sherman small traps were placed perpendicular to the pre-existing drift fence located between wetland and meadow habitat. Traps were pre-baited for two days with peanut butter, birdseed, and Tootsie Roll pieces. Before setting the traps, they were cleaned out to remove any material left by non-target species. Traps were set during daylight hours when voles are most active (Ambrose, 1973) and checked every two hours. Trapped animals were removed from traps and handled using leather gloves. Additionally, three voles unintentionally trapped in pitfall traps (also located along the drift fence fora Field Methods class taught at the UWM Field Station) were sampled.

We collected a small (1-2 mm) tail clip, whiskers (5 whiskers with follicles intact), and fur (10 hairs with follicles intact) from each trapped animal. If fecal pellets were in the trap, we also collected those. Tissue and fecal pellets were placed in 1 ml of 100% ethanol in 1.5 ml tubes and stored at 4°C. Whiskers and fur were placed in small envelopes and stored at 4°C. Multiple samples were collected from each animal to determine if the two approaches were effective

with degraded samples. After sample collection, animals were released back into their habitat within a few feet of where they were trapped.

Table 2. Origin, sample types, specimen status, and species (MIPE = *M. pennsylvanicus*, MIOC = *M. ochrogaster*, MIPI = *M. pinetorum*) of voles used in this study (ET = Ear Tissue, TT = Tail Tissue, MT = Muscle Tissue, LT = Liver Tissue, F = Fecal Pellet, W = Whisker, H = Hair/Fur. Asterisks represent species determined using field-based species identification made by WDNR biologists.

#	ID	Acquired From	Sex	Origin	Sample Type(s)	Specimen Status	Species
1	**01**	WDNR	Male	Trapped	ET	Unknown	MIPE*
2	**02**	WDNR	Male	Trapped	ET	Unknown	MIPE*
3	**03**	WDNR	Female	Trapped	ET	Unknown	MIPE*
4	**04**	WDNR	Female	Trapped	ET	Unknown	MIPE*
5	**05**	WDNR	Female	Trapped	ET	Unknown	MIPE*
6	**06**	WDNR	Male	Trapped	ET	Unknown	MIPE*
7	**07**	WDNR	Male	Trapped	ET	Unknown	MIPE*
8	**08**	WDNR	Male	Trapped	ET	Unknown	MIPE*
9	**09**	WDNR	Male	Trapped	ET	Unknown	MIPE*
10	**10**	WDNR	Female	Trapped	ET	Unknown	MIPE*
11	**11**	WDNR	Male	Trapped	ET	Unknown	MIPE*
12	**12**	WDNR	Female	Trapped	ET	Unknown	MIPE*
13	**13**	WDNR	Male	Trapped	ET	Unknown	MIPE*
14	**14**	WDNR	Female	Trapped	ET	Unknown	MIPE*
15	**15**	WDNR	Female	Trapped	ET	Unknown	MIPE*
16	**16**	WDNR	Male	Trapped	ET	Unknown	MIPE*
17	**17**	WDNR	Male	Trapped	ET	Unknown	MIOC*
18	**18**	WDNR	Male	Trapped	ET	Unknown	MIPE*
19	**19**	WDNR	Female	Trapped	ET	Unknown	MIPE*
20	**20**	WDNR	Female	Trapped	ET	Unknown	MIPE*
21	**21**	WDNR	Male	Trapped	ET	Unknown	MIOC*
22	**22**	WDNR	Female	Trapped	ET	Unknown	MIOC*
23	**23**	WDNR	Female	Trapped	ET	Unknown	MIOC*
24	**24**	WDNR	Male	Trapped	ET	Unknown	MIPE*
25	**25**	WDNR	Male	Trapped	ET	Unknown	MIOC*
26	**FSA**	UWM FS	Unknown	Trapped	TT	Known	MIPE
27	**FS1**	UWM FS	Unknown	Trapped	TT, F, W, H	Known	MIPE
28	**FS2**	UWM FS	Unknown	Trapped	TT, F, W, H	Known	MIPE
29	**FS3**	UWM FS	Unknown	Trapped	TT, F	Known	MIPE
30	**FS4**	UWM FS	Unknown	Trapped	TT, F, W	Known	MIPE
31	**FS5**	UWM FS	Unknown	Trapped	TT, F, W	Known	MIPE
32	**FS6**	UWM FS	Unknown	Trapped	TT, F, W	Known	MIPE

33	**FS7**	UWM FS	Unknown	Trapped	F, W	Known	MIPE
34	**40**	UWZM	Unknown	Museum	MT	Known	MIPE
35	**54**	UWZM	Unknown	Museum	MT	Known	MIPE
36	**59**	UWZM	Unknown	Museum	LT	Known	MIPE
37	**60**	UWZM	Unknown	Museum	MT	Known	MIPE
38	**62**	UWZM	Unknown	Museum	MT	Known	MIPE
39	**68**	UWZM	Unknown	Museum	MT	Known	MIPE
40	**88**	UWZM	Unknown	Trapped	LT	Known	MIPI
41	**A51**	Beery	Male	Laboratory	TT, F	Known	MIOC
42	**A52**	Beery	Male	Laboratory	TT	Known	MIOC
43	**B19**	Beery	Male	Laboratory	TT, F	Known	MIPE
44	**B25**	Beery	Male	Laboratory	TT, F	Known	MIPE
45	**2169**	Beery	Female	Laboratory	F	Known	MIPE
46	**2247**	Beery	Female	Laboratory	F	Known	MIPE
47	**2273**	Beery	Female	Laboratory	F	Known	MIPE
48	**3150**	Beery	Female	Laboratory	TT	Known	MIPE
49	**3172**	Beery	Female	Laboratory	F	Known	MIPE
50	**3193**	Beery	Female	Laboratory	TT, F	Known	MIPE
51	**3194**	Beery	Female	Laboratory	TT	Known	MIPE
52	**3196**	Beery	Female	Laboratory	TT	Known	MIPE
53	**3197**	Beery	Female	Laboratory	TT	Known	MIPE
54	**4066**	Beery	Female	Laboratory	TT	Known	MIOC
55	**4067**	Beery	Female	Laboratory	TT, F	Known	MIOC
56	**4074**	Beery	Male	Laboratory	TT, F	Known	MIOC
57	**4075**	Beery	Male	Laboratory	TT	Known	MIOC
58	**4076**	Beery	Male	Laboratory	TT	Known	MIOC
59	**8214**	Beery	Female	Laboratory	F	Known	MIOC
60	**8219**	Beery	Female	Laboratory	F	Known	MIOC
61	**T1**	TTU-M	Male	Museum	LT	Known	MIPI
62	**T2**	TTU-M	Male	Museum	LT	Known	MIPI
63	**T3**	TTU-M	Female	Museum	LT	Known	MIPI
64	**T4**	TTU-M	Female	Museum	LT	Known	MIPI
65	**T5**	TTU-M	Male	Museum	LT	Known	MIPI

DNA Extraction

DNA was extracted from all samples using DNeasy Blood and Tissue methods, following the manufacturer's protocol (Qiagen, Hilden, Germany). The protocol was adjusted to elute DNA in two steps, using 100 µl AE Buffer each time. DNA concentration was determined using a Qubit 4

Fluorometer following the dsDNA high-sensitivity quantification assay protocol (Thermo Fisher Scientific, Waltham, MA, USA). Extracted DNA was normalized to a concentration of 2 ng/µl.

Length polymorphism in the *avpr1a* gene (HY Method)

Henterly, et al. (2011) found that the *avpr1a* gene could unambiguously distinguish between *M. ochrogaster* and *M. pennsylvanicus* in a sample of male voles from Kansas and Indiana. It was shown that *M. ochrogaster* individuals had a microsatellite (~430 bp) in the regulatory region of the *avpr1a* gene that was absent in *M. pennsylvanicus* individuals, resulting in an easily visible length polymorphism between species. We used the primers outlined in Henterly et al. (2011) to amplify the microsatellite-containing region of the *avpr1a* gene by polymerase chain reaction (PCR; Table 3). Our PCR reagents were identical to those in Henterly et al. (2011), but our reaction conditions differed slightly. Our thermocycler profile was an initial denaturation at 94°C for 2 min; 30 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 30 s; and a final elongation for 5 min at 72°C. We ran each sample on a 1% agarose gel alongside a known size standard (100 bp ladder from New England Biolabs) and visualized bands under UV light. Gel electrophoresis of the resulting PCR product was expected to show a diagnostic product size of ~290 bp in *M. pennsylvanicus* samples and ~720 bp in *M. ochrogaster* samples. We planned to identify the species of samples based on the length of the PCR product: ~290bp for *M. pennsylvanicus* and ~720bp for *M. ochrogaster.* As these methods had not previously been tested on *M. pinetorum*, we did not have an explicit expectation for the product size of amplified *M. pinetorum* samples. However, *M. pinetorum* has a monogamous mating system like *M. ochrogaster* (FitzGerald and Madison, 1983), so we predicted that *M. pinetorum* might

share the longer fragment with *M. ochrogaster*. If fragment sizes are shared between species, the HY method would not be diagnostic.

Table 3. Primers used for *Microtus* species identification.

Primer Name	Sequence (5' → 3')
Microtus_HY-F	GTATTGCCACAAATAGACCAACG
Microtus_HY-R	GTAAGGATGACAGGCGTTACTG
Microtus_BOLD445-F	GGYTTCGGCAACTGACTTGT
Microtus_BOLD445-R	CTCCAGCWGGGTCRAAGAAR
Microtus_BOLD122-F	TTGGAGCACCAGAYATAGCA
Microtus_BOLD122-R	AGTTCAGCCTGTTCCTGCYC

DNA Barcoding using the *COI* gene (COI122 and COI445 Methods)

We downloaded sequences available in BOLD to generate primers for the sequence polymorphism approach (in October 2021, n=125 *M. pennsylvanicus*, n=3 *M. ochrogaster*). We aligned the sequences and generated a consensus sequence for each species in *Geneious* software (Geneious Prime 2023.2.1). Comparing the two consensus sequences revealed 86 base pair differences out of 657 bp of sequence (13%). We designed two primer sets using Primer3 (Untergasser, et al. 2012) to amplify variable regions of the *Microtus COI* gene (Table 3). One set of primers (COI445) was designed to amplify a larger region (445 bp) with more diagnostic sites. The second set of primers (COI122) was designed to amplify a shorter region (122 bp) with fewer variants but a higher chance of amplification in degraded samples. For both fragments, products were amplified in a 10 µl PCR, using 4ng template DNA, 0.25 µM forward primer 0.25 µM reverse primer, 1X of PCR buffer, 0.2 mM of dNTPs, 1U of Taq polymerase. Our thermocycler profile was identical to the HY method. PCR products were purified using ExoSAP-

IT, following the manufacturer's protocol (Thermo Fisher Scientific, Waltham, MA, USA). After PCR all samples were run on a 2% agarose gel to verify amplification and the expected product length. Using standard Sanger sequencing and the BigDye Terminator 3.1 cycle sequencing kit, we sequenced PCR products through the University of Wisconsin Biotechnology Center (UWBC) and the University of Illinois at Urbana-Champaign Roy J. Carver Biotechnology Center. Chromatograms were trimmed and edited manually in Geneious software. Bidirectional sequences were assembled and trimmed to final lengths of 116bp (COI122) and 410bp (COI445). All additional *COI* reference data was downloaded from BOLD.

We identified the species of each specimen in two ways. To demonstrate the accessibility of DNA barcoding we performed a search of all sequences using NCBI's Basic Local Alignment Search Tool (BLAST; Sayer, et al., 2022). BLAST aligns query sequences with GenBank sequences, the NIH database of all publicly available DNA sequences. We performed a megablast search with parameters set to "highly similar sequences" and reviewed the top 10 returned sequences and their percent identity matches. For each sample, we calculated the proportion of the 10 returned sequences that held a percent identity match greater than 95%. If more than half of the returned sequences met these criteria it was considered an accurate match. However, *COI* sequences from *M. ochrogaster* and *M. pinetorum* are underrepresented in the GenBank database (*M. ochrogaster* = 7 sequences; *M. pinetorum* = 2 sequences); for these samples we calculated the percentage of only the top seven (*M. ochrogaster)* and top two (*M. pinetorum)* returned sequences that matched the true species of each sample. Additionally, we identified the species of our sequences with the R package, 'BarcodingR'

(Zhang, et al., 2017), using three methods (BP-based, fuzzy-set based, and Bayesian-based) that compared our sequences to the reference sequences available in BOLD.

Genetic Analysis

To evaluate our reference sequences and determine if there was a local barcoding gap for the three *Microtus* species, we conducted barcoding gap analysis with available BOLD reference sequences (n =149) using Kimura 2-parameter (K2P) genetic distances in the R package 'Barcoding R'. We visualized the barcoding gap with both the R packages 'BarcodingR' and 'spideR' (Brown, et al., 2012). We calculated the optimal threshold for interspecific distance using 'spideR' to determine if the BOLD standard (1%) threshold for species identification was appropriate for our dataset (Meyer and Paulay, 2005).

Pairwise genetic distances of our COI445 barcodes, derived from both known and unknown samples, were calculated using a K2P model and visualized as a heatmap using the R packages 'Ape' (Paradis and Schliep, 2019) and 'ComplexHeatmap' (Gu, et al., 2016). We constructed a Neighbor Joining (NJ) tree using all of our samples to determine if our COI445 barcodes showed species-based clustering. We did not use the tree as a tree-based method for species identification nor did we employ it to analyze evolutionary relationships between samples. In Geneious, we used a Tamura-Nei distance model to create an NJ tree of our COI445 barcodes and bootstrapped the tree using 1000 replicates to assess branch support. *M. fortis* was selected as recently diverged outgroup taxa and used to root the tree (Jackson, et. al., 2020). Branches with <50% consensus support were deleted. We visualized and annotated the tree using iTOL (Letunic and Bork, 2021).

Results

Length polymorphism in the *avpr1a* gene (HY Method)

We successfully quantified the size of the *avpr1a* gene fragment for 70 of the 85 samples. Fifteen samples (21%; 1 of 25 ear tissue, 13 of 19 fecal, 1 of 6 whisker) repeatedly failed to amplify during PCR and were not included in the analysis. Of the 70 samples that successfully amplified, we were able to unambiguously determine the species for 73% (Table 4 & 5). Samples from both male (n = 24) and female (n = 20) individuals of all three species were tested using the HY approach, as expected, allele lengths corresponded to species, with no difference between sexes. This method was 100% accurate when distinguishing known *M. pennsylvanicus* samples (n=33) from the other two species. All known *M. pennsylvanicus* samples produced an allele length of ~290bp, consistent with the results from Henterly et. al. (2011). Species-level discrimination was not possible for *M. ochrogaster* and *M. pinetorum* using the HY Method. Both known *M. ochrogaster* (n = 8) and known *M. pinetorum* samples (n = 5) produced an allele length of ~720bp (Table 4). Unknown samples (n=25) aligned with their field species identification in all but one sample (Table 5). Sample 05 was identified in the field as *M. pennylvanicus* but produced a fragment length of ~720bp (Figure 3), identifying it as *M. ochrogaster/M. pinetorum.* DNA barcoding and genetic analysis also identified this sample as *M. ochrogaster* (described below).

Table 4. Species identity assignments for "known" samples, based on gel electrophoresis results of the length polymorphism in the *avpr1a* gene (HY Method). Sequences are derived from tissue unless noted with an underscore followed by sample type, e.g. "FS1_F" is a fecal pellet from individual FS1 (F= Fecal Pellet, H = Fur, W = Whisker).

Known Sample ID	True Species	HY Fragment Length (bp)	Known Sample ID	True Species	HY Fragment Length (bp)
40	MIPE	~290	FS4	MIPE	~290
54	MIPE	~290	FS4_F	MIPE	~290
59	MIPE	~290	FS5	MIPE	~290
60	MIPE	~290	FS5_W	MIPE	~290
62	MIPE	~290	FS6	MIPE	~290
68	MIPE	~290	FS6_F	MIPE	~290
3150	MIPE	~290	FS6_W	MIPE	~290
3193	MIPE	~290	FS7_F	MIPE	~290
3194	MIPE	~290	FS7_W	MIPE	~290
3196	MIPE	~290	FSA	MIPE	~290
3197	MIPE	~290	88	MIOC	~720
B19	MIPE	~290	4066	MIOC	~720
B25	MIPE	~290	4067	MIOC	~720
FS1	MIPE	~290	4074	MIOC	~720
FS1_F	MIPE	~290	4075	MIOC	~720
FS1_H	MIPE	~290	4076	MIOC	~720
FS1_W	MIPE	~290	A51	MIOC	~720
FS2	MIPE	~290	A52	MIOC	~720
FS2_F	MIPE	~290	TTU1	MIPI	~720
FS2_H	MIPE	~290	TTU2	MIPI	~720
FS2_W	MIPE	~290	TTU3	MIPI	~720
FS3	MIPE	~290	TTU4	MIPI	~720
FS3f	MIPE	~290	TTU5	MIPI	~720

Table 5. Species identity assignments for “unknown” samples, based on gel electrophoresis results of the length polymorphism in the *avpr1a* gene (HY Method). Sample 05, highlighted in yellow, produced the longer ~720bp fragment indicating that it was not *M. pennsylvanicus.* DNA barcoding and genetic analysis confirmed its species as *M. ochrogaster.*

Unknown Sample ID	WDNR ID	HY Fragment Length
02	MIPE	~290
03	MIPE	~290
04	MIPE	~290
06	MIPE	~290
07	MIPE	~290
08	MIPE	~290
09	MIPE	~290
10	MIPE	~290
11	MIPE	~290
12	MIPE	~290
13	MIPE	~290
14	MIPE	~290
15	MIPE	~290
16	MIPE	~290
18	MIPE	~290
19	MIPE	~290
20	MIPE	~290
24	MIPE	~290
17	MIOC	~720
21	MIOC	~720
22	MIOC	~720
23	MIOC	~720
25	MIOC	~720
05	MIPE	~720

Well	1	2	3	4	5	6	7	8	9	10
Contents	Ladder	FS1	T1	T2	T3	T4	T5	88	Negative	Ladder
Species		MIPE	MIPI	MIPI	MIPI	MIPI	MIPI	MIOC	Control	
HY PCR Prod.		~290	~720	~720	~720	~720	~720	~720		

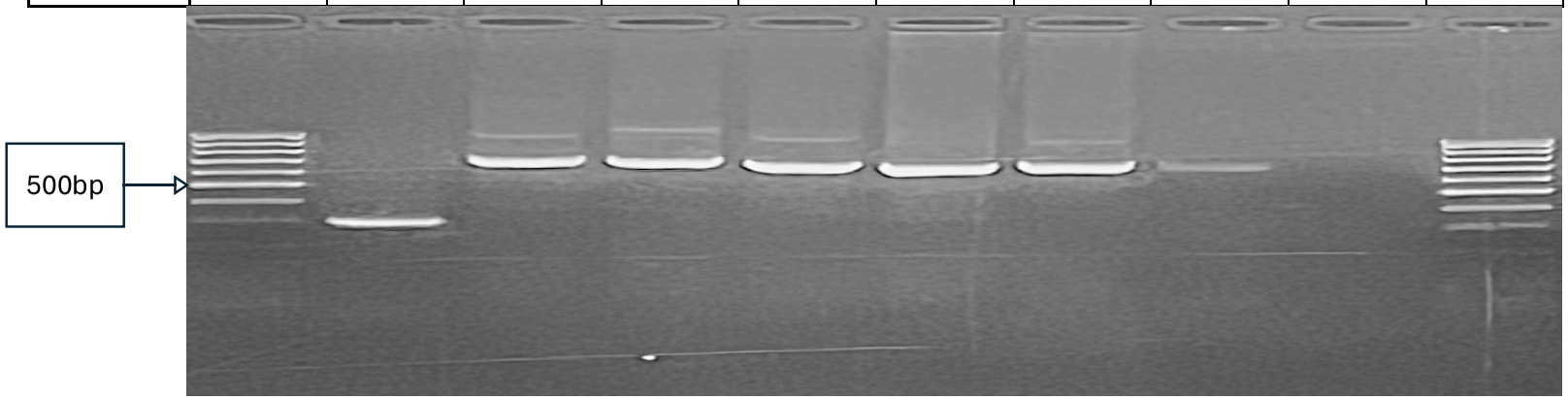

Figure 3. Gel electrophoresis HY method results of known *M. pennsylvanicus, M. pinetorum,* and *M. ochrogaster* alongside two 100bp ladders and one negative control. Well 2 shows amplified product for sample FS1, demonstrating the expected shorter fragment of ~290bp for *M. pennsylvanicus.* Well 8 shows amplified product for 88, demonstrating the expected longer fragment of ~720bp for *M. ochrogaster*. The HY method had not previously been tested on *M. pinetorum.* For the *M. pinetorum* samples included in this study, PCR product was equivalent in length with the *M. ochrogaster* fragment of ~720bp.

DNA Barcoding using the *COI* gene (COI122 and COI445 Methods)

We included all 85 samples collected for this study in DNA barcoding analysis. Barcodes were defined as assembled bidirectional sequences with a minimum of 410bp for COI445 and 116bp for COI122. We sequenced 53 COI122 barcodes (overall 61.2% success rate), 30 from known samples and 22 from unknown samples. No COI122 barcodes were successfully sequenced for *M. pinetorum.* For COI445 we sequenced 74 barcodes (overall 87.1% success rate), 50 from known samples and 24 from unknown samples (Table 6). COI445 primers were superior to COI122 primers in returning successful barcodes (McNemar X^2 (1, N = 85) = 13, $p < .001$). When compared by sample type, tissue samples returned more successful barcodes than noninvasively collected samples (X^2(3, N = 170) = 22.23, $p < .001$; Figure 4). Table 7 shows the number of barcodes created for known samples of each species.

BLAST searches for COI122 and COI445 known samples each returned top target sequences that matched the sample's true species. All known *M. pennsylvanicus* COI122 and COI445 barcodes (n = 45) each returned 10 *M. pennsylvanicus* voucher specimens as their top 10 target sequences (Figure 5). *M. ochrogaster* (n = 16) and *M. pinetorum* (n = 5) known barcodes each returned seven and two voucher specimens, respectively, as their top target sequences. Both COI122 and COI445 barcodes from unknown samples were also clearly identified to species when used in a BLAST search. All unknown samples except Sample 05 were identified by BLAST as the species assigned based on field identification. Sample 05, identified as *M. pennsylvanicus* in the field, was identified as *M. ochrogaster*, which agreed with our genetic results from the HY approach that indicated this sample was *M. ochrogaster/M. pinetorum.* All unknown samples matched 100% (10, 7, or 2 matches) of returned sequences,

except for 4 unknown samples identified in the field as *M. ochrogaster* (17, 21, 23, 25). Although these four BLAST searches returned 4 *M. ochrogaster*, which was sufficient for species identification, they also returned 3 *M. chrotorrhinus (*rock vole) voucher specimens, albeit with lower percent match (>5% lower than the *M. ochrogaster* voucher specimens). The ranges of *M. ochrogaster* and *M. chrotorrhinus* slightly overlap in the southern Appalachian Mountain range (Cassola, 2016c) and therefore would not be considered a viable identification when sampling Wisconsin voles.

In the R package 'BarcodingR' we input 80 barcodes for the 56 known individuals (COI122 n = 31, COI445 n= 49) and 149 reference sequences from BOLD, aligned and trimmed to the minimum length of each barcode. We omitted one *M. ochrogaster* reference sequence from BOLD (ABRCR301-06) that had a mean intraspecific genetic distance (17.6%) with *M. ochrogaster* that was greater than the highest interspecific distance between all three species (15.8%). We assigned each sequence to a species using the BP-based, fuzzy-set based, and Bayesian-based methods. Each method returned a species designation and produced a species assignment value between 0.0 – 1.0. The Fuzzy Membership Function (FMF) had a greater range in assignment value than the BP-based or Bayesian-based methods, with values from 0.366 - 1. All known barcodes were assigned to their true species in all three methods (Figure 6). *M. pinetorum* samples had the lowest FMF values (T1 = 0.393, T2 = 0.366, T3 = 0.366, T4 = .475, T5 = .393), most likely because there were only two *M. pinetorum* reference sequences available for comparison. We analyzed 46 barcodes (COI122 n=22, COI445 n=24) for the 24 unknown individuals in BarcodingR, using the same 149 reference sequences mentioned above. Sample 05 was the only unknown sample that returned a species designation that differed from

the field-based species identification, once more identifying the sample as *M. ochrogaster,* in contrast to its field identification as *M. pennsylvanicus* (Figure 7).

Table 6. Barcodes created for each sample. Samples that repeatedly failed to amplify and/or sequence are marked as "FAIL." Samples listed with barcodes had both COI122 and COI445 barcodes, only COI122, or only COI445. Asterisks represent species determined using field species identification made by the WDNR.

Sample ID	Sample type	Barcode(s)	Species
01	T	FAIL	MIPE*
02	T	COI122 & COI445	MIPE*
03	T	COI122 & COI445	MIPE*
04	T	COI122 & COI445	MIPE*
05	T	COI122 & COI445	MIPE*
06	T	COI122 & COI445	MIPE*
07	T	COI122 & COI445	MIPE*
08	T	COI122 & COI445	MIPE*
09	T	COI122 & COI445	MIPE*
10	T	COI122 & COI445	MIPE*
11	T	COI122 & COI445	MIPE*
12	T	COI122 & COI445	MIPE*
13	T	COI122 & COI445	MIPE*
14	T	COI122 & COI445	MIPE*
15	T	COI122 & COI445	MIPE*
16	T	COI122 & COI445	MIPE*
17	T	COI122 & COI445	MIOC*
18	T	COI445	MIPE*
19	T	COI122 & COI445	MIPE*
20	T	COI122 & COI445	MIPE*
21	T	COI122 & COI445	MIOC*
22	T	COI445	MIOC*
23	T	COI122 & COI445	MIOC*
24	T	COI122 & COI445	MIPE*
25	T	COI122 & COI445	MIOC*
40	T	COI122 & COI445	MIPE
54	T	COI122 & COI445	MIPE
59	T	COI445	MIPE

60	T	COI445	MIPE
62	T	COI445	MIPE
68	T	COI445	MIPE
88	T	COI122 & COI445	MIOC
3150	T	COI122 & COI445	MIPE
3193	T	COI122 & COI445	MIPE
3194	T	COI122 & COI445	MIPE
3196	T	COI122 & COI445	MIPE
3197	T	COI122 & COI445	MIPE
4066	T	COI122 & COI445	MIOC
4067	T	COI122 & COI445	MIOC
4074	T	COI122 & COI445	MIOC
4075	T	COI445	MIOC
4076	T	COI445	MIOC
2169_F	F	COI122 & COI445	MIPE
2247_F	F	COI445	MIPE
2273_F	F	COI122 & COI445	MIPE
3172_F	F	FAIL	MIPE
3193_F	F	COI445	MIPE
4067_F	F	COI122	MIOC
4074_F	F	COI122	MIOC
8214_F	F	COI445	MIOC
8219_F	F	COI122	MIOC
A51	T	COI445	MIOC
A51_F	F	FAIL	MIOC
A52	T	COI445	MIOC
B19	T	COI122 & COI445	MIPE
B19_F	F	COI445	MIPE
B25	T	COI122 & COI445	MIPE
B25_F	F	COI445	MIPE
FS1	T	COI122 & COI445	MIPE
FS1_F	F	COI122 & COI445	MIPE
FS1_H	H	COI122 & COI445	MIPE
FS1_W	W	COI445	MIPE
FS2	T	COI122 & COI445	MIPE
FS2_F	F	COI445	MIPE
FS2_H	H	COI122	MIPE
FS2_W	W	COI445	MIPE
FS3	T	COI122 & COI445	MIPE
FS3_F	F	COI445	MIPE

FS4	T	COI122 & COI445	MIPE
FS4_F	F	FAIL	MIPE
FS4_W	W	COI122	MIPE
FS5	T	COI445	MIPE
FS5_F	F	COI445	MIPE
FS5_W	W	COI445	MIPE
FS6	T	COI122 & COI445	MIPE
FS6_F	F	FAIL	MIPE
FS6_W	W	COI445	MIPE
FS7_F	F	COI122	MIPE
FS7_W	W	COI122 & COI445	MIPE
FSA	T	COI122 & COI445	MIPE
T1	T	COI445	MIPI
T2	T	COI445	MIPI
T3	T	COI445	MIPI
T4	T	COI445	MIPI
T5	T	COI445	MIPI

Table 7. Summary of barcodes for COI122 and COI445 created from known samples, organized by species. No COI122 barcodes were successfully created for *M. pinetorum.*

Species	# of Samples	COI122	COI445
MIPE	42	19	36
MIOC	13	7	9
MIPI	5	0	5

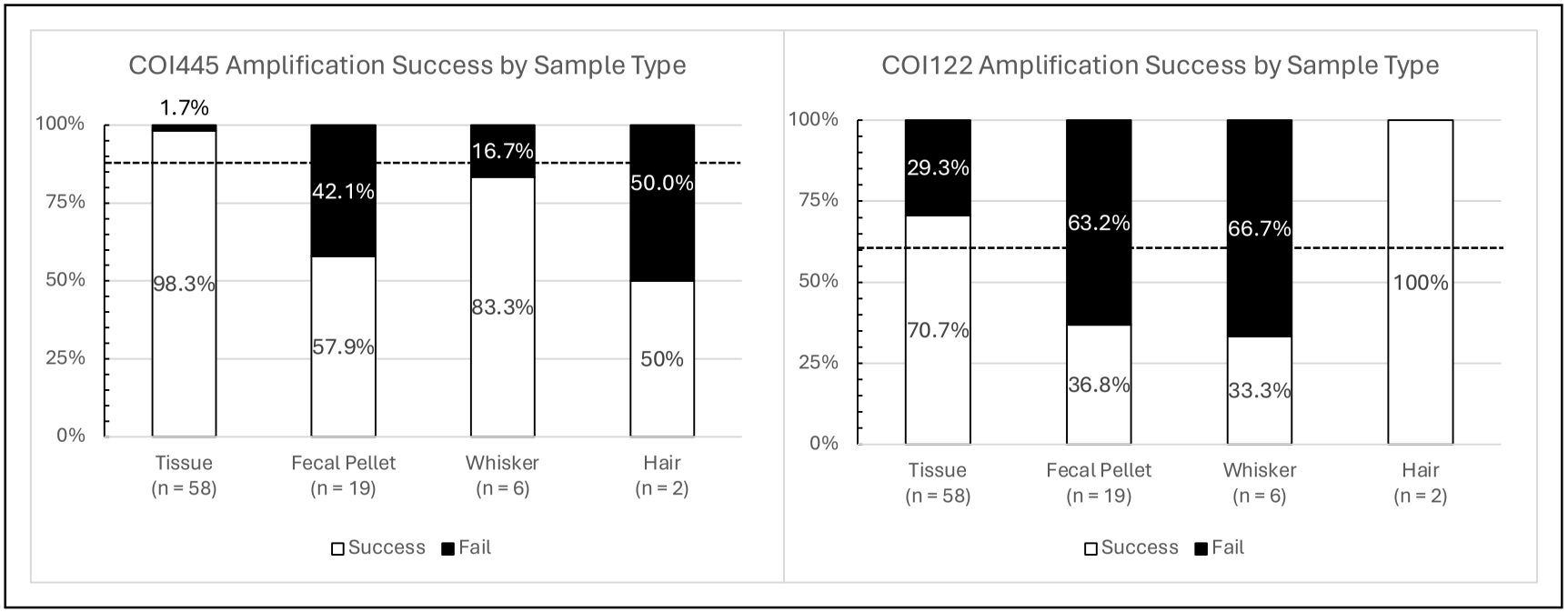

Figure 4. Barcoding success rates for COI445 and COI122 barcodes. COI445 primers (left) were more successful at producing barcodes than COI122 (right) (X^2 (1, N = 85) = 11, $p < .01$). When compared by sample type, tissue samples returned more successful barcodes than noninvasively collected samples (X^2(3, N = 170) = 22.23, $p < .01$). Dashed lines represent the overall success rate for all samples: COI445 = 87.1%, COI122 = 61.2%.

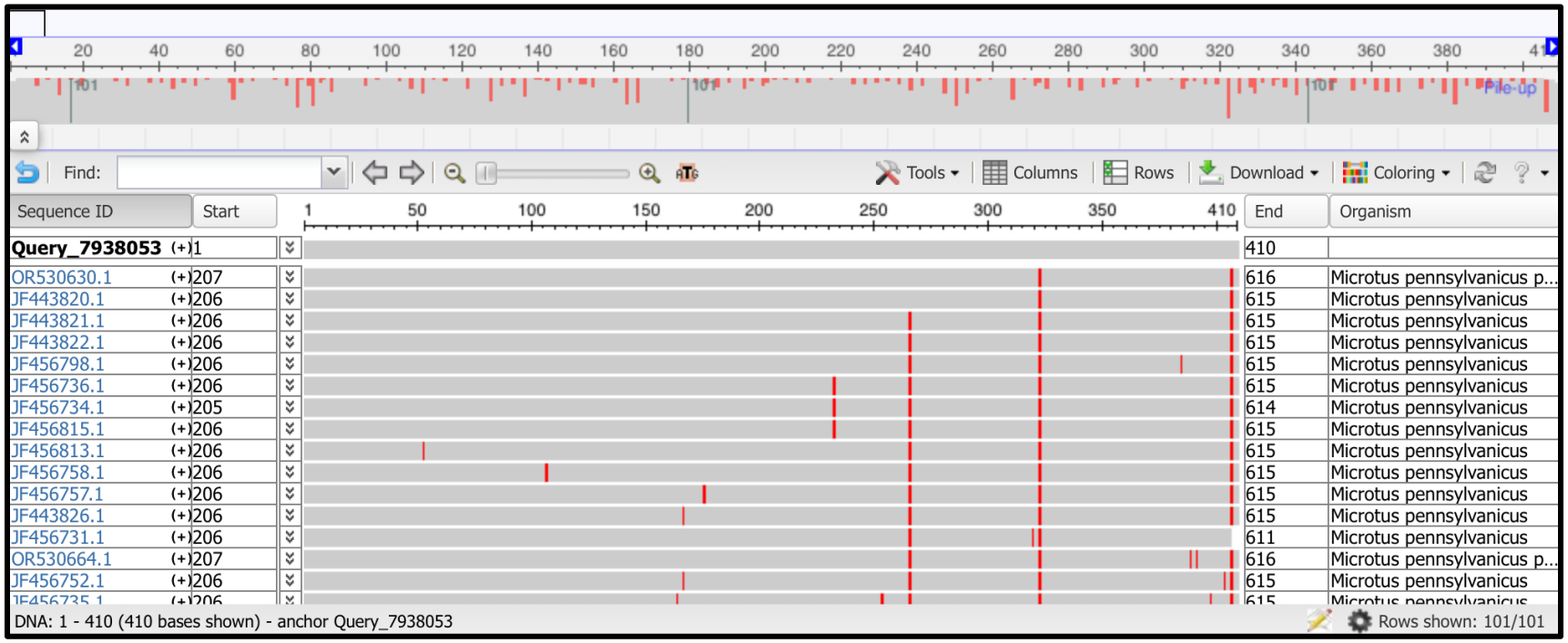

Figure 5. Screenshot of NCBI MSA Viewer 1.25.0, showing BLAST search results and alignments for COI445 barcode FS1, labeled as Query_7938053. The top 10 target sequences that were returned are aligned in percent match ID descending order. The organism of the voucher specimen is listed to the right of each sequence. Red bars indicate differences between the query sequence, FS1, and voucher specimen sequences.

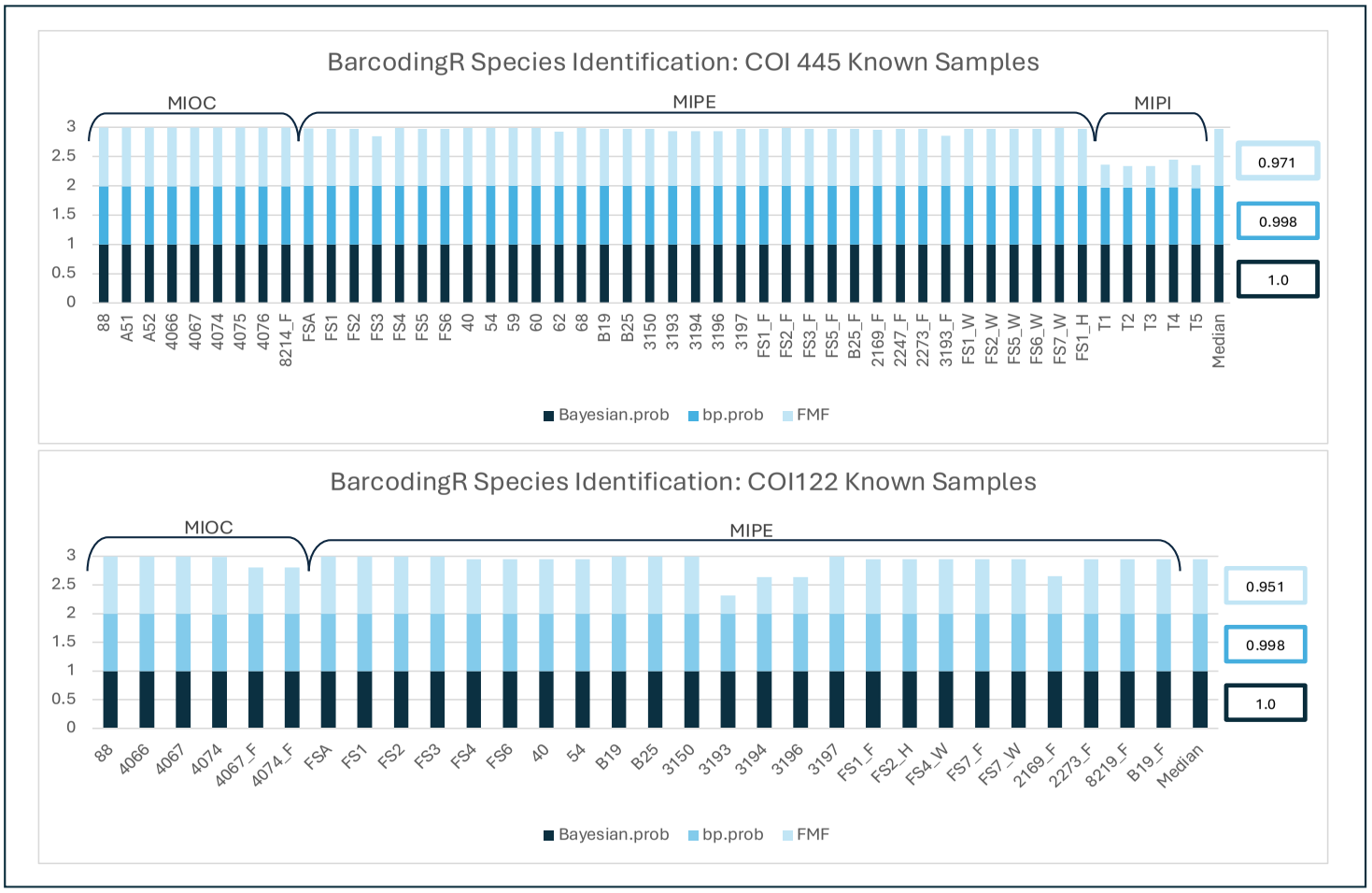

Figure 6. Results of three species identification methods for COI445 (above) and COI122 (below) known sample sequences including a Bayesian-based, BP-based, and fuzzy-set based methods, performed in the R package "BarcodingR". Each method produced a value of probability of assignment between 0.0 – 1.0. The median of these values is displayed on the righthand side. All COI445 and COI122 sequences returned species identification that matched their true species. Sequences are derived from tissue unless noted with an underscore followed by sample type, e.g. "FS1_F" is a fecal pellet from individual FS1 (F= Fecal Pellet, H = Fur, W = Whisker).

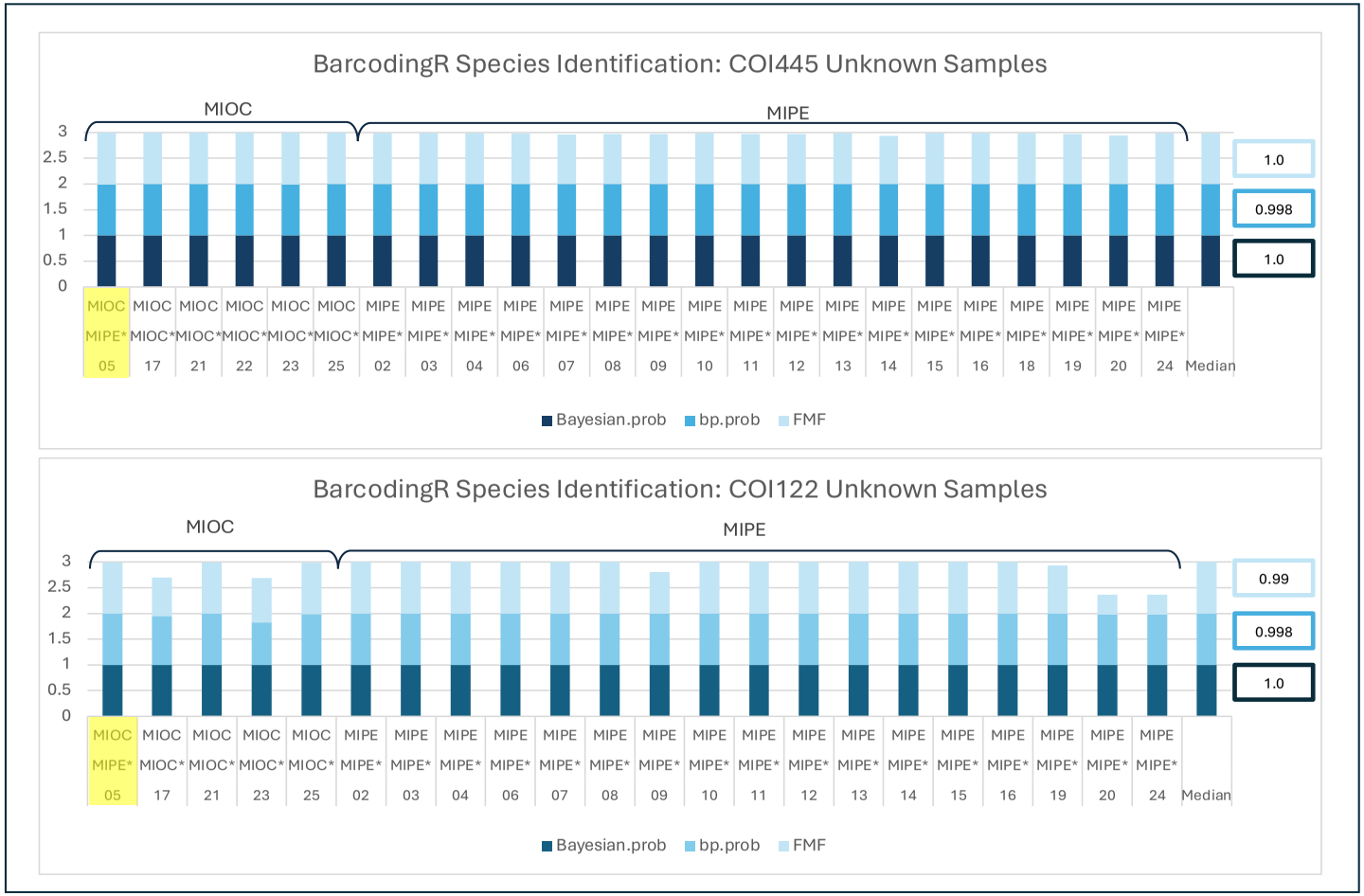

Figure 7. Results of three species identification methods for COI445 (above) and COI122 (below) unknown sample sequences including a Bayesian-based, BP-based, and fuzzy-set based methods, performed in the R package "BarcodingR". Each method produced a value of probability of assignment between 0.0 – 1.0. The median of these values is displayed on the righthand side. BarcodingR species designations are listed above WDNR species identifications that are marked with an asterisk. Sample 05 was the only sample that did not match its WDNR species identification, it was identified as *M. pennsylvanicus* in the field but as *M. ochrogaster* in BarcodingR, it is highlighted in yellow.

Genetic Analysis

Barcoding gap analysis showed a local gap between species for the BOLD reference sequences (n = 149) included in this study. Pairwise intraspecific genetic distances ranged from 0 – 6.8%, with a mean of 1.76%. Interspecific genetic distances ranged from 13.6% – 18.2%, with a mean of 15.6% (Figure 8). Overall mean interspecific distances were 8 times higher than mean intraspecific distances between species (Table 8). We visualized the local barcoding gap by plotting the difference between maximum intra- and minimum interspecific distances for each reference sequence using the R package 'spideR'. No individual reference sequence exhibited an intraspecific distance that was greater than its interspecific distances (Figure 9). We found that the BOLD standard 1% threshold for interspecific genetic distance for species identification was not sufficient for our specific dataset. Of the 149 reference sequences used in our study, more than half (59.2%) of pairwise intraspecific genetic distances were greater than 1%. The range for an optimal threshold is where our barcoding gap exists based on samples currently available in the database, with the least cumulative error was 6.81 - 12.85% (Figure 10).

Pairwise genetic differences between COI445 barcodes for all individuals, including known and unknown samples, was visualized as a heatmap (Figure 11) with unknown samples sorted by their barcode-assigned species. An NJ tree showed all of our samples clustered with their barcode-assigned species reference sequences (Figure 12). There was low consensus support between *M. pinetorum* and *M. ochrogaster* clades (51%), most likely caused by the smaller sample sizes of these two species in the dataset. We observed some substructure within species in both the heatmap and NJ tree. For example, T2-4 were collected in Tennessee and show higher similarity to each other than to T1 collected in Oklahoma and T5 collected in

Texas. Additionally, samples 2169, 3196, 3193, and 3194, which are known *M. pennsylvanicus* from Beery Lab, showed higher intraspecific differences in comparison to other *M. pennsylvanicus* samples and clustered together with 100% consensus support.

Table 8. Matrix of mean intraspecific (red) and mean interspecific (blue) pairwise genetic distances for BOLD reference sequences (n = 149).

Species	MIPE	MIOC	MIPI
MIPE	1.82%	15.83%	14.31%
MIOC	15.83%	1.36%	14.35%
MIPI	14.31%	14.35%	6.82%

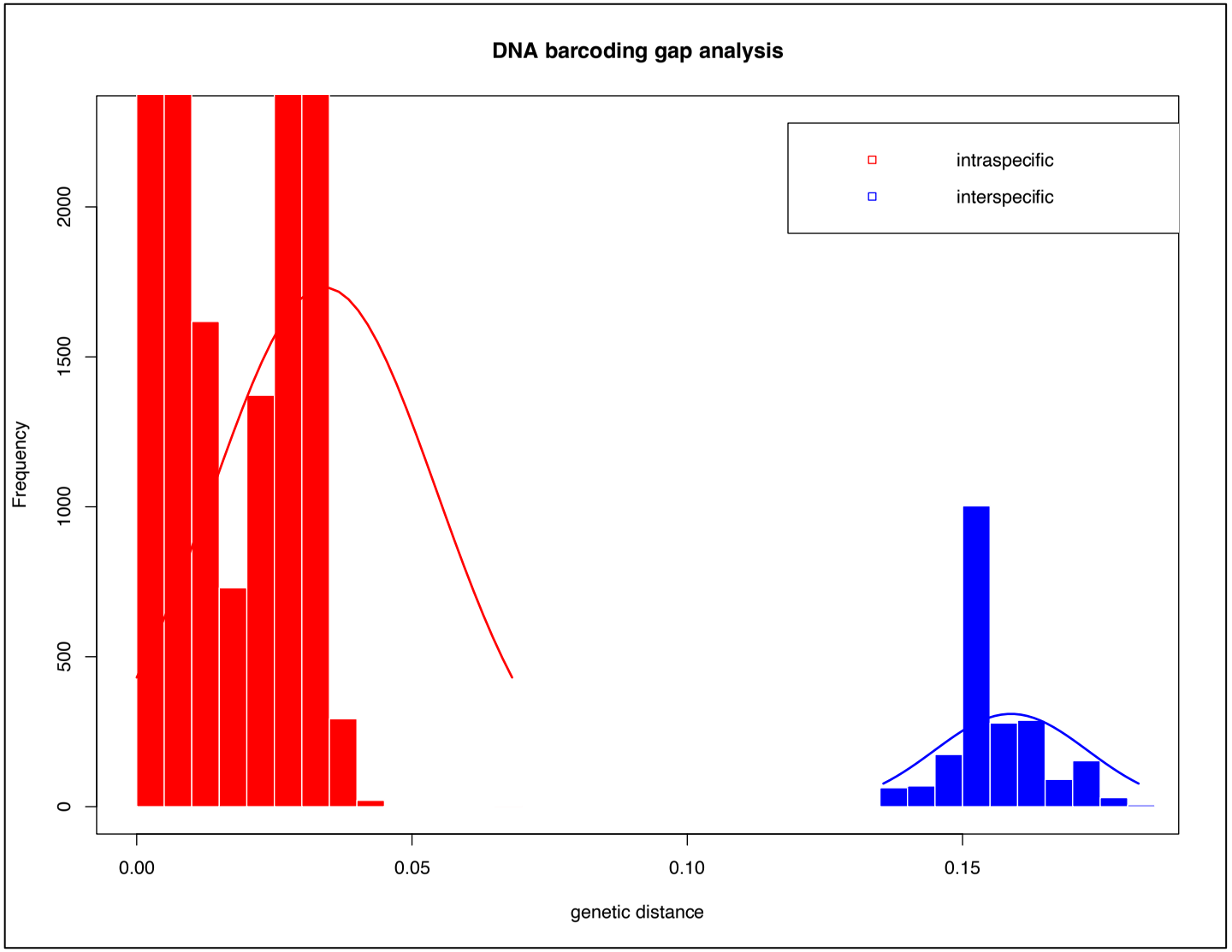

Figure 8. Barcoding gap analysis of BOLD reference sequences performed using BarcodingR. Intraspecific genetic distance ranged from 0.000 - 0. 0.068, with a mean of 0.0176. Interspecific genetic distance ranged from 0.136 – 0.182, with a mean of 0.156.

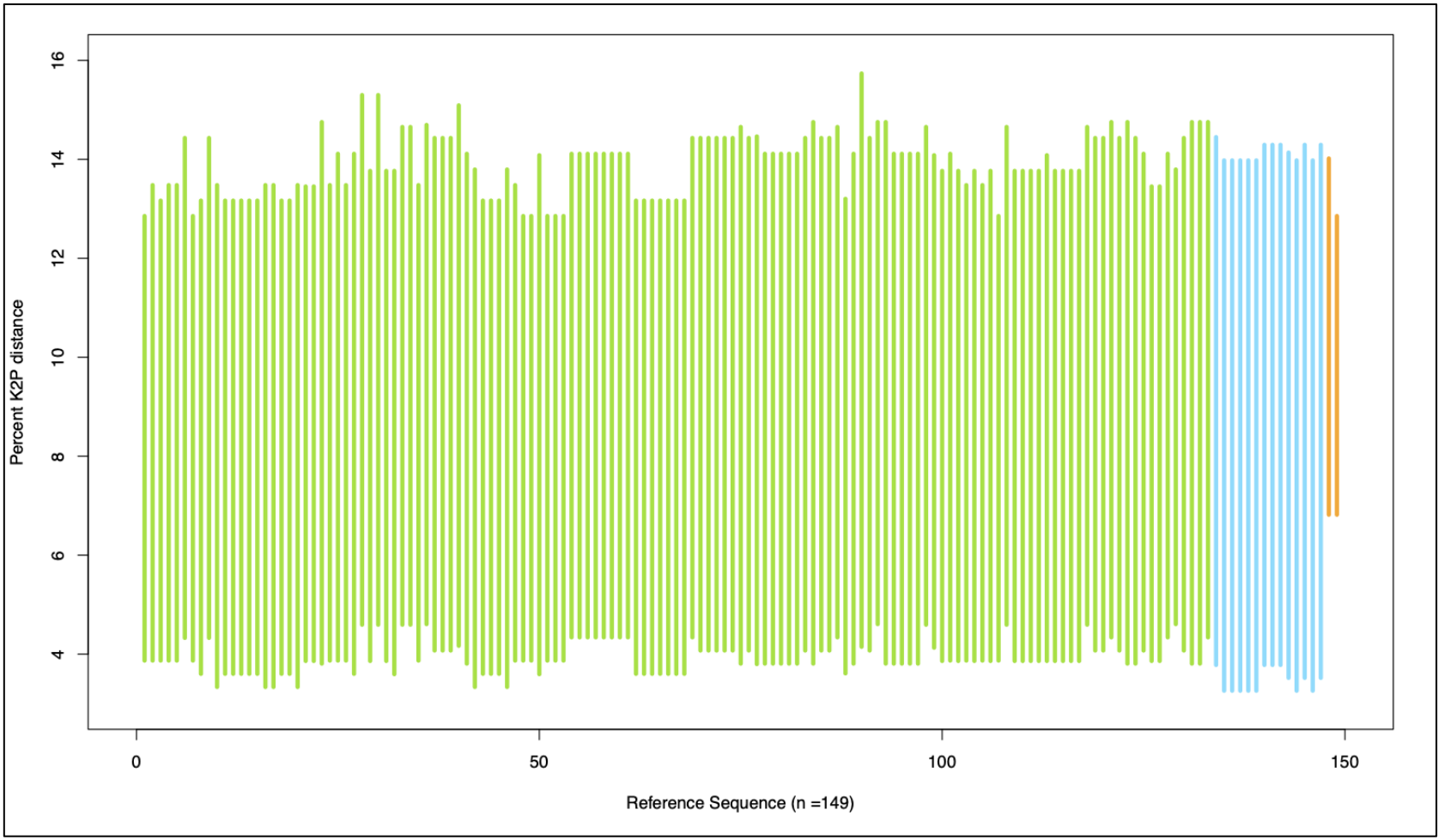

Figure 9. Line plot of the three *Microtus* species reference sequences (n = 149; bp = 410). Each line represents one individual reference sequence. Reference sequences from MIPE are in green, MIOC in blue, MIPI in orange. The maximum intraspecific distance (bottom of each bar) and the minimum interspecific distance (top of each bar) visualizes the size of the barcoding gap of each individual. All had a minimum interspecific distance that was greater than the maximum intraspecific distance, indicating the presence of a local barcoding gap across our reference sequences.

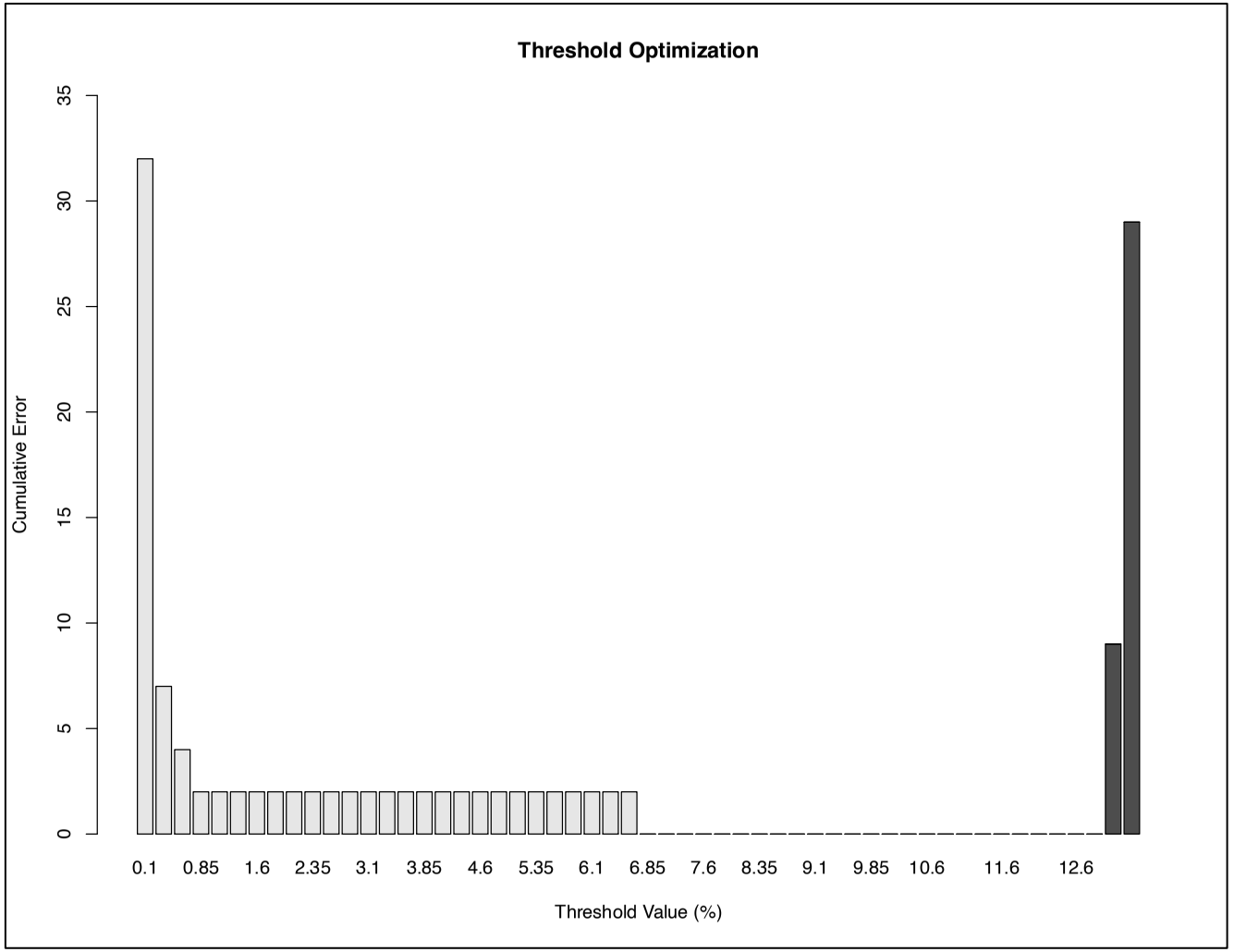

Figure 10. Bar plot showing false positive (light grey) and false negative (dark grey) rates of identification of Wisconsin's three *Microtus* species for different species identification threshholds.

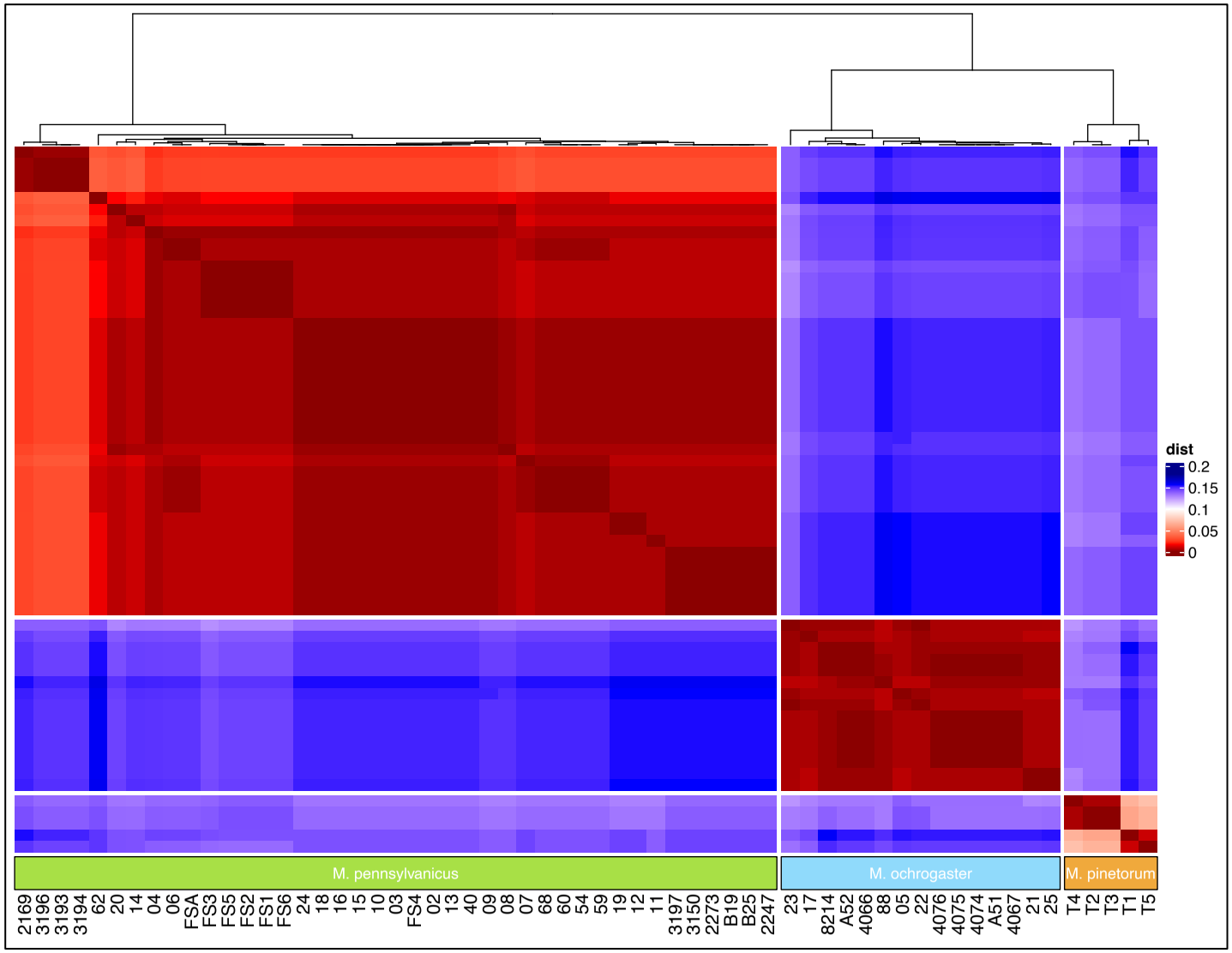

Figure 11. Heatmap of pairwise K2P genetic distances of COI445 query sequences. The Heatmap has been split by known species and are titled above the sample IDs. The dendrogram (left) clusters groups based on pairwise genetic distance.

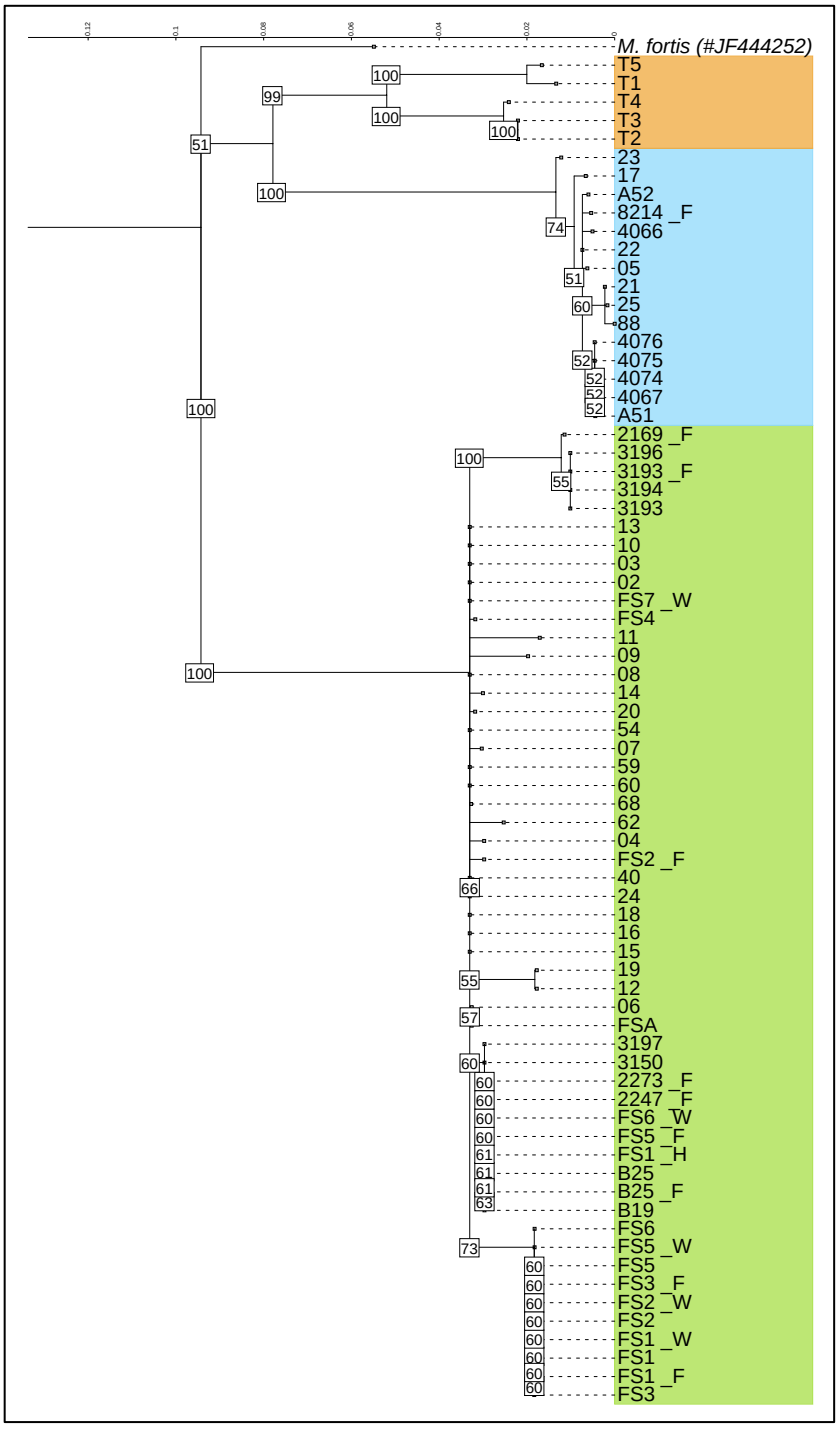

Figure 12. NJ phylogenetic consensus tree of samples included in this study constructed using an alignment of 410bp portion of the *COI* gene. *M. fortis* was chosen as an outgroup based on phylogeny reported in Jackson et. al. (2020). Node values represent bootstrap support consensus (%) from 1000 replicates. Branches with less than 50% support were collapsed. Species clades are boxed by color: *M. pinetorum* = orange, *M. ochrogaster* = blue, *M. pennsylvanicus* = green.

Discussion

Conservation strategies rely on accurate species identification. In Wisconsin's three *Microtus* species, traditional field-based species identification can be difficult due to overlapping morphological traits. Additionally, there is an interest in approaches that do not require trapping or specimen preparation of species of conservation concern. We evaluated two distinct DNA-based approaches to species identification for this system. The HY approach had not been previously tested on females, on voles in Wisconsin, or on *M. pinetorum*. Encouragingly, we found no evidence for geographic- or sex-based differences in success or accuracy. However, we confirmed our initial prediction that *M. pinetorum* would produce the longer (~720bp) fragment of the *avpr1a* gene like *M. ochrogaster*, rendering the HY approach ineffective for distinguishing *M. ochrogaster* and *M. pinetorum.* The HY approach could be effective in species identification of *M. pennsylvanicus* if it is used in areas where *M. pennsylvanicus* is sympatric with just one of the other two species (Figure 2). By contrast, the DNA barcoding approach was successful in distinguishing among all three species.

COI445 primers (85.9%) had higher successful amplification rates than COI122 primers (62.4%). Initially, we predicted that shorter amplicons would more successfully amplify, especially for non-invasive samples. However, the sequences generated from COI122 primers often had weak and noisy signal traces resulting in shorter forward and reverse sequences that did not overlap, including all COI122 barcodes for *M. pinetorum* museum samples (n = 5). PCR optimization, including adjusting cycling conditions and/or PCR reagent concentrations could potentially improve sequencing results for COI122 primers (Hajibabaei, et al., 2005), but primer

redesign might also be necessary to successfully sequence *M. pinetorum*. Without adequate reference sequences from one or more of the target species, designing a single set of primers that successfully amplifies products from all species can be elusive.

Tissue samples, expectedly, had higher successful amplification than non-invasively collected samples in both the HY and DNA barcoding approaches. Further refinement of our methods to improve success rates in non-invasive samples would provide two main advantages for future applications. First, collection of whisker, hair, and fecal pellet samples requires animals to be trapped and/or handled, but minimizes disturbance to our target species, two of which are SGCN in the state. Second, non-invasive genetic sampling saves time and effort in the field compared to morphology-based species identification and tissue collection (Zemanova, 2021). PCR amplification success from fecal DNA extractions has been improved by increasing PCR reaction volumes (Verkuil, et al., 2018), sampling the outermost pellet material (Wehausen, et al., 2004), and increasing DNA yield in extraction using adjusted protocols (Costa, 2017; Zhang, et al., 2006). Though our study's non-invasive sample amplification success rates were lower than anticipated, barcodes that were sequenced from non-invasive samples demonstrated 100% accuracy in identifying species, matching the performance of tissue-derived barcodes.

There were advantages for each of the two analytical approaches we evaluated for DNA barcoding, NCBI's BLAST search and BarcodingR. BLAST searches were easy to complete and did not require the collection, alignment, and analysis of reference sequences that were necessary for BarcodingR. However, the BLAST search approach required interpretation to draw conclusions, for example if multiple species were returned among the top hits. This

interpretation could be biased, for example if a putative species is over- or under-represented in the GenBank database (Munch, et al., 2008; Spouge and Mariño-Ramírez, 2012). In contrast, BarcodingR produced assignment statistics that facilitated objective species identification, though these too are affected by species representation (Zhang, et al., 2017).

Among the three species identification statistics generated by BarcodingR, the BP-based and Bayesian-based methods returned assignment values with less variation among individuals in our dataset than the fuzzy-based method. This trend stems from the theoretical difference among the three statistics. In terms of the underlying assignment algorithm, the BP-based and Bayesian-based statistics return a probability of assignment for each query sequence to each set of reference sequences, taken as the likelihood that an individual belongs to a species. In contrast, the fuzzy-based statistic returns a membership function that defines how each input reference sequence is mapped to a degree of membership. The three statistics also differ in how they consider uncertainty, which is especially relevant to studies with low or uneven reference species representation. Fuzzy inference considers uncertainty and lack of precision for each species designation directly, by comparing each query sequence to the reference sequences. Bayesian-based relies on the assumption of a prior distribution of sequence variation between species and the probability of a query sequence belonging to one or none of those species and does not consider uncertainty (Kościelny, et al., 2021; Nielsen and Matz, 2006). The BP-based method makes no *a priori* assumptions on the frequency distributions of nucleotide variants for each species, as it uses back-propagation neural networks, a form of artificial intelligence trained by the input of reference sequences. Each nucleotide base and its position in reference sequences is weighted and adjusted as more sequences are read.

Simultaneously query sequences are read through the same process and values of probability of species assignment are returned (Zhang, et al., 2008). All three methods correctly assigned each known sample to its true species, but the fuzzy-based method highlighted the uncertainty in assigning M. pinetorum samples, likely due to a weak reference sequence database for this species. Combining the results of all three methods and making species assignments when there is agreement across approaches is an approach we favor, in agreement with other studies that have used these approaches (Wibowo, et al., 2021; Carugati, et al., 2022; Zheng, et al., 2022).

Analysis of the reference sequences used in this study revealed a local barcoding gap between the species pairs. Our barcoding gap was large and far exceeded the 1% threshold recommended by BOLD, but our barcoding gap is also biased by the availability of reference sequences. Lack of reference sequences usually leads to underestimation of intraspecific genetic distance (Phillips, et al., 2019), which can lead to overestimation of interspecific genetic distances and an overly large barcoding gap (Meyer and Paulay, 2005). However, *M. pinetorum* ran counter to this expectation, with only two references but the highest mean intraspecific genetic distance of the three species (6.82%). This value represents the genetic distance between two *M. pinetorum* individuals rather than serving as an estimate for the overall intraspecific variation within *M. pinetorum.* The two individuals, BOLD Records MAMN8542-23 and MAMN8263-22, were sampled in Kansas and Texas, respectively. The geographic distance between these individuals may explain the elevated genetic distance. Mean intraspecific distances for *M. pennsylvanicus* (1.82%) and *M. ochrogaster* (1.36%) are likely a slight underestimation compared to expected K2P genetic distances of the *COI* gene in similar

species. Nicolas, et al. (2012) found a mean intraspecific distance of 2.03% in Praomyini, a tribe of muroid rodents and Çetintürk, et al. (2021) found average genetic distances between geographic regions of 2.17% in *M. arvalis*. The inclusion of additional sequences in reference databases, which expands both the number of individuals and the geographic coverage, better captures the full range of genetic variation within a species and thus tends to lead to an increase in intraspecific distances and smaller barcoding gaps (Bertolazzi, et al., 2009; Lou and Golding, 2012; Clare et al., 2007). If, however, limited reference sampling spans large geographic distances, as we found with *M. pinetorum*, sampling additional reference species sampling might be predicted to decrease mean intraspecific genetic distances and increase the barcoding ga. Regardless of the shift direction, the end result of additional reference species sampling will be more accurate representation of genetic variation within species, a more accurate estimation of the barcoding gap, and improved DNA barcoding. More reference sequences for sister species will also improve the resolution of genetic distances in Wisconsin's *Microtus* species, as is the case for most species (Gostel and Kress, 2022). But as we have shown in our study, despite under-sampling, DNA barcoding remains an effective means of species identification.

Morphology-based species identification was 96% accurate on the 25 voles sampled from sympatric habitats in Wisconsin. The one vole misidentified in the field as the more common species *M. pennsylvanicus* was consistently identified as the Wisconsin SGCN *M. ochrogaster* in all of our genetic approaches. Species misidentifications that underestimate the population and/or fail to confirm the presence of a species of concern can impair conservation efforts (Costa, et al. 2015; Hodges and Kirchhoff, 2012; Peres, et al., 2021). Effective species

identification becomes especially important in systems like Wisconsin's three *Microtus* species, where rare, threatened, and endangered species overlap in morphology with more common species DNA barcoding of the *COI* gene provided accurate species identification for all three species and its application will aid in the WDNR's *M. ochrogaster* and *M. pinetorum* conservation efforts. DNA barcoding does not replace morphology-based species identification nor negate the importance of taxonomy, rather, it offers an effective solution to challenging species identification. If used in conjunction, DNA barcoding and morphometric analysis could be used to define a DNA-based taxonomy framework among the three species (Vogler and Monaghan, 2007) and potentially be extended to other similar vole species. It is encouraging that the DNA barcoding approach worked well despite a paucity of reference sequences for two of the three species, suggesting that such limitations can be overcome to incorporate DNA-based approaches into species monitoring and the implementation of well-informed conservation strategies.

References

Adams, et al. 2017. “Range-wide microsatellite analysis of the genetic population structure of prairie voles (*Microtus ochrogaster*),” *The American Midland Naturalist* 177, 183-199. https://doi.org/10.1674/0003-0031-177.2.183

Ambrose, 1973. “An experimental study of some factors affecting the spatial and temporal activity of *Microtus* pennsylvanicus,” *Journal of Mammalogy*, 54: 79-110. https://doi.org/10.2307/1378874

Becker, et al. 2011. “Five years of FISH-BOL: brief status report,” *Mitochondrial DNA*, 22(S1), 3-9. https://doi.org/10.3109/19401736.2010.535528

Bertolazzi, et al. 2009. “Learning to Classify Species with Barcodes,” *BMC Bioinformatics,* 10(S14): S7. https://doi.org/10.1186/1471-2105-10-s14-s7

Brown, et al. 2012. “SPIDER: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding.” *Molecular Ecology Resources*, 12, 562–565. Version 1.5.0 https://doi.org/10.1111/j.1755-0998.2011.03108.x

Carugati, et al. 2022. "Combined *COI* barcode-based methods to avoid mislabeling of threatened species of deep-sea skates," *Animal Conservation,* 25(1):38-52. https://doi.org/10.1111/acv.12716

Cassola, 2016. (a) “*Microtus ochrogaster*,” *The IUCN Red List of Threatened Species* 2016: e.T42631A115196932. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T42631A22347375.en

Cassola. 2016. (b) “*Microtus pinetorum*,” *The IUCN Red List of Threatened Species* 2016: e.T42633A115197344. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T42633A22346051.en

Cassola. 2016. (c) “*Microtus chrotorrhinus* (errata version published in 2017)” *The IUCN Red List of Threatened Species* 2016: https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T42626A22347958.enran

Çetintürk, et al. 2021. "Inferring phylogenetic relationships in the common vole (*Microtus arvalis*) based on mitochondrial and nuclear sequence diversities," *Turkish Journal of Zoology*, 45(2):5. https://doi.org/10.3906/zoo-2008-3

Clare, et al., 2007. “DNA barcoding of Neotropical bats: species identification and discovery within Guyana,” *Molecular Ecology Notes*, 7: 184-190. https://doi.org/10.1111/j.1471-8286.2006.01657.x

Collins and Cruickshank. 2013. “The seven deadly sins of DNA barcoding,” *Molecular Ecology Resources,* 13, 969-975 https://doi.org/10.1111/1755-0998.12046

Costa, et al. 2015. "Impacts of Species Misidentification on Species Distribution Modeling with Presence-Only Data." *ISPRS International Journal of Geo-Information*, 4(4)2496 – 2518. https://doi.org/10.3390/ijgi4042496

Costa, et al. 2017. "Improving DNA quality extracted from fecal samples—a method to improve DNA yield,' *European Journal of Wildlife Research,* 63(3). https://doi.org/10.1007/s10344-016-1058-1

Dayrat. 2005. "Towards integrative taxonomy," *Biological Journal of the Linnean Society*, 85(3):407–417, https://doi.org/10.1111/j.1095-8312.2005.00503.x

DeSalle, et al. 2017. "MtDNA: The small workhorse of evolutionary studies," *Frontier in Bioscience (Landmark Ed),* 22(5), 873–887. https://doi.org/10.2741/4522

Everson, et al. 2023. "Disentangling morphology and genetics in two voles (*Microtus pennsylvanicus* and *M. ochrogaster*) in a region of sympatry," *Journal of Mammalogy,* 104 (3), 532–545. https://doi.org/10.1093/jmammal/gyac119

FitzGerald and Madison. 1983. "Social organization of a free-ranging population of pine voles, *Microtus pinetorum*," *Behavioral Ecology and Sociobiology* 13, 183–187. https://doi.org/10.1007/BF00299921

Ford, et al. 2007. "The land manager's guide to mammals of the south: meadow Vole, *Microtus pennsylvanicus*" *The Nature Conservancy*, Durham, North Carolina, USA. 284–288. https://doi.org/10.1093/bioinformatics/btw313

Geneious Prime 2023.2.1. https://www.geneious.com

Gostel and Kress. 2022. "The Expanding Role of DNA Barcodes: Indispensable Tools for Ecology, Evolution, and Conservation," *Diversity*, 14(3) 213. https://doi.org/10.3390/d14030213

Gu, et al. 2016. "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, 32(8), 2847 – 2849. https://doi.org/10.1093/bioinformatics/btw313

Hajibabaei, et al. 2005."Critical Factors for Assembling a High Volume of DNA Barcodes," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462):1959-1967. https://doi.org/10.1098/rstb.2005.1727

Hammock and Young. 2005. "Microsatellite instability generates diversity in brain and sociobehavioral traits," *Science*, 308(5728):1630 – 1634. https://doi.org/10.1126/science.1111427

Hall. 1981. *The Mammals of North America.* 2nd ed. Wiley 1981.

Hebert, et al. 2003. "Biological identifications through DNA barcodes," *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 270(1512), 313-321.

Hebert and Gregory. 2005. “The promise of DNA barcoding for taxonomy,” *Systematic Biology* 54: 852-859. https://doi.org/10.1080/10635150500354886

Henterly, et al. 2011. "Comparison of morphological versus molecular characters for discriminating between sympatric meadow and prairie voles," *The American Midland Naturalist* 165(2), 412-420. https://doi.org/10.1674/0003-0031-165.2.412

Hodges and Kirchhoff. 2012. "Kittlitz's murrelet *Brachyramphus Brevirostris* population trend in Prince William Sound, Alaska: Implications of species misidentification," *Marine Ornithology,* 40(2):7. https://digitalcommons.usf.edu/marine_ornithology/vol40/iss2/7

Hubert, et al. 2008. “Identifying Canadian freshwater fishes through DNA barcodes,” *PLOS ONE* 3(6): e2490. https://doi.org/10.1371/journal.pone.0002490

Jackson and Cook. 2019. “A precarious future for distinctive peripheral populations of meadow voles (*Microtus pennsylvanicus*),” Journal of Mammalogy, 101(1) 36-51. https://doi.org/10.1093/jmammal/gyz196

Jackson and van Aarde. 2003. “Sex- and species-specific growth patterns in cryptic African rodents, *Mastomys natalensis* and *M. coucha*,” *Journal of Mammalogy*, 84(3):851–860. https://doi.org/10.1644/BPR-001

Kingsley, et al., 2017. "The ultimate and proximate mechanisms driving the evolution of long tails in forest deer mice." *Evolution*, 71(2): 261-273 https://doi.org/10.1111/evo.13150

Kirkland and Farré. 2021. “Mitochondrial genome evolution, genetic diversity, and population structure in British water voles (Arvicola amphibius),” *Genes.* 12(2):138. https://doi.org/10.3390/genes12020138

Kościelny, et al. 2021. "Diagnosing with a hybrid fuzzy–Bayesian inference approach," *Engineering Applications of Artificial Intelligence,* 104: 104345. https://doi.org/10.1016/j.engappai.2021.104345

Ladoukakis & Zouros. 2016. “Evolution and inheritance of animal mitochondrial DNA: rules and exceptions,” *Journal of Biological Research-Thessaloniki.* 24, 2. https://doi.org/10.1186/s40709-017-0060-4

Laerm and Ford. 2007. “The land manager’s guide to mammals of the south: prairie vole *Microtus ochrogaster*” *The Nature Conservancy*, Durham, North Carolina, USA. 280–283.

Lee and Beery. 2022. “Selectivity and sociality: aggression and affiliation shape vole social relationships,” *Frontiers in Behavioral Neuroscience*. 16:8. https://doi.org/10.7554/eLife.72684

Letunic and Bork. 2021. "Interactive tree of life (ITOL) V5: An online tool for phylogenetic tree display and annotation," *Nucleic Acids Research*, 49(W1): W293 – 296. https://doi.org/10.1093/nar/gkab301

Lou and Golding. 2012. "The Effect of Sampling from Subdivided Populations on Species Identification with DNA Barcodes Using a Bayesian Statistical Approach," *Molecular Phylogenetics and Evolution*, 65(2)765-773. https://doi.org/10.1016/j.ympev.2012.07.033

Meiklejohn, et al. 2019. "Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials" *PLOS ONE* 14(6): e0217084. https://doi.org/10.1371/journal.pone.0217084

Meyer and Paulay. 2005. "DNA barcoding error rates based on comprehensive sampling," *PLOS Biology* 3 e422. https://doi.org/10.1371/journal.pbio.0030422

Meusnier, et al. 2008. "A universal DNA mini-barcode for biodiversity analysis," *BMC Genomics* 9, 214. https://doi.org/10.1186/1471-2164-9-214

Miller, et al. 2010. "Size-associated morphological variation in the red tree vole (*Arborimus longicaudus*)." *Northwestern Naturalist*, 91(1):63–73. https://www.jstor.org/stable/i40038298

Munch, et al. 2008. "Statistical assignment of DNA sequences using Bayesian phylogenetics," *Systematic Biology*, 57(5):750757. https://doi.org/10.1080/10635150802422316

Nekrutenko, et al. 2000. "Isolation of binary species-specific PCR-based markers and their value for diagnostic applications," *Gene,* 249(1-2):47-51. https://doi.org/10.1016/S0378-1119(00)00168-2

Nicolas, et al. 2012. "Assessment of Three Mitochondrial Genes (16S, Cytb, CO1) for Identifying Species in the Praomyini Tribe (Rodentia: Muridae)," *PLOS ONE*, 7(5): e36586. https://doi.org/10.1371/journal.pone.0036586

Nielsen and Matz. 2006. "Statistical Approaches for DNA Barcoding," *Systematic Biology*, 55(1)162-169. https://doi.org/10.1080/10635150500431239

Paradis and Schliep. 2019. "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R," *Bioinformatics*, 35:526-528. doi:10.1093/bioinformatics/bty633.

Peres, et al. 2021. "Implications of Unreliable Species Identification Methods for Neotropical Deer Conservation Planning," *Perspectives in Ecology and Conservation*, 19(4):435-442. https://doi.org/10.1016/j.pecon.2021.08.001

Phillips, et al. 2019. "Incomplete Estimates of Genetic Diversity within Species: Implications for DNA Barcoding," *Ecology and Evolution*, 9(5):2996-3010, https://doi.org/10.1002/ece3.4757

Phillips, et al. 2022. "Lack of statistical rigor in DNA barcoding likely invalidates the presence of a true species' barcode gap," *Frontiers in Ecology and Evolution,* 10, 859099. https://doi.org/10.3389/fevo.2022.859099

Ratnasingham and Hebert. 2007. "BOLD: The Barcode of Life Data System (www.barcodinglife.org)," *Molecular Ecology Notes* 7, 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Sayers, et al. 2022. "Database resources of the national center for biotechnology information," *Nucleic Acids Research* 50(D1): D20-D26. https://doi.org/10.1093/nar/gkab1112

Schwartz and Schwartz. 2001. "The wild mammals of Missouri," *University of Missouri Press*, Columbia, Missouri, USA.

Smolen. 1981. "Mammalian species: *Microtus pinetorum*" *The American Society of Mammologists* 147, 1-7. https://doi.org/10.1093/mspecies/147.1.

Spouge and Mariño-Ramírez. 2012. "The practical evaluation of DNA barcode efficacy," *Methods in Molecular Biology,* 858:365-77. https://doi.org/10.1007%2F978-1-61779-591-6_17

Tyagi, et al. 2019. "Identification of Indian spiders through DNA barcoding: cryptic species and species complex," *Scientific Reports* 9, 14033. https://doi.org/10.1038/s41598-019-50510-8

Wisconsin Department of Natural Resources. 2013. "Wisconsin woodland vole species guidance" Bureau of Natural Heritage Conservation, Wisconsin Department of Natural Resources, Madison, Wisconsin. PUB-ER-691.

Wisconsin Department of Natural Resources. 2015. "2015-2025 Wisconsin Wildlife Action Plan" Madison, Wisconsin. PUB-NH-938.

Wibowo, et al. 2021. "Assessing temporal patterns and species composition of glass eel (*Anguilla* spp.) cohorts in Sumatra and Java using DNA barcodes," *Diversity*, 13(5):193. https://doi.org/10.3390/d13050193

Untergasser, et al. 2012. "Primer3—new capabilities and interfaces" *Nucleic Acids Research* 40(15): e115. https://doi.org/10.1093/nar/gks596

Verkuil, et al. 2018. "Molecular identification of temperate *Cricetidae* and *Muridae* rodent species using fecal samples collected in a natural habitat," *Mammal Research,* 63:379–385. https://doi.org/10.1007/s13364-018-0359-z

Vogler and Monaghan. 2006. "Recent advances in DNA taxonomy," *Journal of Zoological Systematics and Evolutionary Research,* 45(1), 1-10.

https://doi.org/10.1111/j.1439-0469.2006.00384.x

Waits and Paetkau. 2005. “Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection," *The Journal of Wildlife Management*, 69(4):1419-1433. https://doi.org/10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2

Wehausen, et al. 2004. “Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method,” *Journal of Heredity*, 95(6):503–509. https://doi.org/10.1093/jhered/esh068

Zemanova. 2021. "Noninvasive genetic assessment is an effective wildlife research tool when compared with other approaches." *Genes*, 12(11). https://doi.org/10.3390/genes12111672

Zhang, et al. 2012. “A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding,” *Molecular Ecology,* 21(8):1848-63. https://doi.org/10.1111/j.1365-294x.2011.05235.x

Zhang, et al. 2006. “A widely applicable protocol for DNA isolation from fecal samples,” *Biochemical Genetics*, 44:494–503. https://doi.org/10.1007/s10528-006-9050-1

Zhang, et al. 2017. “BarcodingR: an integrated r package for species identification using DNA barcodes,” *Methods in Ecology & Evolution,* 8: 627-634. https://doi.org/10.1111/2041-210X.12682

Zheng, et al. 2022. "The identification of fish eggs from four economically important species in Guanghai Bay (China)," *Frontiers in Marine Science*, 9:815473. https://doi.org/10.3389/fmars.2022.815473