# Using environmental (e)DNA techniques to better understand aquatic ecosystems


**Kim Scribner**

**and**

**John Robinson**

**Department of Fisheries and Wildlife**
**Michigan State University**




*email:*
*scribne3@msu.edu*

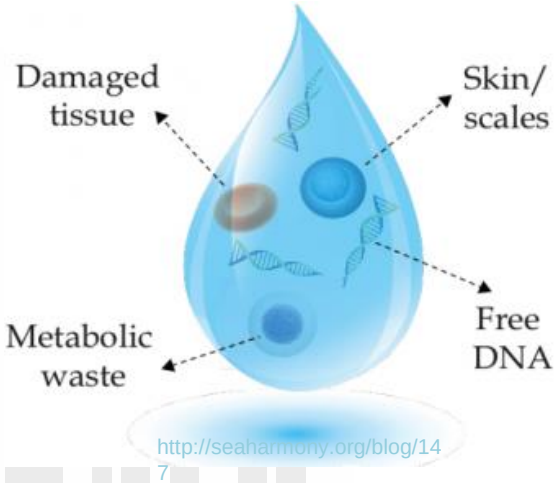
# Road map for talk

Part I – Lessons learned and nuances of single species and multiple species eDNA methodology for aquatic species

Part II – Applications mostly focusing on aquatic plant and animal multi-species methods or ‘amplicon sequencing’ (metabarcoding)

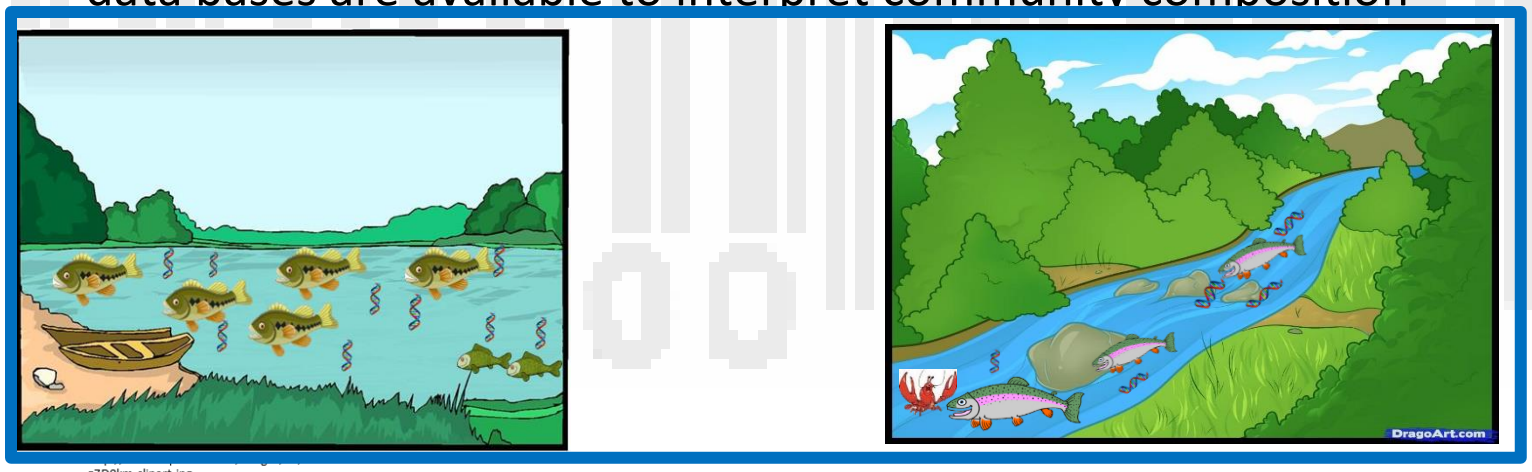
# What is environmental (e)DNA and how is it used?

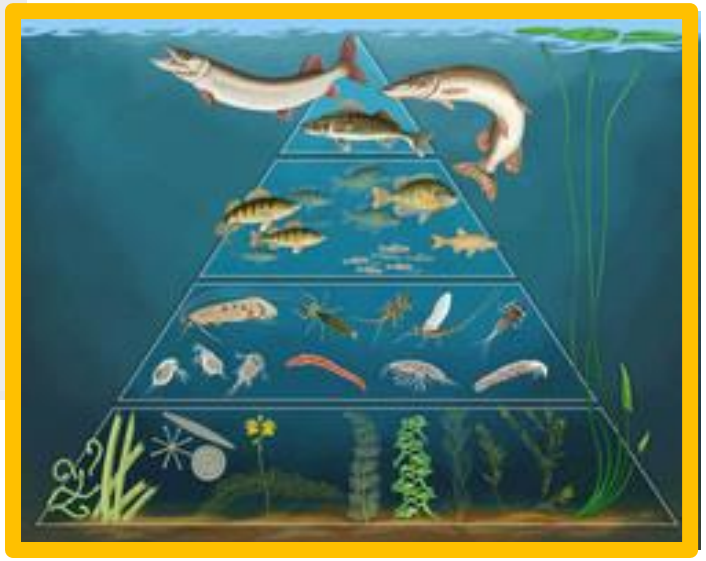
- All living organisms leave traces of cell debris and extracellular DNA in their environment
- Can be used as a quantitative tool to estimate **species** *presence/absence* and *relative biomass* across sampling sites, including **AIS** and **T&E** species
- eDNA is an effective *non-invasive method* to quantify **community** composition and diversity
- One sample contains information about the whole community – from fish to invertebrates to plants to micro-organisms
- Physical and biotic aquatic *lake-/river-scape* and terrestrial data bases are available to interpret community composition




http://seaharmony.org/blog/147

https://www.estuarydna.org\





# Environmental (e)DNA in Environmental Sciences

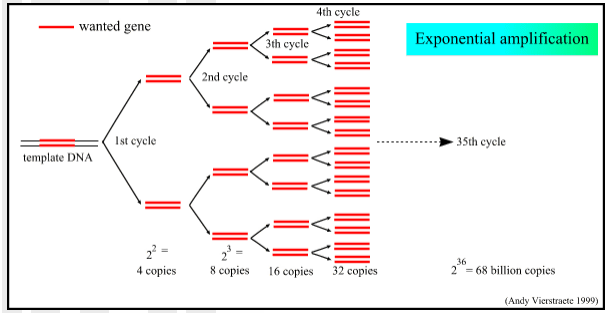
## Step 1 – obtain environmental sample

Burchfieldpenny.org

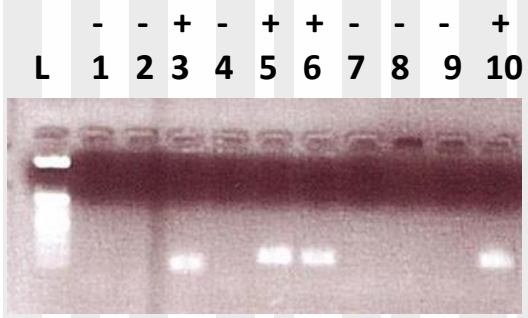
or

or

## Step 2 – PCR amplification of DNA




## Step 3 – screen DNA for presence or taxonomic composition




Modified from Herder et al. 2014

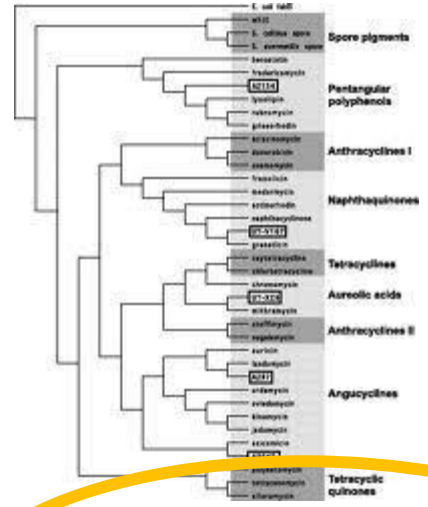
Visualization of amplification products

=


Single Species Analyses

Species (yes/no) and 'relative abundance'

or



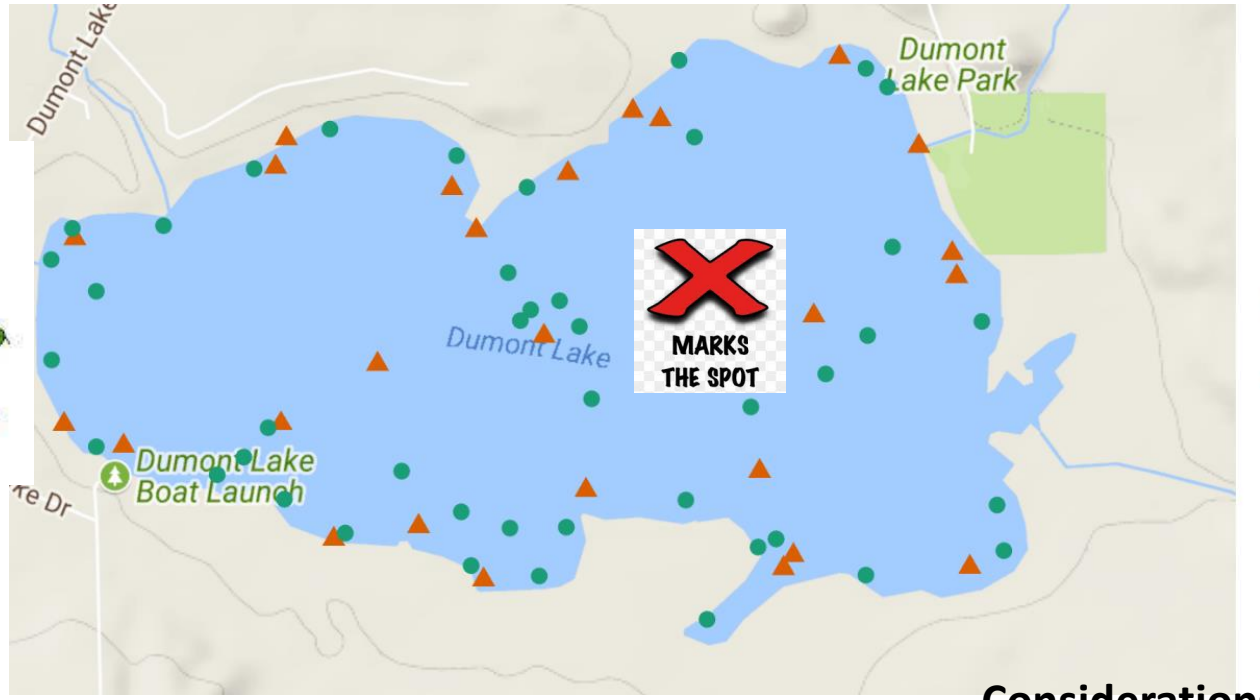

Multiple Species Analyses

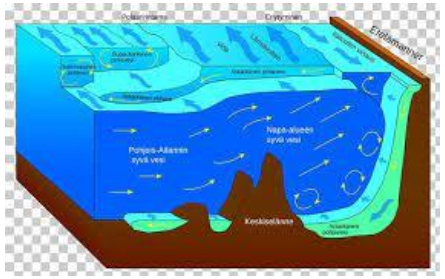
- Community composition
- Species relative abundance

Questions

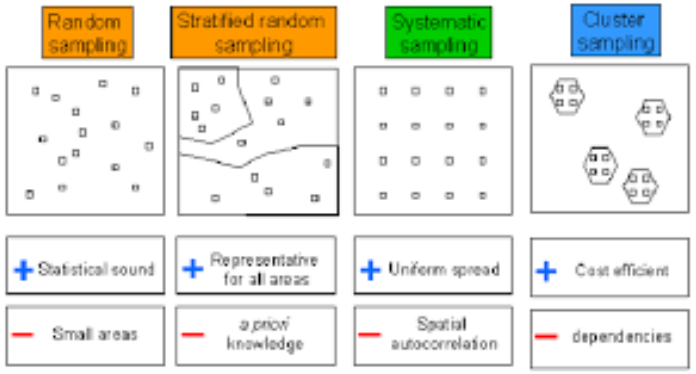
# Factors to consider when designing an aquatic eDNA study --- forces that determine amount and location of eDNA in an aquatic system

## Factors affecting eDNA location



How to develop a sampling design?
Where and how many?
What are study objectives?




## Factors affecting eDNA deposition and loss

Gear type
- eDNA
- Traditional

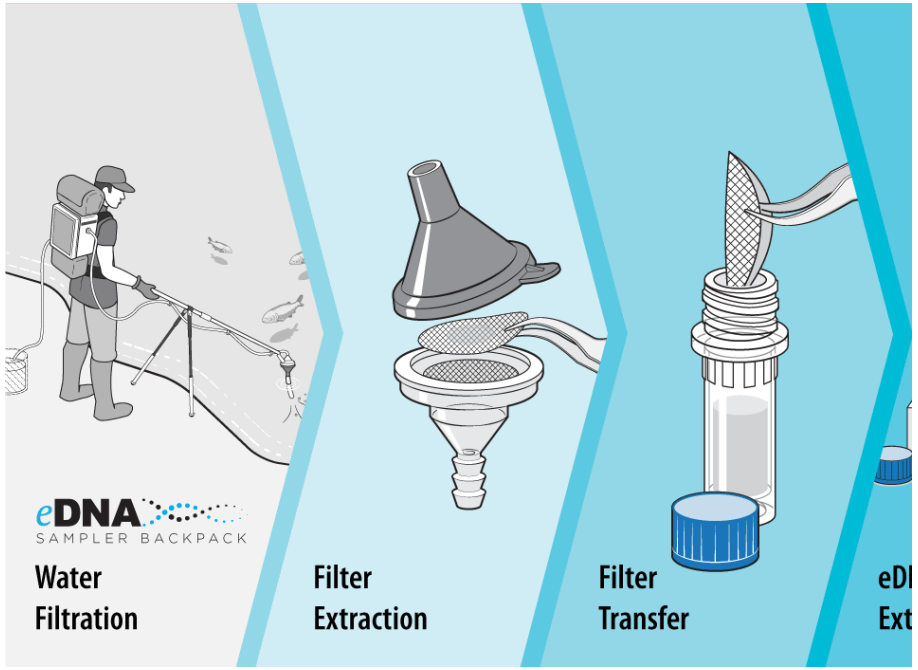
**Considerations**

- Lake area
- Depth profile - surface and benthic samples
- Terrestrial sources and surrounding land-use
- Season of the year
- Filter characteristics

# Multiple methods of sample collection

+

or



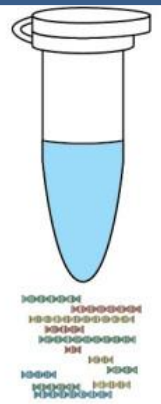

https://www.smith-root.com/edna/edna-sampler

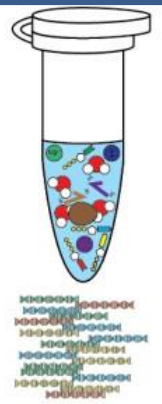
# eDNA QA/QC needs: antiseptic methods and verification


1 Collect an environmental sample

2 DNA extraction from environmental sample

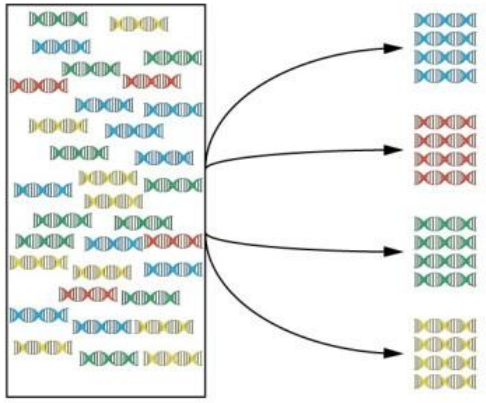
3 Amplify DNA markers





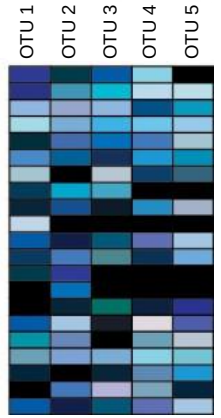
4 High-throughput sequencing

5 Bioinformatic processing



6 * Species identification

7 Ecological analysis




https://mresbec.wordpress.com/2017/10/26/saving-the-world-from-alien-invasion-early-detection-with-edna/

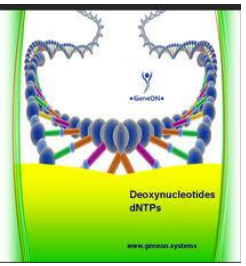
# The Power of PCR

# How can you obtain detectable quantities of DNA when starting quantities are low?


Isolate total RNA or mRNA

**Step 1**
**- Extract DNA**
**- Remove inhibitors**

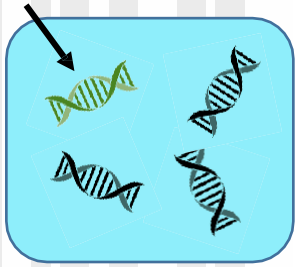

**Step 2**
**- PCR**



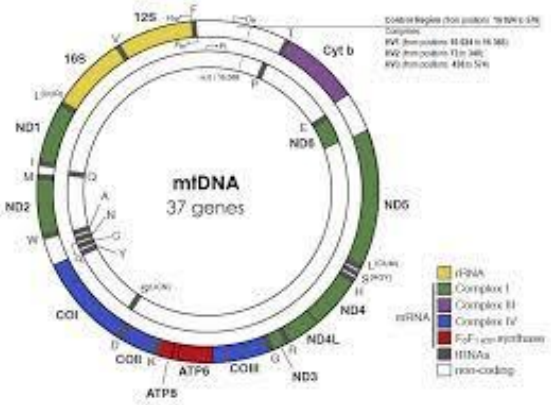

**+** Enzyme


DifferenceBetween.com

Option 1 - Target 1 species DNA

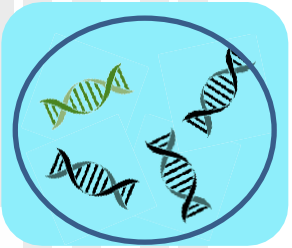


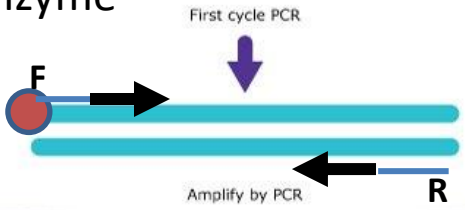
**F** **R**



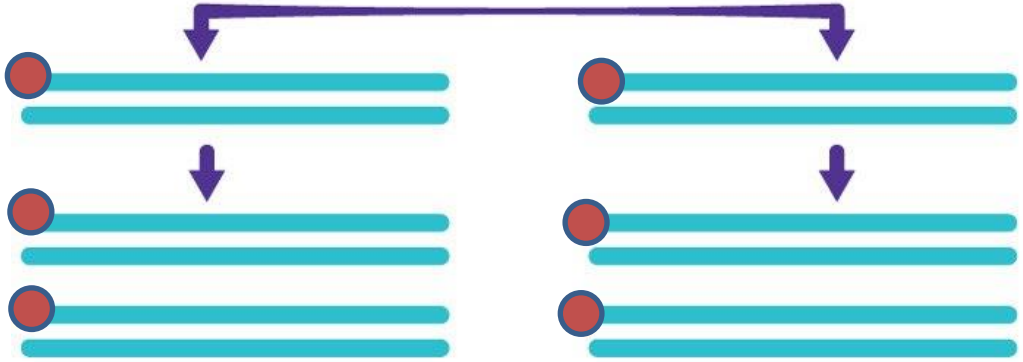

Option 2 - Target all species DNA

**F** **R**






Cycle 1
(1 to 2 copies)



Cycle 2
(2 to 4 copies)

Sequencing primer

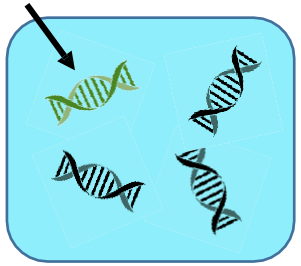
Bar-code

PCR product from eDNA

High throughput sequencing

# PCR or qPCR allows detection of a *specific* target species

Target species DNA



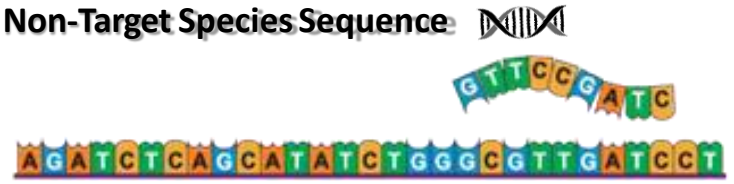
Water sampling & eDNA isolation

DNA sequence collection & comparison

Design of species specific primers (& probe if desired)

Testing for species specificity of primers/probe

**Non-Target Species Sequence**



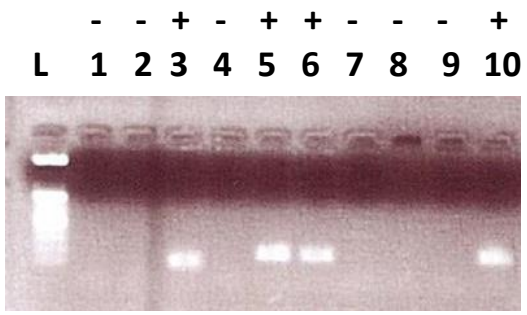

Mismatched sequence = No Amplification Product

**Target Species Sequence**

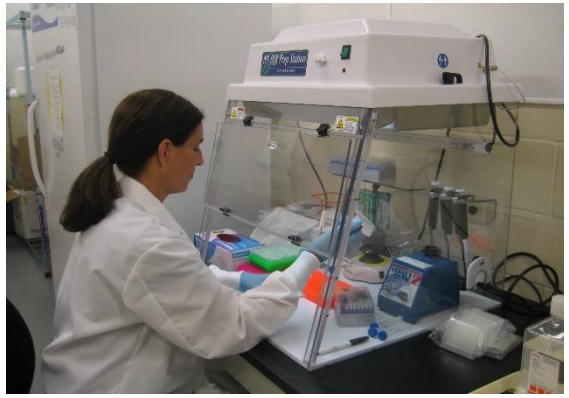

Matching sequence = Amplification Product

Modified from www.gmotesting.com

Sample scoring: Only samples 3, 5, 6 & 10 show evidence of presence of target species DNA




Modified from Herder et al. 2014

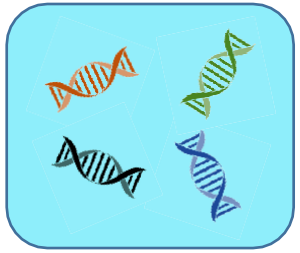


From www.vet.upenn.edu

Amplification of eDNA samples using PCR or qPCR

Visualization of amplification products

# Metabarcoding allows detection of multiple target species



Water sampling & eDNA isolation

DNA sequence collection & comparison

Design of "Universal" Primers

Testing for ability of selected gene region(s) to amplify & discriminate all target species

Simultaneous PCR amplification of all target DNAs in each eDNA sample

Amplicon library prep for each eDNA sample

Data indexing, cleaning, filtering & bioinformatic analysis

Comparison of eDNA sample sequence data to reference sequence database

Typically done *"in silico"*

| Species/Abbrv | ********** |
|---|---|
| 1. Alosa pseudoharengus | TTAGATACCCCACTATGC |
| 2. Pomoxis nigromaculatus | TTAGATACCCCACTATGC |
| 3. Lepomis cyanellus | TTAGATACCCCACTATGC |
| 4. Lepomis macrochirus | TTAGATACCCCACTATGC |
| 5. Lepomis gibbosus | TTAGATACCCCACTATGC |
| 6. Pimephales notatus | TTAGATACCCCACTATGC |
| 7. Etheostoma nigrum | TTAGATACCCCACTATGC |
| 8. Percina maculata | TTAGATACCCCACTATGC |
| 9. Ameiurus nebulosus | TTAGATACCCCACTATGC |
| 10. Ameiurus melas | TTAGATACCCCACTATGC |
| 11. Ameiurus natalis | TTAGATACCCCACTATGC |
| 12. Ictalurus punctatus | TTAGATACCCCACTATGC |
| 13. Salvelinus fontinalis | TTAGATACCCCACTATGC |
| 14. Salmo trutta | TTAGATACCCCACTATGC |
| 15. Oncorhynchus mykiss | TTAGATACCCCACTATGC |
| 16. Cyprinus carpio | TTAGATACCCCACTATGC |
| 17. Notropis atherinoides | TTAGATACCCCACTATGC |
| 18. Notemigonus crysoleucas | TTAGATACCCCACTATGC |
| 19. Cyprinella spiloptera | TTAGATACCCCACTATGC |
| 20. Catostomus catostomus | TTAGATACCCCACTATGC |
| 21. Catostomus commersoni | TTAGATACCCCACTATGC |
| 22. Micropterus salmoides | TTAGATACCCCACTATGC |
| 23. Micropterus dolomieu | TTAGATACCCCACTATGC |
| 24. Umbra limi | TTAGATACCCCACTATGC |
| 25. Dorosoma cepedianum | TTAGATACCCCACTATGC |
| 26. Lepisosteus osseus | TTAGATACCCCACTATGC |
| 27. Amia calva | TTAGATACCCCACTATGC |
| 28. Acipenser fulvescens | TTAGATACCCCACTATGC |

= A. fulvescens

= S. trutta

= C. commersonii

= L. osseus

ID of target species' DNA sequences in each sample

# Important issues for single species eDNA barcoding

- **Specificity** – establishing confidence that the exact species (and no other) is being amplified

  *Goal – minimize false positives (amplification of wrong species) --- typically quantified across multiple DNA sample concentrations*

- **Sensitivity** – the levels (quantities) of DNA that can be consistently detected

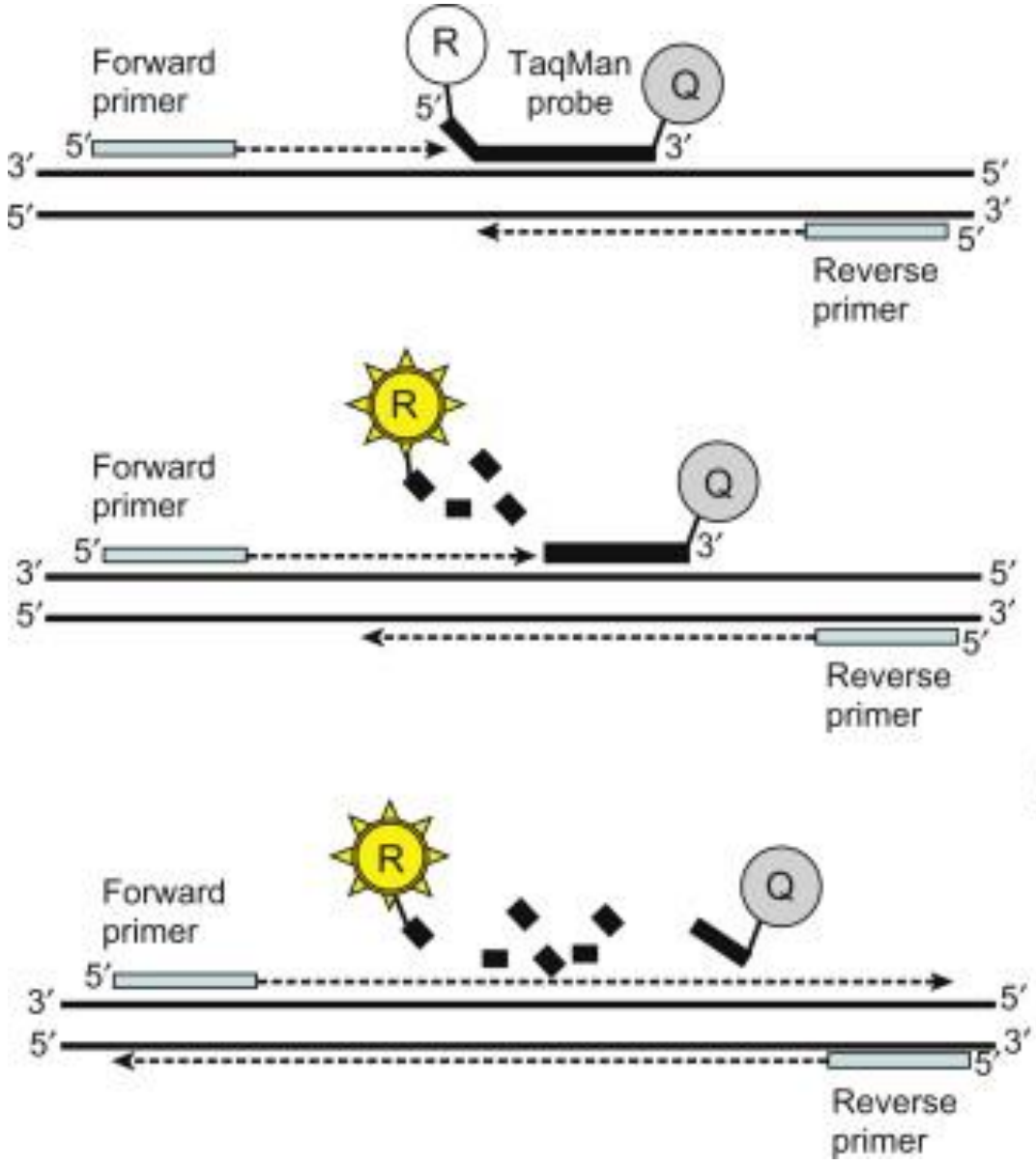
  *Goal – minimize false negatives (failure to amply DNA when present but in low concentration) ------ typically quantified using qPCR*

***Limits of detection (LOD)*** *- LOD is defined as the lowest copy number where 95% of the replicates per concentration were positive.*

***Limits of quantification (LOQ)*** *– LOQ represents the lowest concentration of target DNA that can be quantified within an assay. Klymus et al. (2019).*

- *Protocols should be evaluated in the context of the environmental conditions expected in the field (e.g., with inhibitors, etc present)*
- *Protocols should include tests for contamination (e.g., with no DNA controls)*

# Taqman assay – a more sensitive *single species* PCR assay




Polymerization and Strand Displacement

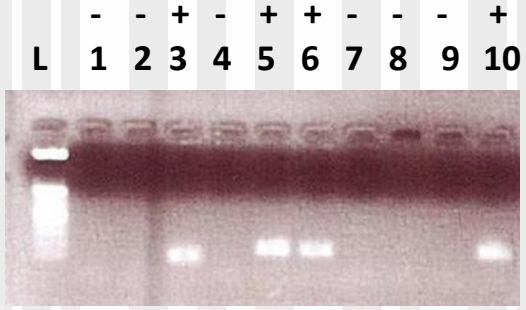
Probe Cleavage (release of reporter dye)

*Fluorescence occurs when reporter dye and quencher dye are no longer in close proximity*

Completion of Polymerization

science.direct

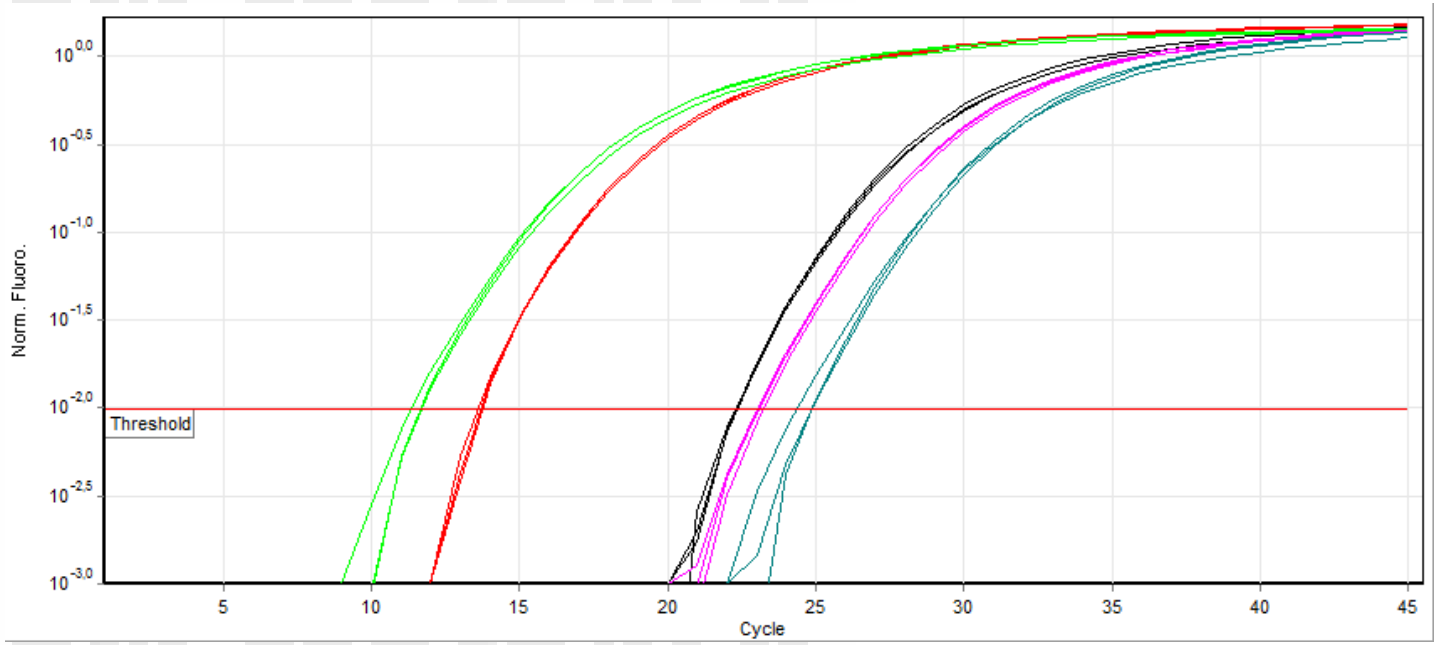
# Visualization of PCR products for single species detections (yes/no)




Modified from Herder et al. 2014

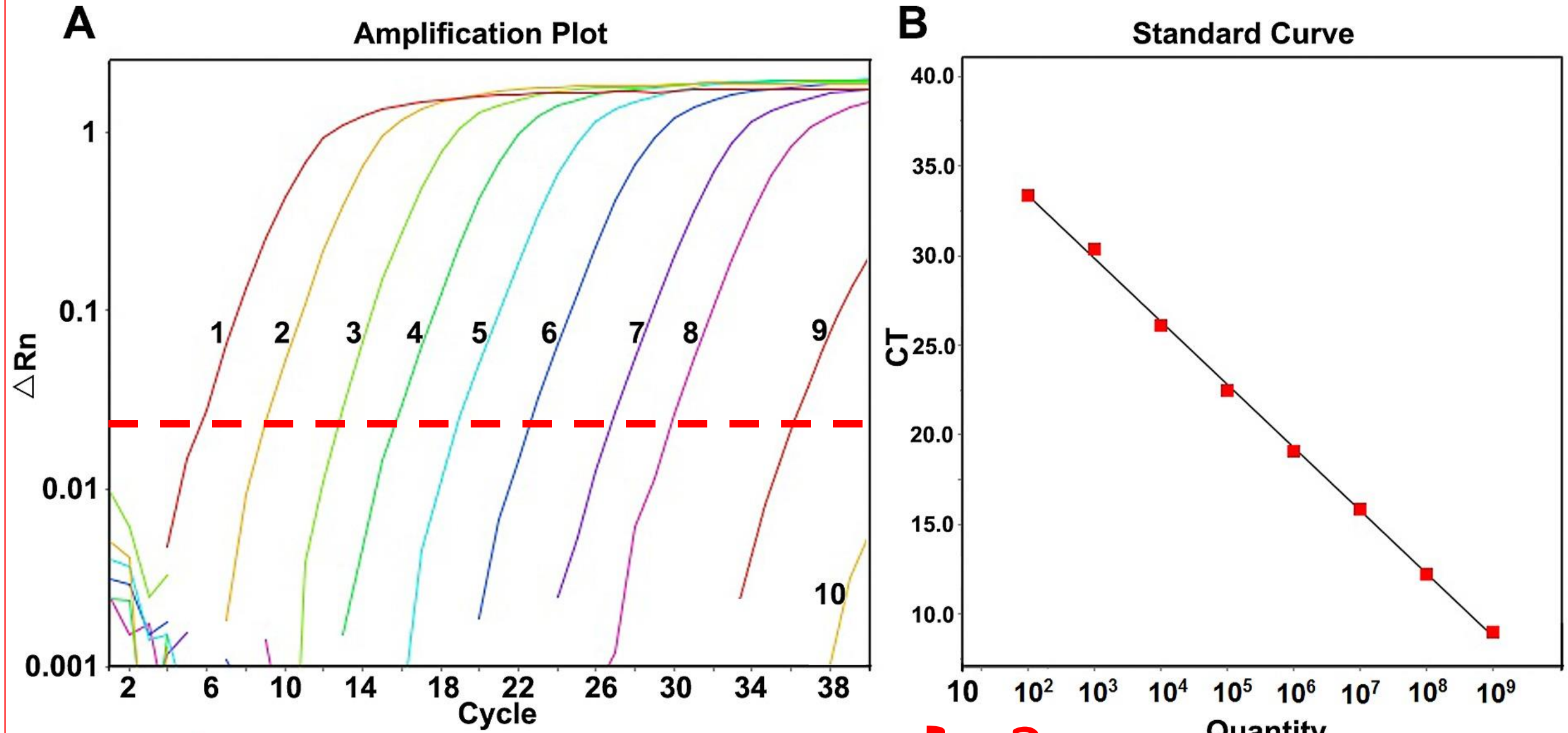
Visualization of amplification products using ‘staining’ and gel electrophoresis

**More sensitive taqman qPCR reactions**



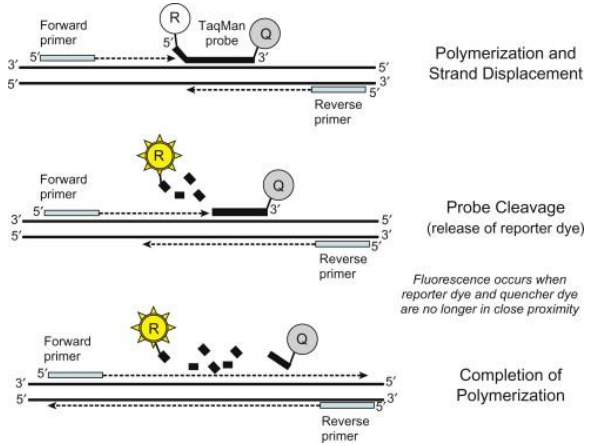

Higher Initial DNA copy number

Lower Initial DNA copy number


A
Amplification Plot
△Rn
1
0.1
0.01
0.001
1
2
3
4
5
6
7
8
9
10
2
6
10
14
18
22
26
30
34
38
Cycle
B
Standard Curve
CT
40.0
35.0
30.0
25.0
20.0
15.0
10.0
10
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹
Quantity
1: 1×10⁹ copies/μl
2: 1×10⁸ copies/μl
3: 1×10⁷ copies/μl
4: 1×10⁶ copies/μl
5: 1×10⁵ copies/μl
6: 1×10⁴ copies/μl
7: 1×10³ copies/μl
8: 1×10² copies/μl
9: 1×10¹ copies/μl
10: no-template control
?
(1) PCR product length,
(2) nucleotide composition

Results of grass carp qPCR – 3 samples, each assayed with 4 replicates of each of 3 probes. Positive results are given as positive (not UD) CT values……..Question-----Is there a difference in ‘reliability’ between positive Results across the 3 samples?

| Rep | Sample1 | probe | dye | CT | Sample2 | probe | dye | CT | Sample3 | probe | dye | CT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | DR0718-02 | GCTM10 | VIC | UD | DR0818-01 | GCTM10 | VIC | UD | HP1018-01 | GCTM10 | VIC | 38.959 |
| 2 | DR0718-02 | GCTM10 | VIC | UD | DR0818-01 | GCTM10 | VIC | UD | HP1018-01 | GCTM10 | VIC | UD |
| 3 | DR0718-02 | GCTM10 | VIC | UD | DR0818-01 | GCTM10 | VIC | 40.238 | HP1018-01 | GCTM10 | VIC | UD |
| 4 | DR0718-02 | GCTM10 | VIC | UD | DR0818-01 | GCTM10 | VIC | UD | HP1018-01 | GCTM10 | VIC | 40.885 |
| 1 | DR0718-02 | GCTM22 | FAM | UD | DR0818-01 | GCTM22 | FAM | UD | HP1018-01 | GCTM22 | FAM | 36.123 |
| 2 | DR0718-02 | GCTM22 | FAM | UD | DR0818-01 | GCTM22 | FAM | UD | HP1018-01 | GCTM22 | FAM | UD |
| 3 | DR0718-02 | GCTM22 | FAM | 37.082 | DR0818-01 | GCTM22 | FAM | UD | HP1018-01 | GCTM22 | FAM | UD |
| 4 | DR0718-02 | GCTM22 | FAM | UD | DR0818-01 | GCTM22 | FAM | UD | HP1018-01 | GCTM22 | FAM | UD |
| 1 | DR0718-02 | GCTM32 | NED | UD | DR0818-01 | GCTM32 | NED | UD | HP1018-01 | GCTM32 | NED | 38.895 |
| 2 | DR0718-02 | GCTM32 | NED | 39.294 | DR0818-01 | GCTM32 | NED | UD | HP1018-01 | GCTM32 | NED | UD |
| 3 | DR0718-02 | GCTM32 | NED | 38.209 | DR0818-01 | GCTM32 | NED | UD | HP1018-01 | GCTM32 | NED | UD |
| 4 | DR0718-02 | GCTM32 | NED | UD | DR0818-01 | GCTM32 | NED | UD | HP1018-01 | GCTM32 | NED | 38.300 |

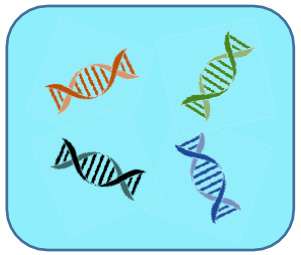



Primers developed by USFWS – LaCrosse lab

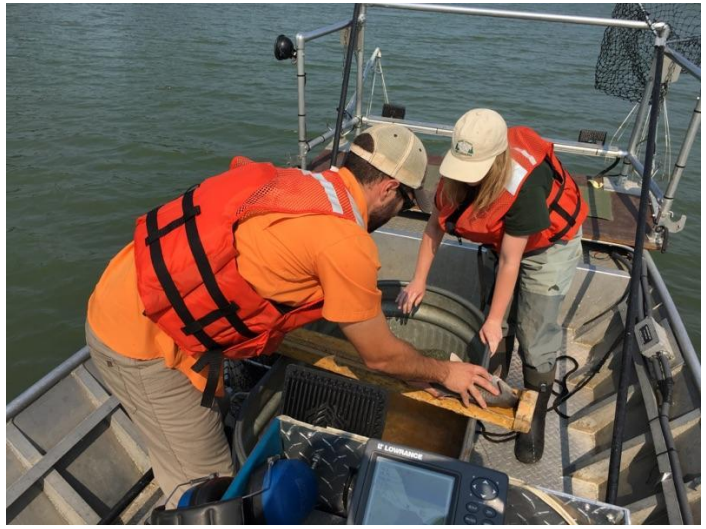
Bopp et al. (in review)

# Single species detection using eDNA-- a field application for grass carp in Lake Erie

Boat Electro-Fishing or boat Electro-Fishing-trammel net combination (“combo”)
Collected species including surrogate species
Common carp (COC) Bigmouth (BMBF)
and smallmouth buffalo (SMBF)
10 eDNA samples/site/month



eDNA



N=81 eDNA detections

N=0 fish captured

Bopp et al. (in review)

# How to assemble a DNA sequence baseline data base?
# First – Determine what species we can expect to observe

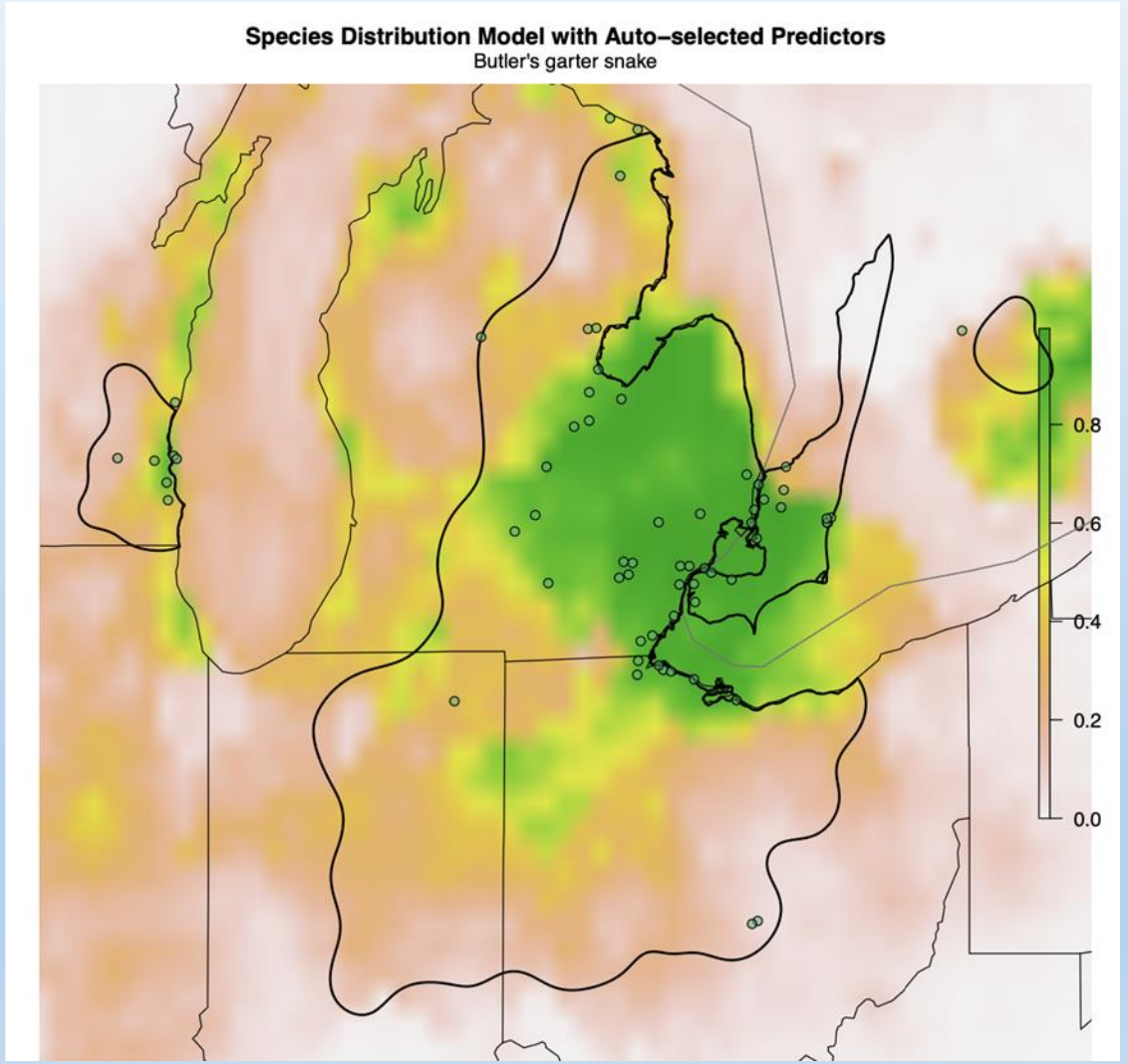
VerNet

GBIF

IUCN RED LIST

USGS
science for a changing world

PRISM CLIMATE GROUP

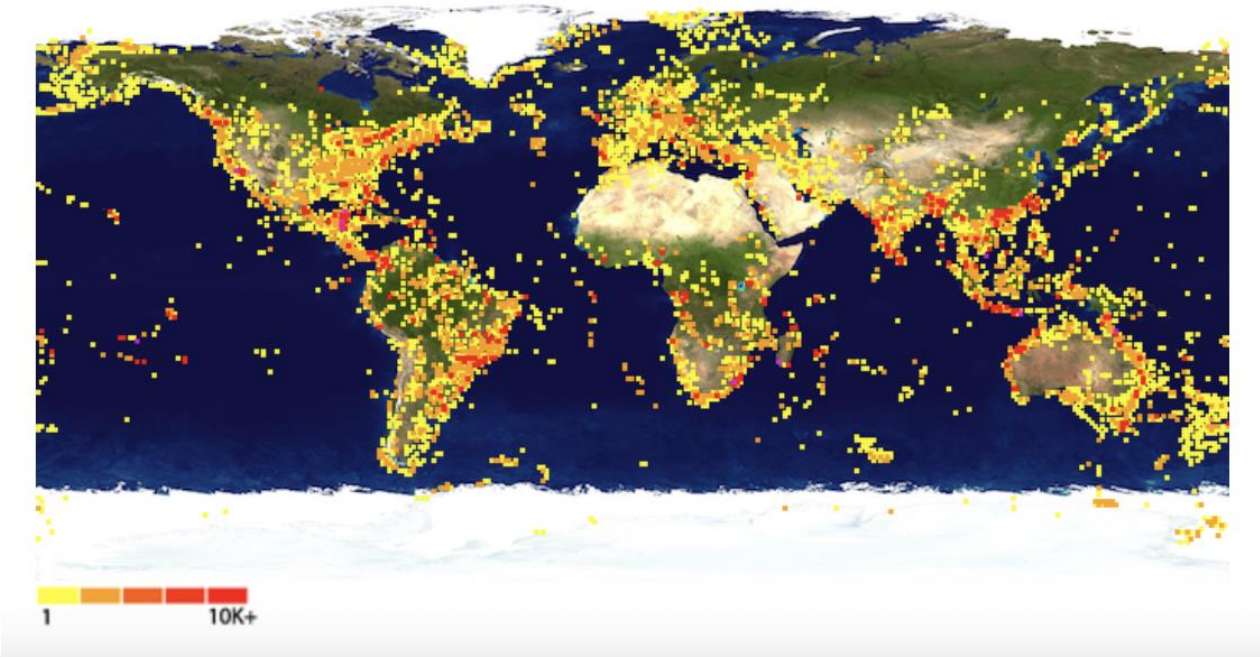


# Considerations for eDNA Metabarcoding

Second, DNA marker choice is critical

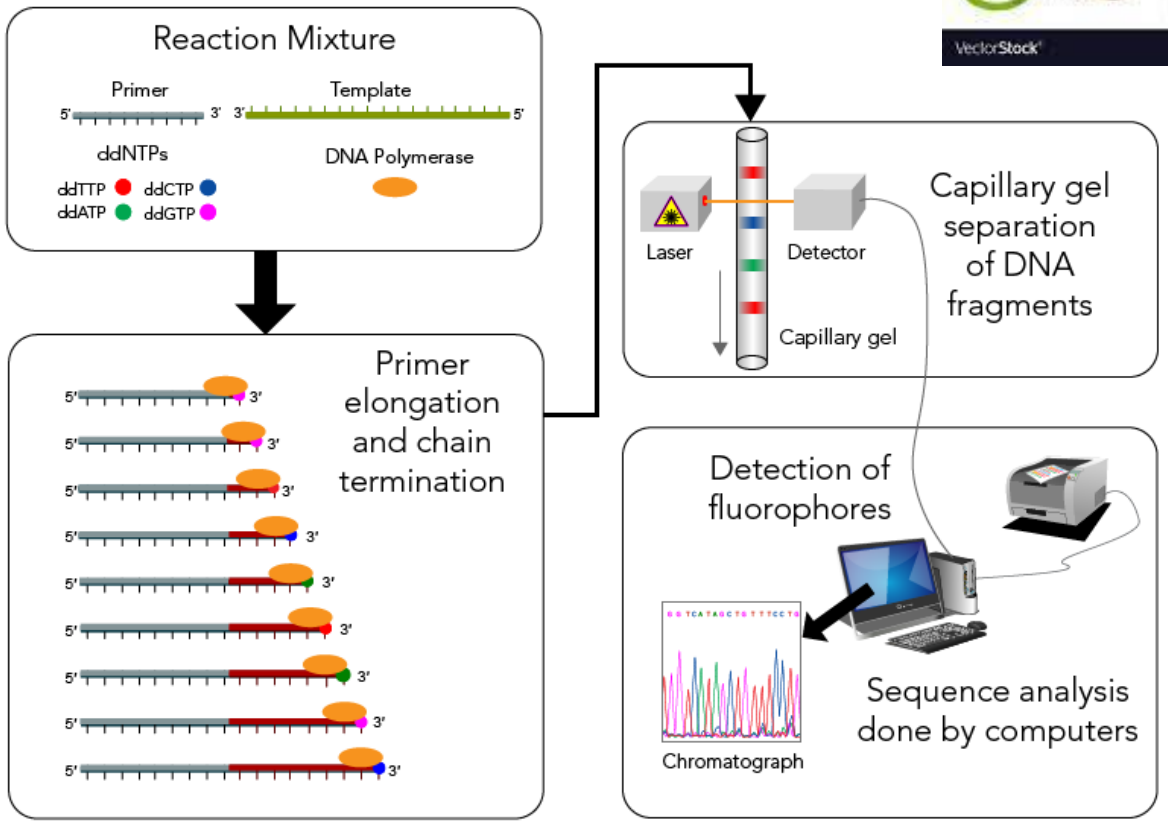
- Short gene fragments likely persist longer in the environment
- Long fragments allow for more species-level identification
- Need a ‘base-line’ database of sequences at the chosen marker (BOLD, GenBank)

BOLD SYSTEMS

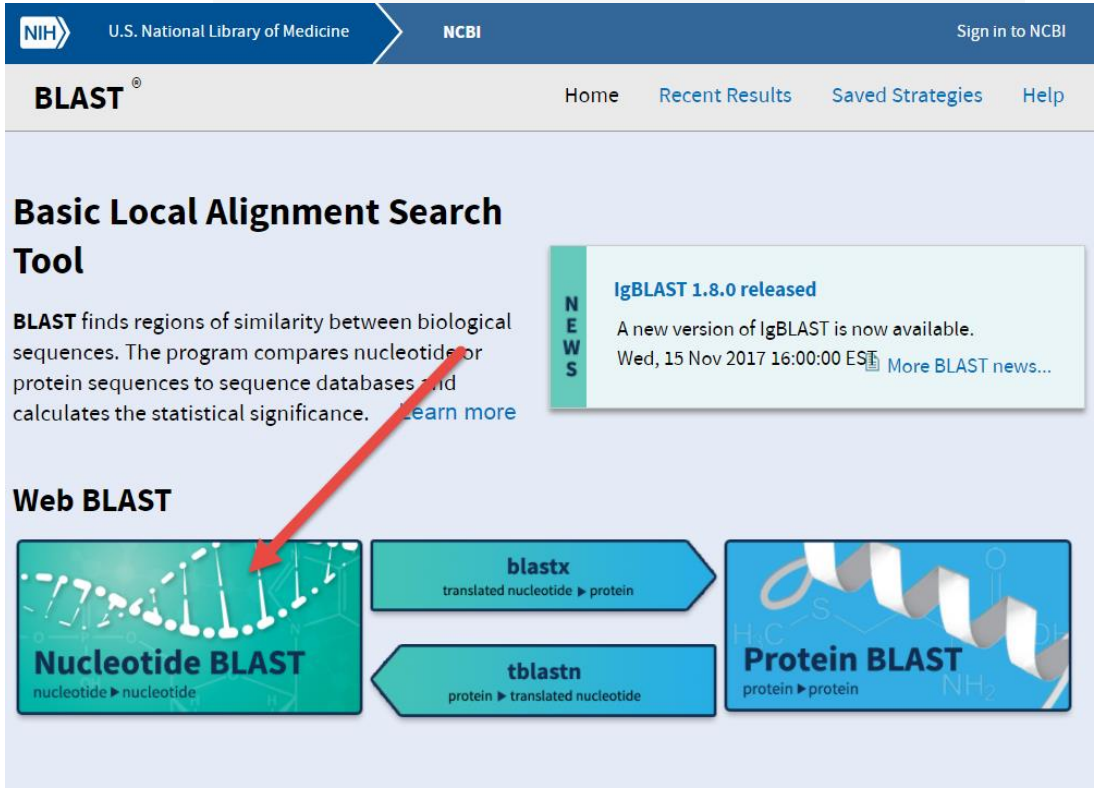


# Third - Developing sequence baseline via GenBank or Sanger sequencing

National Center for Biotechnology Information




or




Option 1 – Do your own sequencing

Option 2 - Query sequence repository by species, gene, gene region

# Michigan State/Michigan DNR/EGLE eDNA metabarcoding Advances: from aquatic vertebrates to plants to entire ecosystems

SCIENTIFIC REPORTS

OPEN

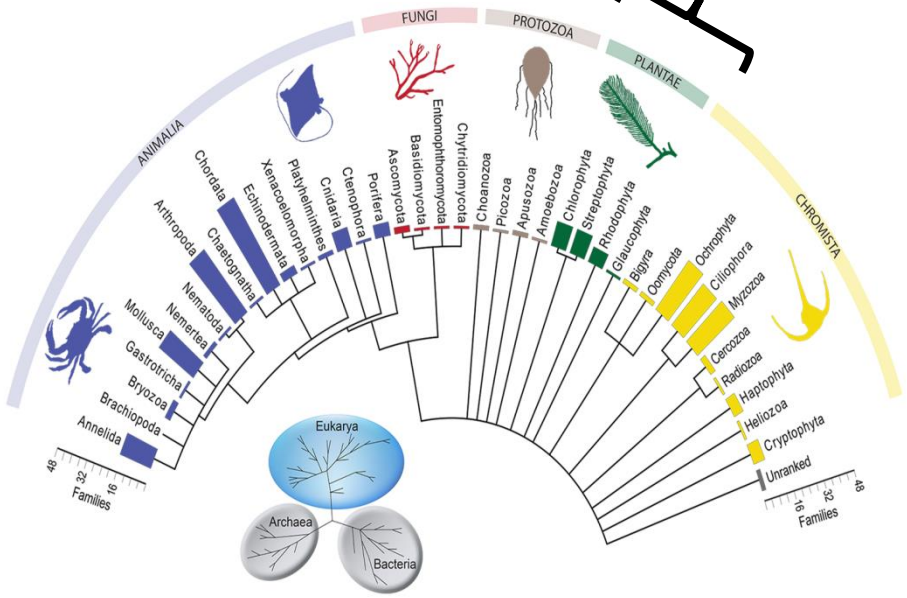
## Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment

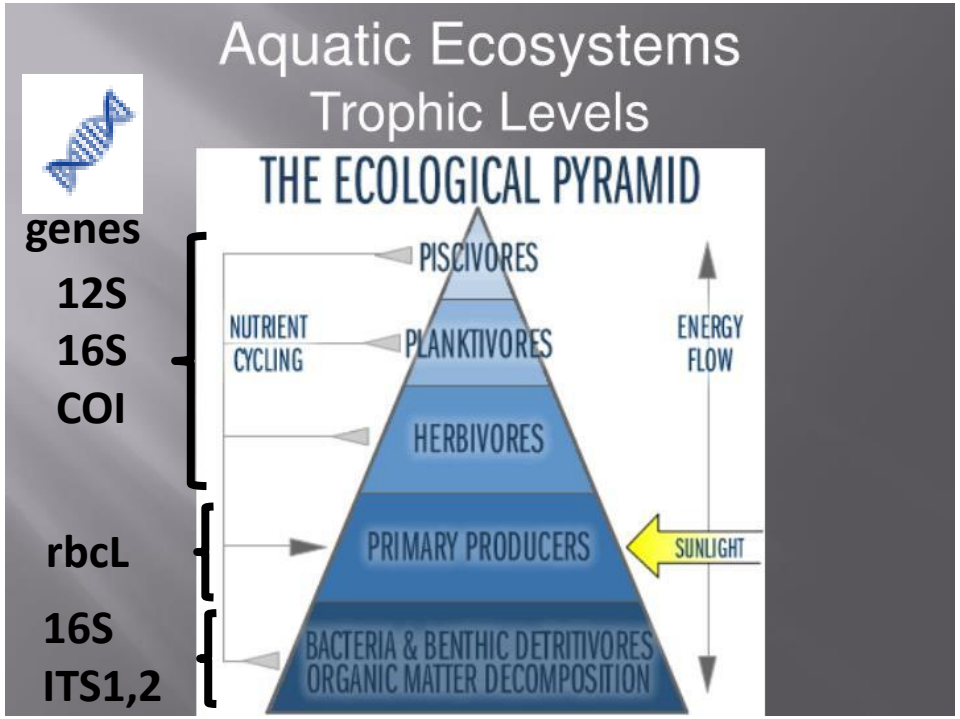
Received: 21 June 2017
Accepted: 11 September 2017
Published online: 25 September 2017

Michael Stat[1], Megan J. Huggett[2], Rachele Bernasconi[2], Joseph D. DiBattista[1], Tina E. Berry[1], Stephen J. Newman[1,3], Euan S. Harvey[1] & Michael Bunce[1]


12S, 16S, COI

rbcL



# Fourth - Taxonomic Database Development

12S & 16S sequences gathered from GenBank and produced in lab

- ➢136 and 140 native fish species
- ➢34 and 36 non-native ***(AIS)*** fish species
- ➢98 and 102 non-fish vertebrate species (e.g. mammals, birds, reptiles)

Primary focus on invasive fish species

- COI sequences gathered from GenBank and produced in lab
  - ➢ 1,865 arthropod
  - ➢ 200 mollusk and gastropod species
  - ➢ (including 27 and 18 ***AIS***, respectively)

- rbcL sequences gathered from GenBank and produced in lab
  - ➢ 2,211 plant species (including 80 AIS)

| **AIS Common name** | **AIS Sci.name** |
|---|---|
| Alewife | *Alosa pseudoharengus* |
| Northern snakehead | *Channa argus* |
| Grass carp | *Ctenopharyngodon idella* |
| Common carp | *Cyprinus carpio* |
| Ruffe | *Gymnocephalus cernua* |
| Silver carp | *Hypophthalmichthys molitrix* |
| Bighead carp | *Hypophthalmichthys nobilis* |
| Orange spotted sunfish | *Lepomis humilis* |
| Ide | *Leuciscus idus* |
| Pond loach | *Misgurnus anguillicaudatus* |
| Black carp | *Mylopharyngodon piceus* |
| Round goby | *Neogobius melanostomus* |
| Sea lamprey | *Petromyzon marinus* |
| Stone moroko | *Pseudorasbora parva* |
| Amur bitterling | *Rhodeus sericeus* |
| Zander | *Sander lucioperca* |
| Rudd | *Scardinius erythrophthalmus* |
| Wels catfish | *Silurus glanis* |
| Tench | *Tinca tinca* |

Pukk et al. in 2021; Diversity & Distributions

# Building your 'baseline' sequence data base via GENBANK or your own Sanger sequencing

## Crayfish example

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KJ6458271_Procambarus_clarkii | | | | | | | | | | | | | | | | | | | | | | | |
| KT0364441_Procambarus_clarkii | 1 | | | | | | | | | | | | | | | | | | | | | | |
| KJ6458241_Procambarus_clarkii | 2 | 1 | | | | | | | | | | | | | | | | | | | | | |
| KJ6458231_Procambarus_clarkii | 2 | 1 | 0 | | | conspecific basepair differences out of 220 (0-3 base pairs or 0.00% to 1.36% within species | | | | | | | | | | | | | | | | | |
| AF4360401_Procambarus_clarkii | 1 | 0 | 1 | 1 | | | | | | | | | | | | | | | | | | | |
| KJ6458221_Procambarus_clarkii | 2 | 1 | 0 | 0 | 1 | | | | | | | | | | | | | | | | | | |
| KJ6458341_Procambarus_clarkii | 2 | 1 | 2 | 2 | 1 | 2 | | | | | | | | | | | | | | | | | |
| KJ6458181_Procambarus_clarkii | 3 | 2 | 3 | 3 | 2 | 3 | 1 | | | | | | | | | | | | | | | | |
| KJ6458321_Procambarus_clarkii | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | | | | | | | | | | | | | | | |
| JX127827_Procambarus_acutus | 11 | 10 | 11 | 11 | 10 | 11 | 11 | 12 | 9 | | Congeneric basepair differences out of 220 (5-13 base pairs or 2.27% to 5.45%) | | | | | | | | | | | | |
| KF771116_Procambarus_acutus | 8 | 7 | 8 | 8 | 7 | 8 | 8 | 9 | 6 | 5 | | | | | | | | | | | | | |
| FJ619805_Procambarus_acutus | 8 | 7 | 8 | 8 | 7 | 8 | 8 | 9 | 6 | 5 | 0 | | | | | | | | | | | | |
| FJ619804_Procambarus_acutus | 8 | 7 | 8 | 8 | 7 | 8 | 8 | 9 | 6 | 5 | 0 | 0 | | | | | | | | | | | |
| EU433915_Procambarus_acutus | 12 | 11 | 12 | 12 | 11 | 12 | 12 | 13 | 10 | 9 | 8 | 8 | 8 | | | | | | | | | | |
| EF012354_Procambarus_acutus | 7 | 6 | 7 | 7 | 6 | 7 | 7 | 8 | 5 | 4 | 1 | 1 | 1 | 7 | | | | | | | | | |
| KC1634911_Fallicambarus_fodiens | 23 | 22 | 21 | 21 | 22 | 21 | 22 | 21 | 23 | 26 | 24 | 24 | 24 | 24 | 23 | | confamilial differences are considerably higher (Procambarus vs others) | | | | | | |
| KC1634701_Fallicambarus_fodiens | 25 | 26 | 25 | 25 | 26 | 25 | 26 | 27 | 27 | 24 | 24 | 24 | 24 | 24 | 23 | 14 | | | | | | | |

## Considerations

- Often times sequences in DNA data base repositories originate from locales far removed from your study area
- Often times sequences from the same gene regions do not overlap completely, reducing the length of homologous regions
- Classification should embrace intra-specific diversity (how similar is a sequence– no. base pair differences) to be treated as the same species vs higher taxonomic classification?

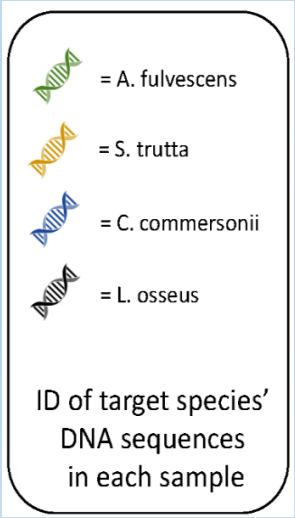
# Fifth - Bioinformatic Data Analysis

- Sequence data was analysed using computer pipeline – e.g., *Mothur*
- Reads were aligned to the taxonomic database
- Sequences were clustered into OTUs
- Prior to the analysis negative controls were used to account for contamination
- Unclassified OTUs subjected to BLAST



Community matrix

| Sample |  |  | ... |  |
|---|---|---|---|---|
| ID1 | 11100 | 690 | ... | 1232 |
| ID2 | 3334 | 45 | ... | 0 |
| ... | ... | ... | .. | ... |
| ID413 | 2 | 3076 | .. | 5 |




Analyses done at **HPCC** – high performance computing cluster

# Community Matrix of metabarcoding eDNA (16S rDNA) for crayfish in Detroit metro-area wetlands

| Location | Sample Type | Procambarus clarkii | Procambarus acutus | Procambarus unclassified | Cambarus diogenes | Cambarus robustus | Cambarus unclassified | Orconectes virilis | Orconectes immunis | Orconectes propinquus | Orconectes rusticus | Orconectes unclassified |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sheraton | Control | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 2 | 0 |
| | Surface | 8433 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 1 |
| | | 15445 | 0 | 2 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 2 |
| | | 17566 | 0 | 2 | 0 | 0 | 0 | 10 | 1 | 1 | 0 | 0 |
| Rhodson | Control | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 1 | 0 | 0 |
| | Surface | 1 | 0 | 0 | 0 | 0 | 0 | 7146 | 1 | 0 | 0 | 96 |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 16744 | 3 | 0 | 0 | 220 |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 19023 | 0 | 0 | 1 | 267 |
| Lujon | Control | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 1 | 1 |
| | Surface | 0 | 0 | 0 | 0 | 0 | 0 | 10182 | 2 | 1656 | 2 | 175 |
| | | 391 | 0 | 0 | 0 | 2 | 0 | 17957 | 2 | 0 | 1 | 181 |
| | | 1 | 0 | 0 | 0 | 1005 | 0 | 12012 | 0 | 1680 | 4 | 179 |
| Fitzgerald | Control | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 0 |
| | Surface | 0 | 0 | 0 | 0 | 1455 | 0 | 7 | 0 | 0 | 0 | 1 |
| | | 1 | 0 | 0 | 0 | 6521 | 1 | 13 | 0 | 0 | 0 | 2 |
| | | 0 | 0 | 0 | 0 | 5039 | 3 | 9 | 0 | 0 | 6 | 0 |
| Fairway | Control | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 |
| | Surface | 0 | 0 | 0 | 0 | 0 | 0 | 9071 | 0 | 0 | 1 | 89 |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 11418 | 1 | 0 | 0 | 111 |
| | | 1 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 5 | 0 | 1 |
| Fox Creek 1 | Control | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 3 | 1 | 0 |
| | Surface | 0 | 0 | 0 | 0 | 0 | 0 | 7100 | 1 | 0 | 0 | 100 |
| | | 2 | 0 | 0 | 0 | 0 | 0 | 17264 | 0 | 0 | 1 | 241 |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 3 | 2 |

# Database developments for plants

- = A. fulvescens
- = S. trutta
- = C. commersonii
- = L. osseus

- Reviewed literature and availability of sequences
  - ITS2, MatK, rbcL, tRNA-Leu p6 loop region
- Michigan's AIS Watchlist, and the Michigan EGLE p51 form
- queried NCBI for sequences of interest

Our experiences

- We reviewed the plant barcoding literature and found that most of the single-species assays developed thus far amplified loci that are too large (i.e., > 600 bp) for sequencing on Illumina's MiSeq platform (Hollingsworth et al. 2011) from eDNA.
- Interrogation of aquatic plant communities was conducted based on eDNA samples collected from fish surveys
- For tRNA-Leu and rbcL where metabarcoding data was collected from inland lakes, many aquatic plant AIS species had native congeners that could not be distinguished to species
- Based on the timing and location of sampling, in many samples and inland lakes that majority of sequences were from terrestrial and wetlands plant taxa

# Part II – Applications – Michigan inland lake communities

# MSU eDNA Metabarcoding Projects

## Objectives

- Detect aquatic invasive species
- Compare with traditional gear surveys
- What lake characteristics predict diversity? Non-native prevalence?

## 22 inland lakes sampled (2016-2018)

- Sard et al. (2019) – 8 lakes (fish)
- Pukk et al. (2021) – 22 lakes (fish)
- Costello et al. (in prep) – 22 lake (plants)*


ORIGINAL ARTICLE

Environmental DNA

WILEY

**Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears**

Nicholas M. Sard[1,2] | Seth J. Herbst[3] | Lucas Nathan[3] | Genelle Uhrig[4] | Jeannette Kanefsky[2] | John D. Robinson[2] | Kim T. Scribner[2,5]

RESEARCH ARTICLE

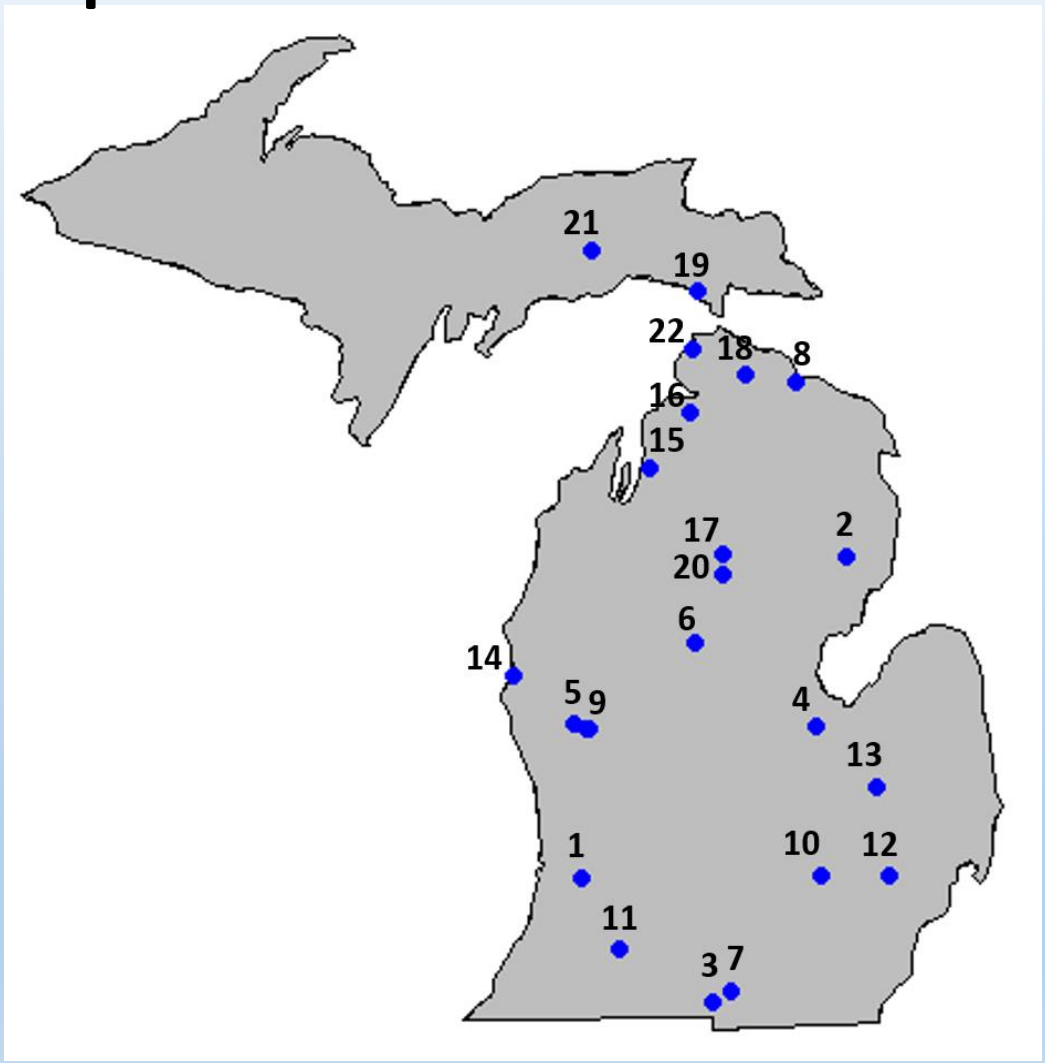
Diversity and Distributions

WILEY

**eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity**

Lilian Pukk[1] | Jeannette Kanefsky[1] | Amanda L. Heathman[1] | Ellen M. Weise[1] | Lucas R. Nathan[2] | Seth J. Herbst[2] | Nicholas M. Sard[3] | Kim T. Scribner[1,4] | John D. Robinson[1]

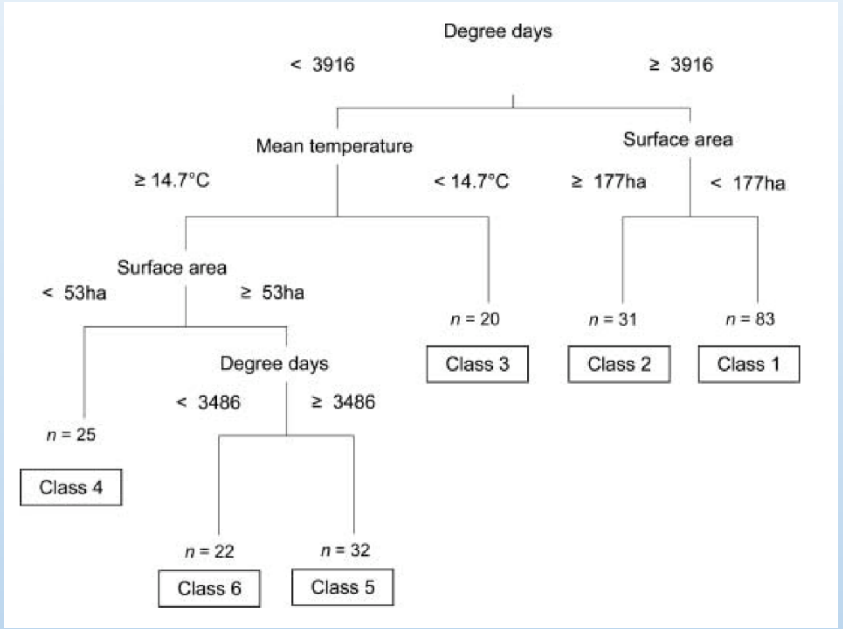
# Sampled Lakes




ARTICLE

**A Landscape-Based Classification of Fish Assemblages in Sampled and Unsampled Lakes**

**Kevin E. Wehrly,* James E. Breck, Lizhu Wang, and Lidia Szabo-Kraft**
*Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 212 Museums Annex, 1109 North University Avenue, Ann Arbor, Michigan 48109, USA*




## Sampled Lake Classes

| Lake Class | Number Total | Percent |
|---|---|---|
| 1 | 10 | 45 |
| 2 | 6 | 27 |
| 3 | 2 | 9 |
| 4 | 0 | 0 |
| 5 | 2 | 9 |
| 6 | 2 | 9 |

# Lake Selection

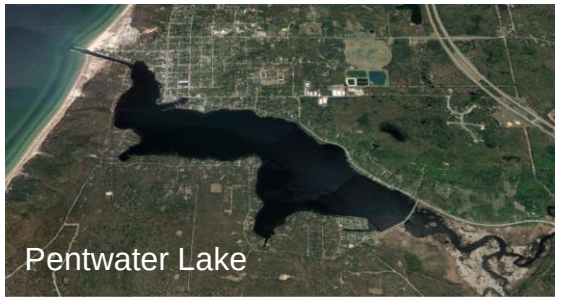
Represent environmental variation

- Lake area - 14 to 8000 hectares
- Max depth - 2 to 87 m
- Development - 1.5% to 90%+
- Connectivity - isolated to highly connected

Include…

- Lakes popular with anglers (Houghton / Higgins)
- Highly developed areas (Cass / Livingston Co.)
- Deep, cold lakes (Torch Lake)

Prioritized lakes with fish survey data available





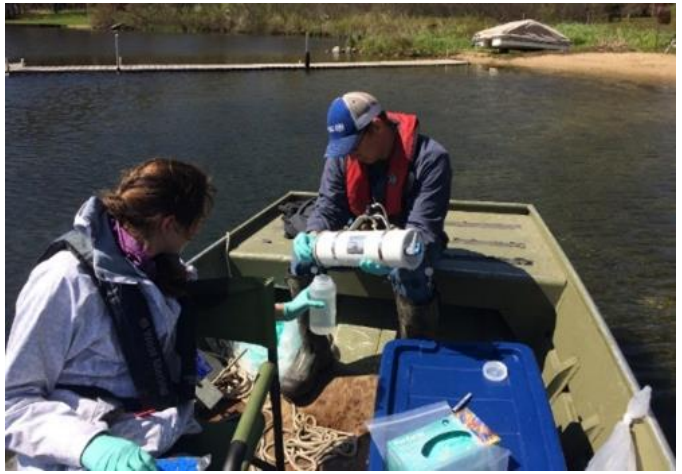

# eDNA Sampling Methods

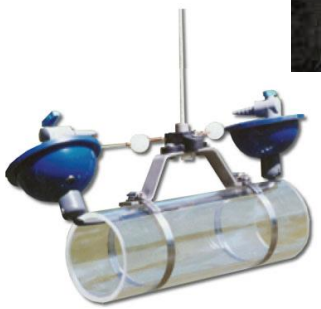
Paired samples with Status and Trends (13 lakes)

- 30 to 57 samples per lake, over 950 total samples
- Collections include surface and benthic water samples
- Negative controls processed in the field and lab

1L filtered on site with Smith-Root ANDe backpack

- Single-use filter housings limit potential contamination
- DNA extraction → PCR (2 markers) → sequencing





# Objective 1 – evaluate eDNA relative to traditional gear

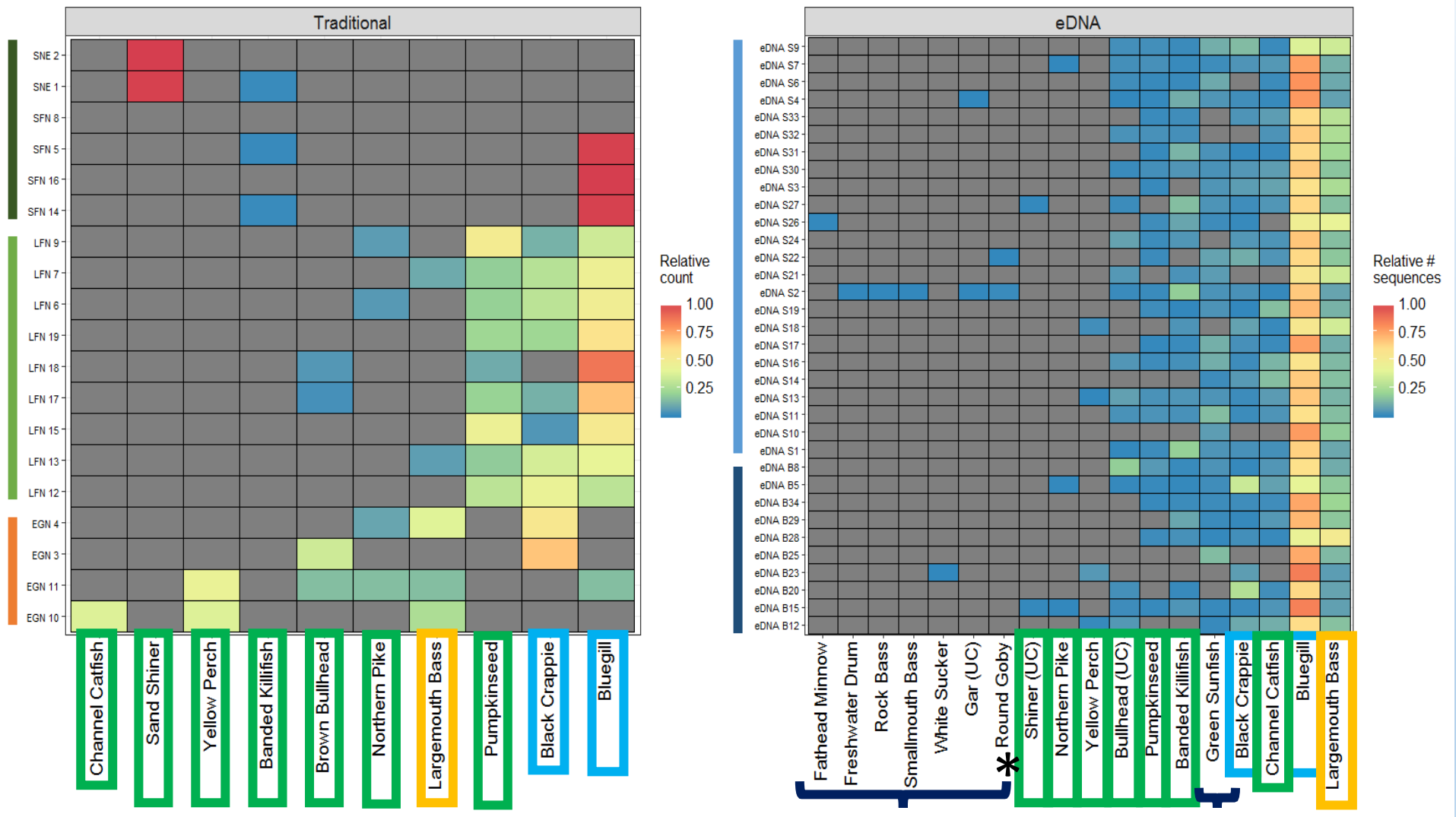





Methods

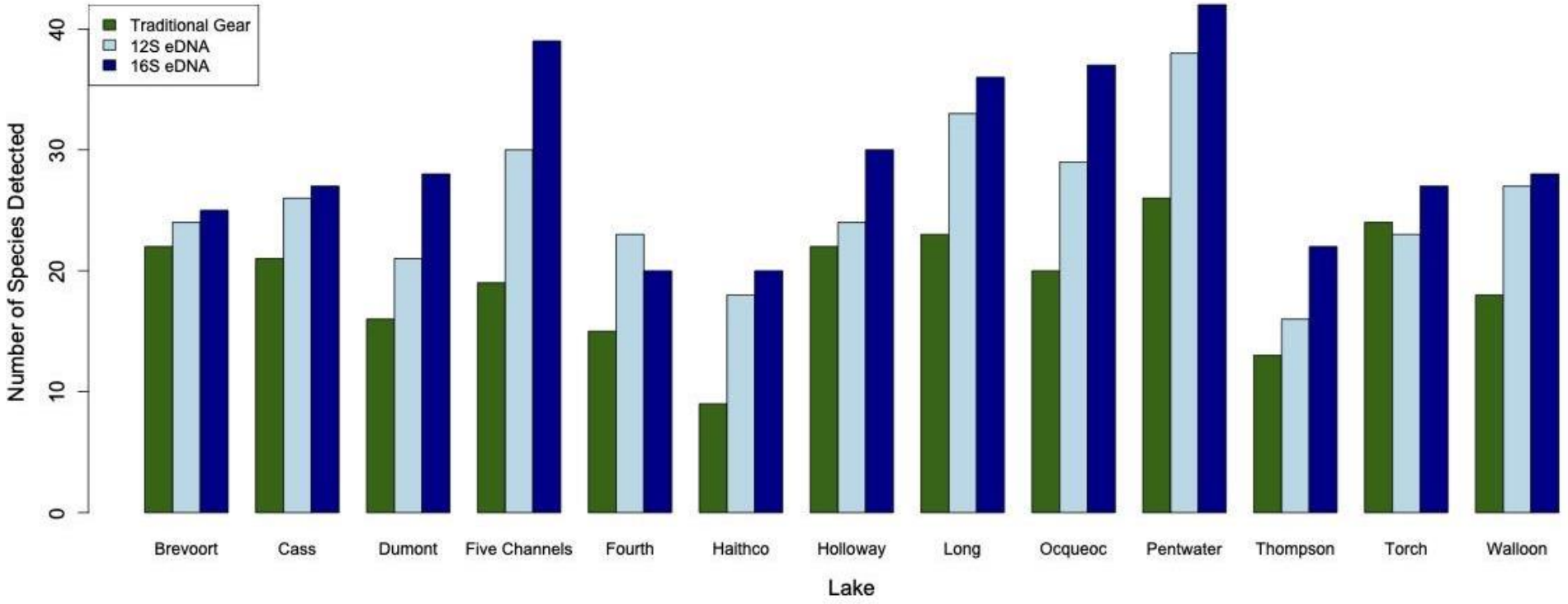
1. With traditional gear – paired and supplemented
2. Without traditional gear - random

# Results – example for 1 lake of raw eDNA and traditional community matrices



# Fish Species Detected

## eDNA vs. Traditional Gear




eDNA

vs

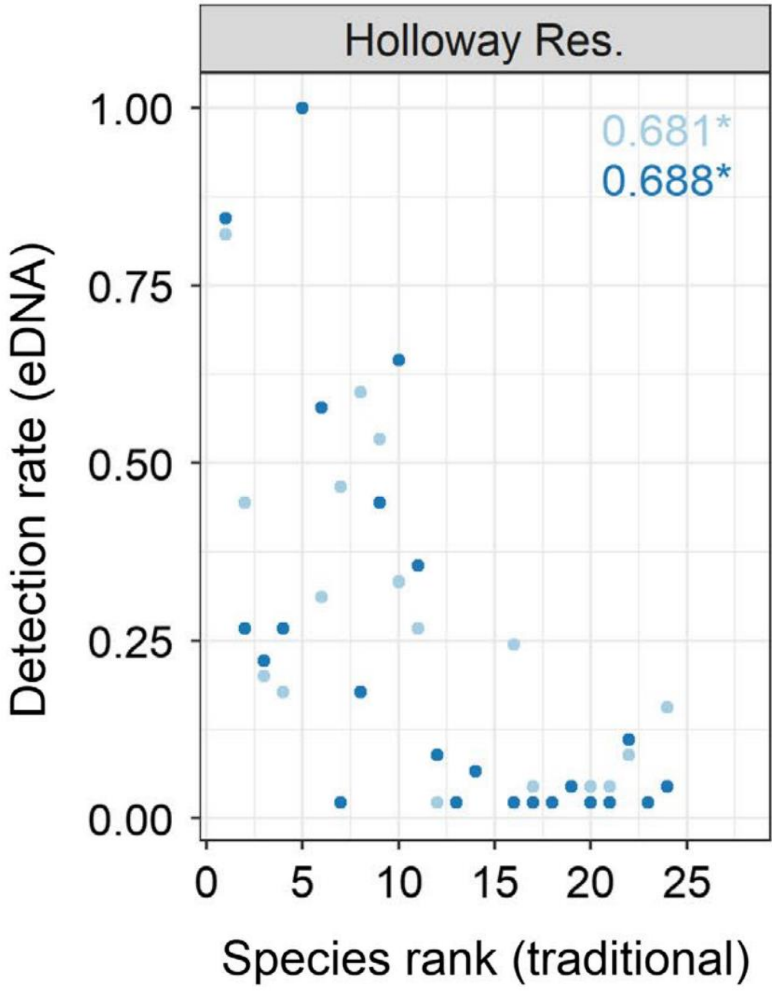
Traditional

*Both* eDNA markers detect more species than traditional gear (12 of 13 lakes)

# Information on Abundance / Biomass
## eDNA vs. Traditional Gear

Comparing estimates of relative abundance from eDNA to those from Status and Trends

- Significant correlations b/w rank abundance (traditional gear) and eDNA detection rate




Sard et al. (2019)

# eDNA vs Traditional gear

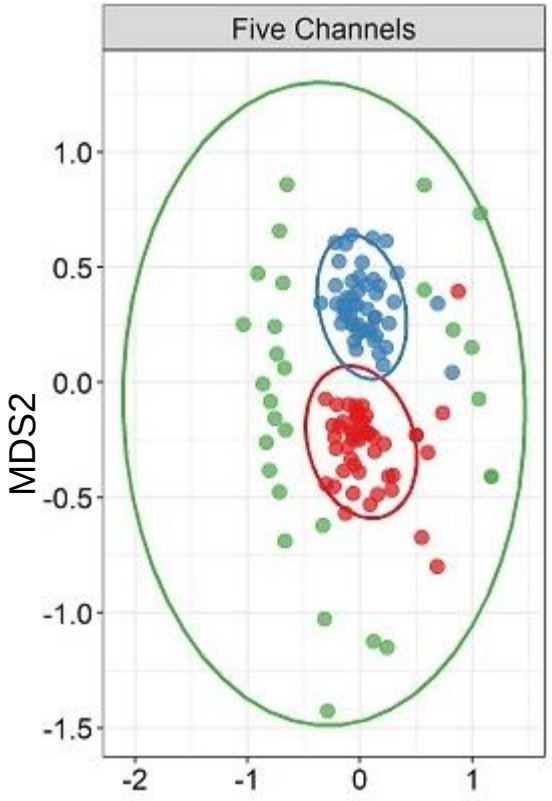
eDNA

vs

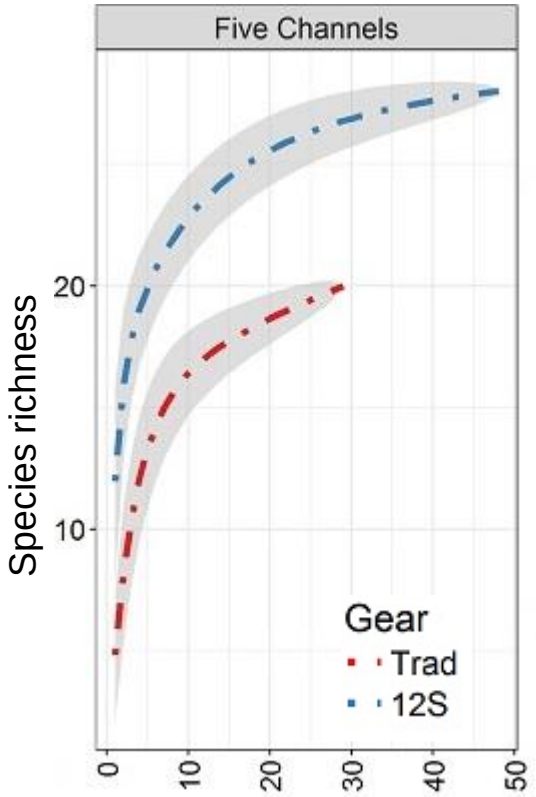
Traditional

- eDNA samples have less inter-sample variation in fish species
- Different gear types are known to be selective for specific fish species

Sard et al. (2020) Environmental DNA

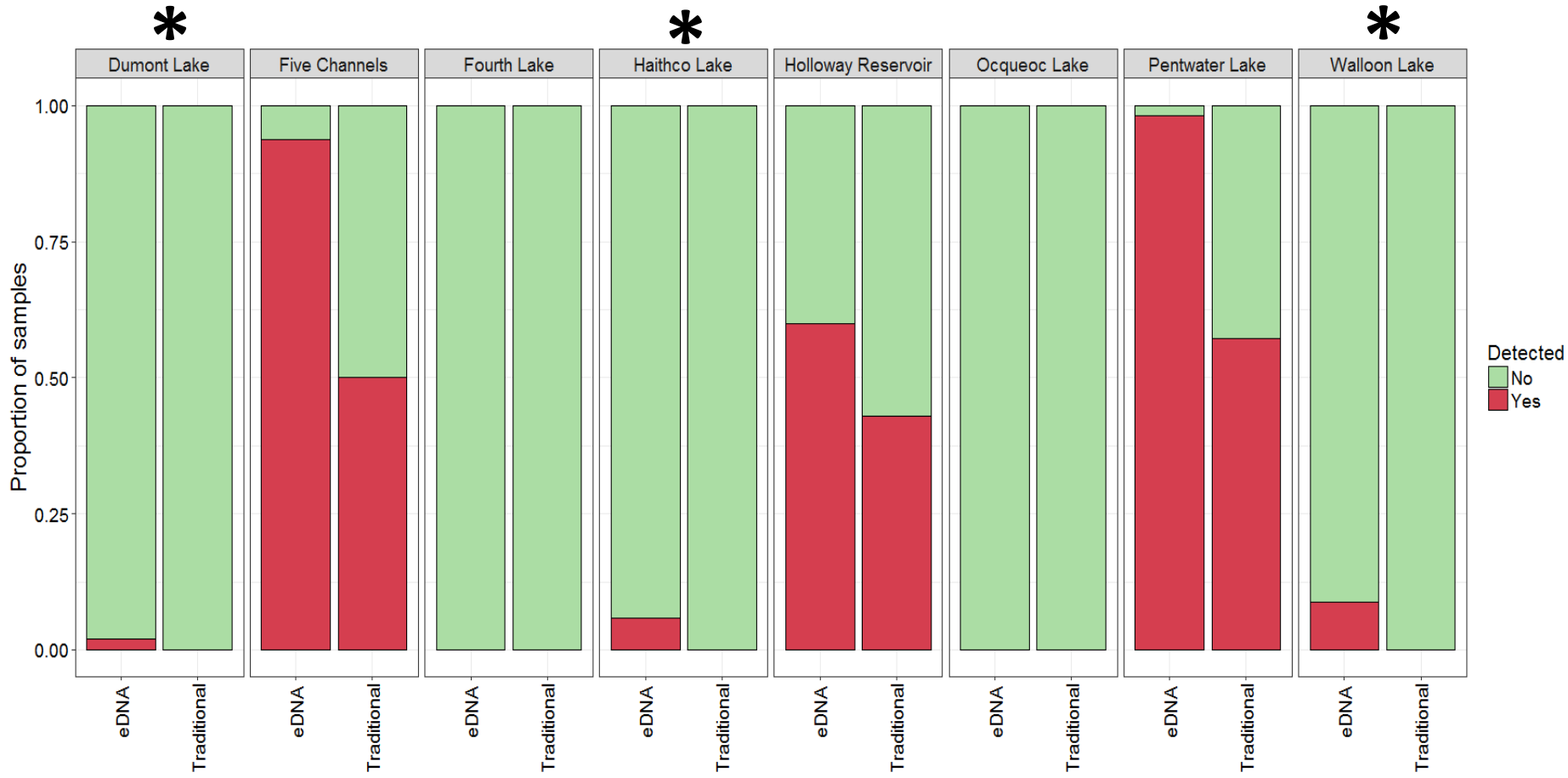



Non-metric multidimensional scaling plot, comparing **Bray-Curtis distance** matrices among 12S (red), 16S (blue) and traditional (green) sampling approaches.

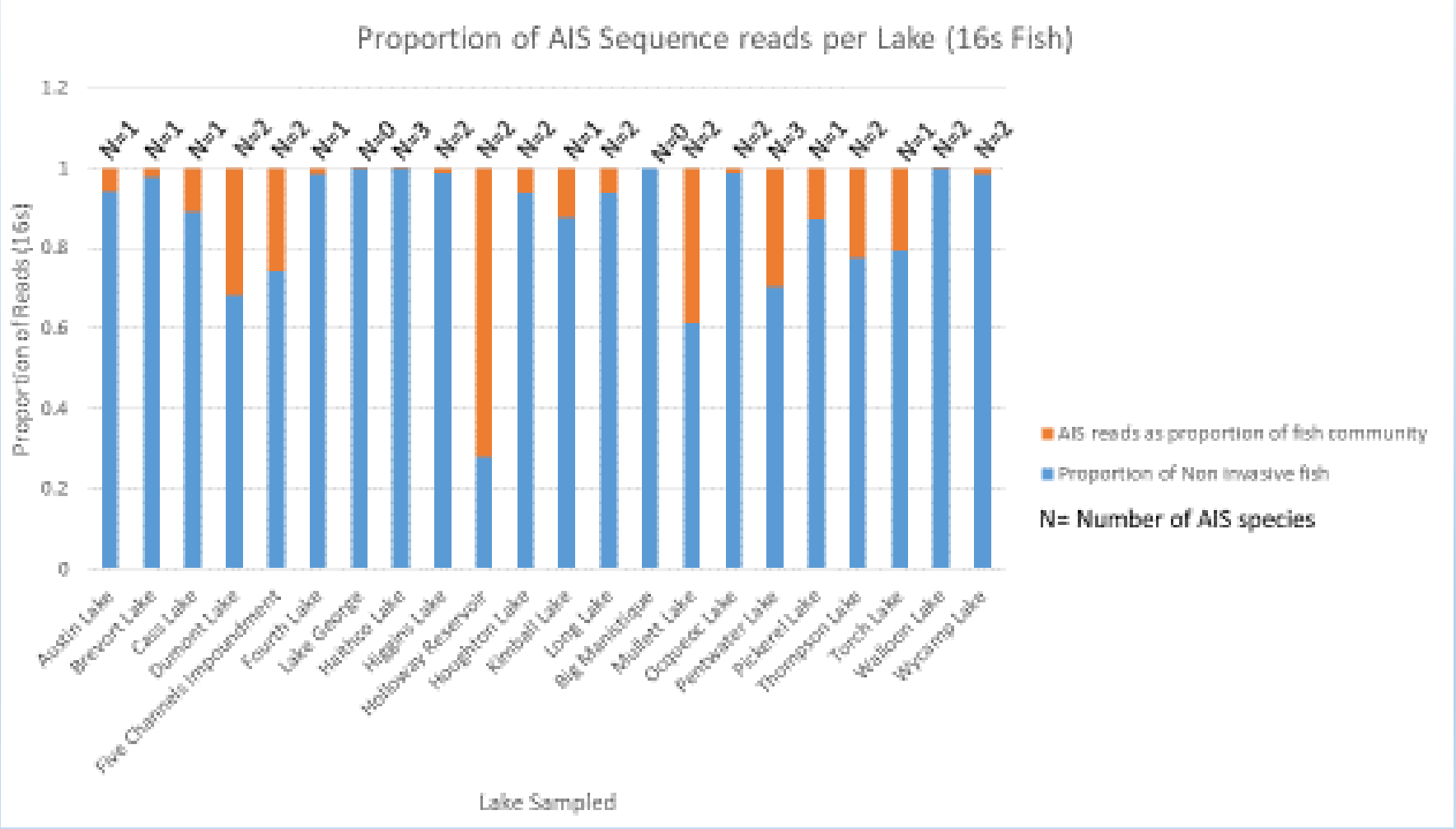



Species accumulation curve for traditional and eDNA sampling approaches

# The eDNA approach detected round goby in more samples than traditional gear (eDNA 6 of 8 lakes vs traditional 3 of 8 lakes)

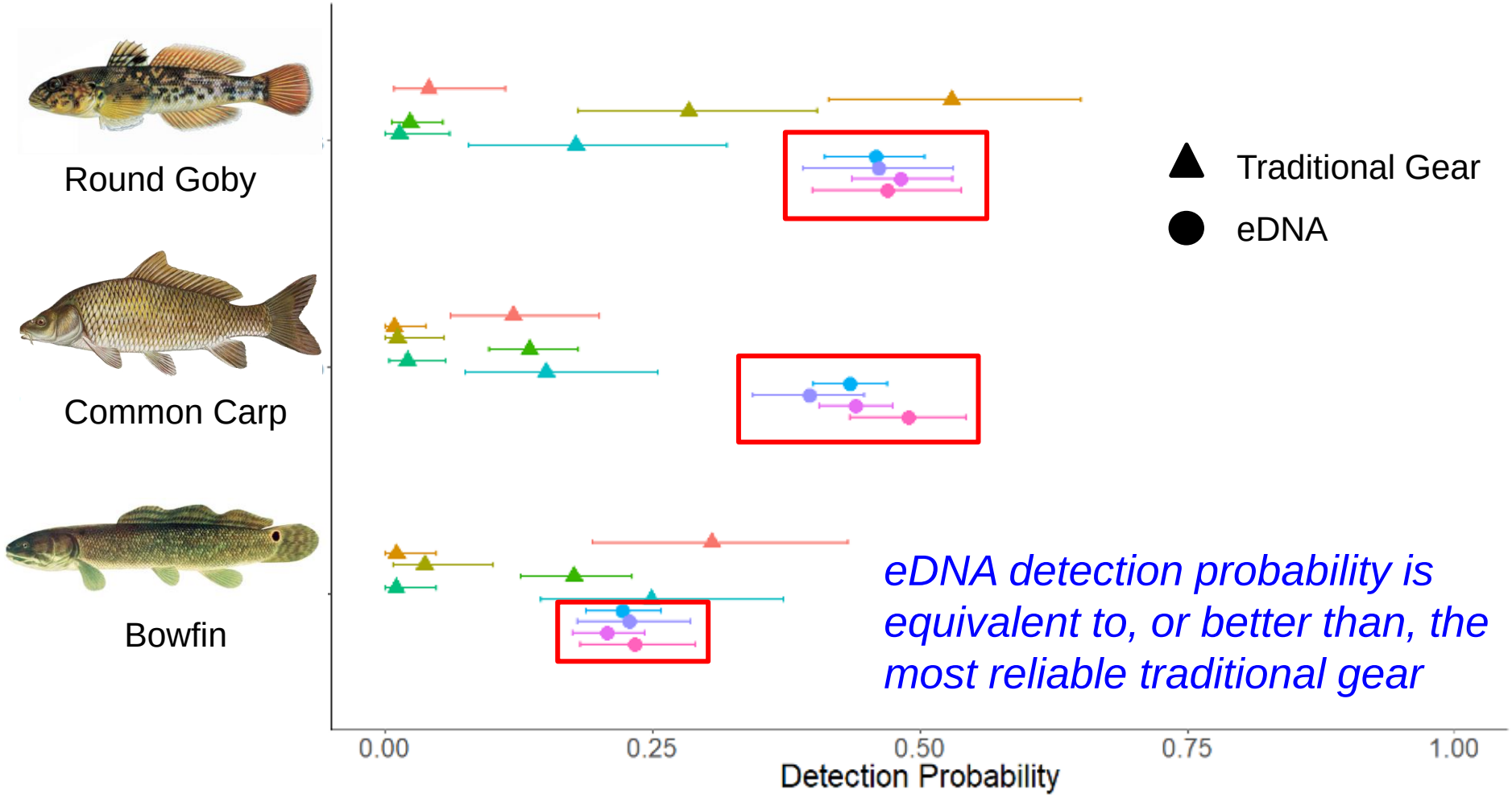


Proportion of aquatic invasive fish species reads and the counts of AIS per lake (16S) for 22 Michigan inland lakes.



# Objective 2 – Estimation of Detection Probability

## Occupancy Modeling - eDNA vs. Traditional Gear




*eDNA detection probability is equivalent to, or better than, the most reliable traditional gear*

Pukk et al. (2021)

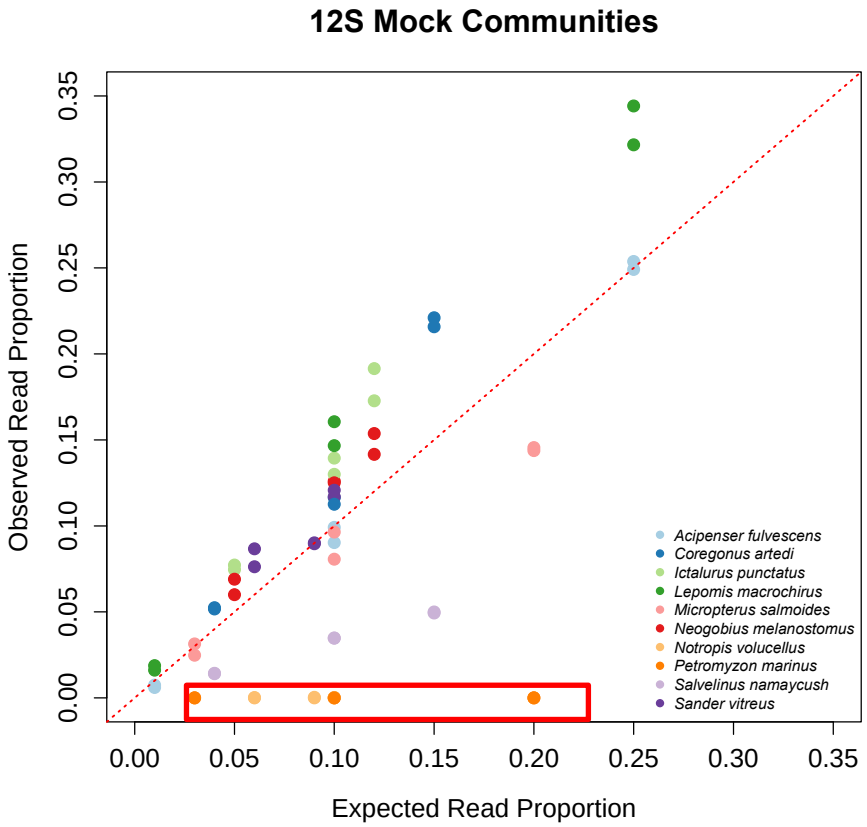
# Information on Abundance / Biomass
## Mock Community Samples

“Mock communities”

- Mixtures of DNA from 10 species
- Three mixtures with different relative concentrations

Results

- DNA concentration tightly correlated with read counts (r = 0.68)
- Some evidence of amplification bias for particular taxa (sea lamprey)



# Objective 3 - Influences of Lake Characteristics

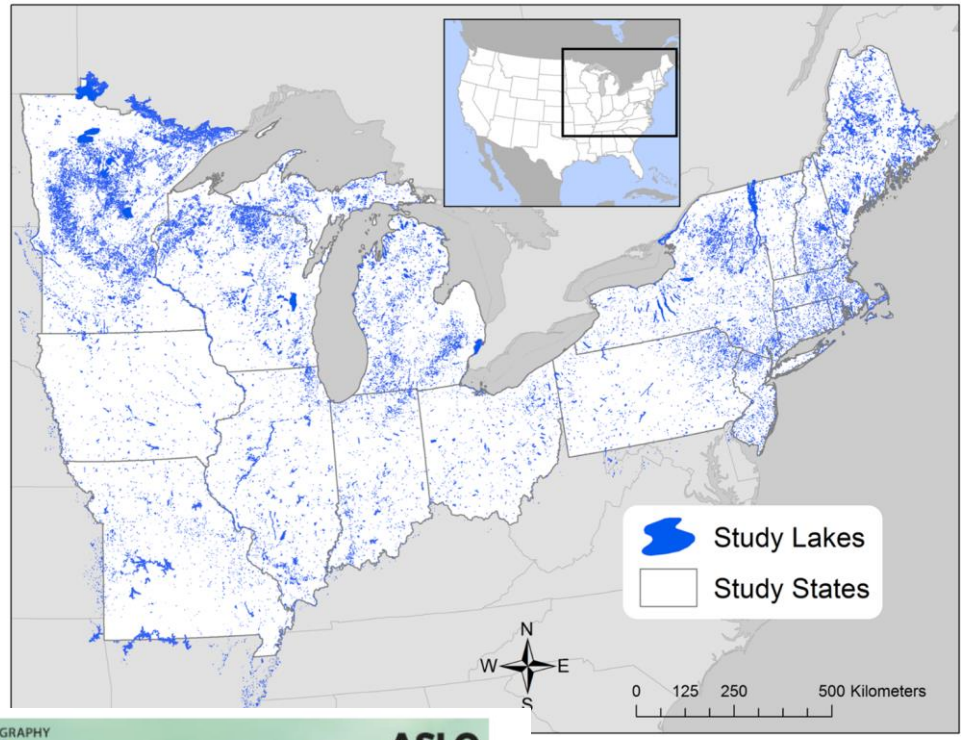
What lake attributes contribute to variation in…

- Diversity of the fish community?
- Prevalence of non-native species?

Water quality / ecological data from LAGOS database

- 51,000 total lakes (> 4 ha)
- 15,000+ lakes in Michigan
- Climate, land use, area, depth, connectivity, water quality…

LAGOS – lake multi-scaled geospatial and temporal database




www.lagoslakes.org
bigdatalimno.org

**LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes**

Patricia A Soranno, Linda C Bacon, Michael Beauchene, Karen E Bednar, Edward G Bissell, Claire K Boudreau, Marvin G Boyer, Mary T Bremigan, Stephen R Carpenter, Jamie W Carr ... Show more

*GigaScience*, Volume 6, Issue 12, December 2017, gix101,

LIMNOLOGY AND OCEANOGRAPHY LETTERS ASLO Open Access

Data Article | Open Access

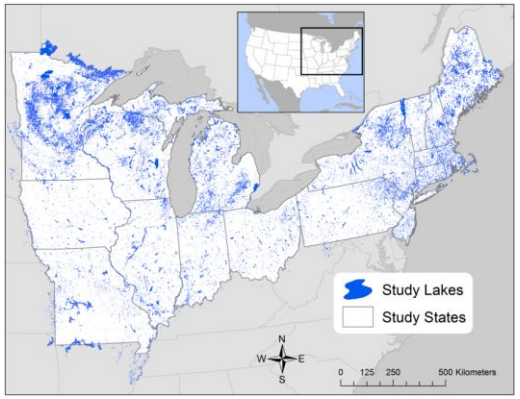
**LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S.**

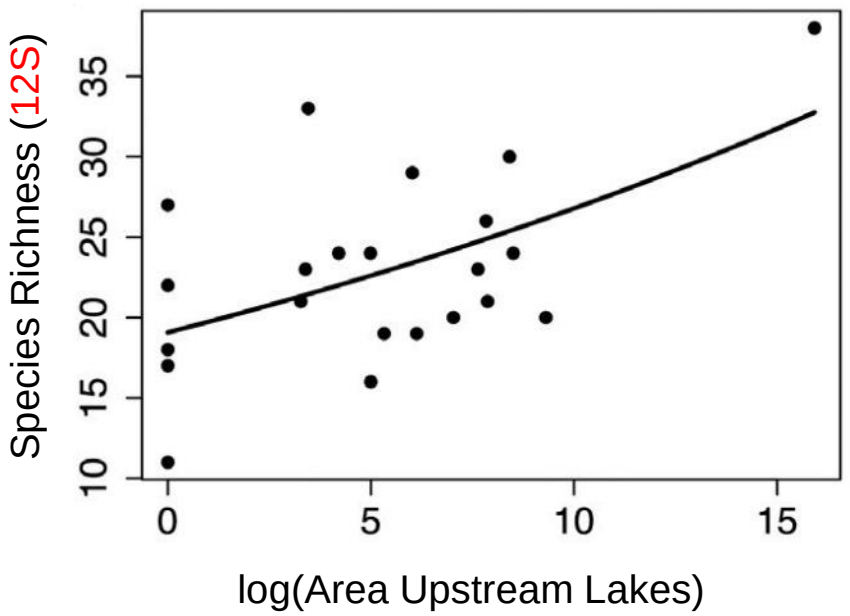
Kendra Spence Cheruvelil, Patricia A. Soranno, Ian M. McCullough, Katherine E. Webster, Lauren K. Rodriguez, Nicole J. Smith

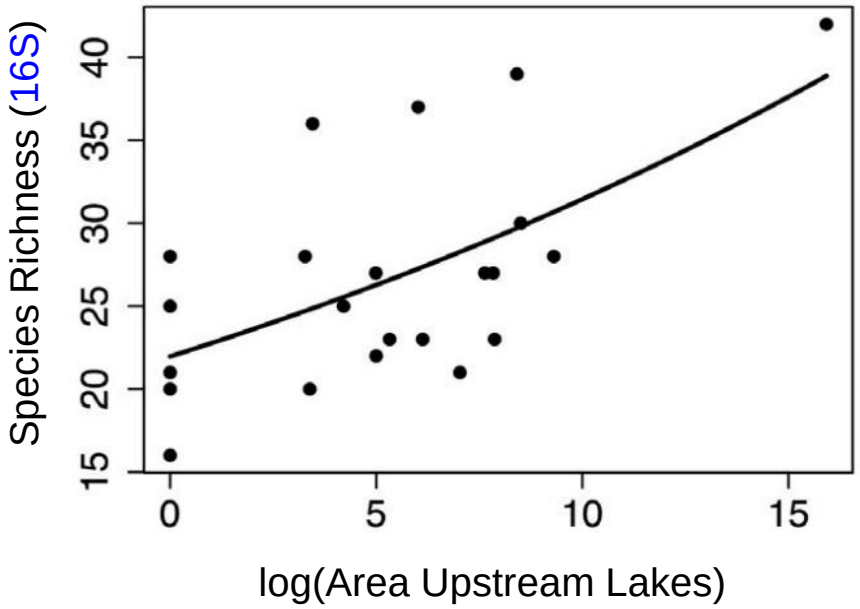
# Species Richness

Area of upstream lakes (connectivity) was positively related to fish species richness

- Both eDNA markers show this relationship
- Connectivity, rather than lake size, drives diversity



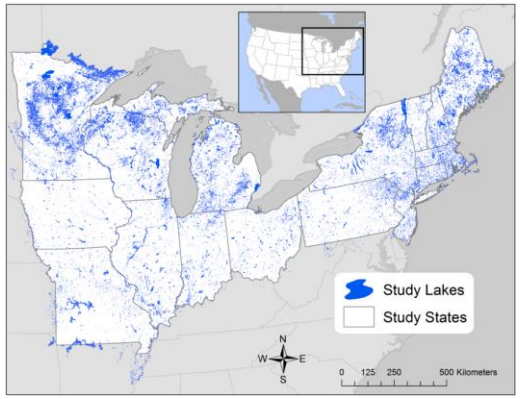



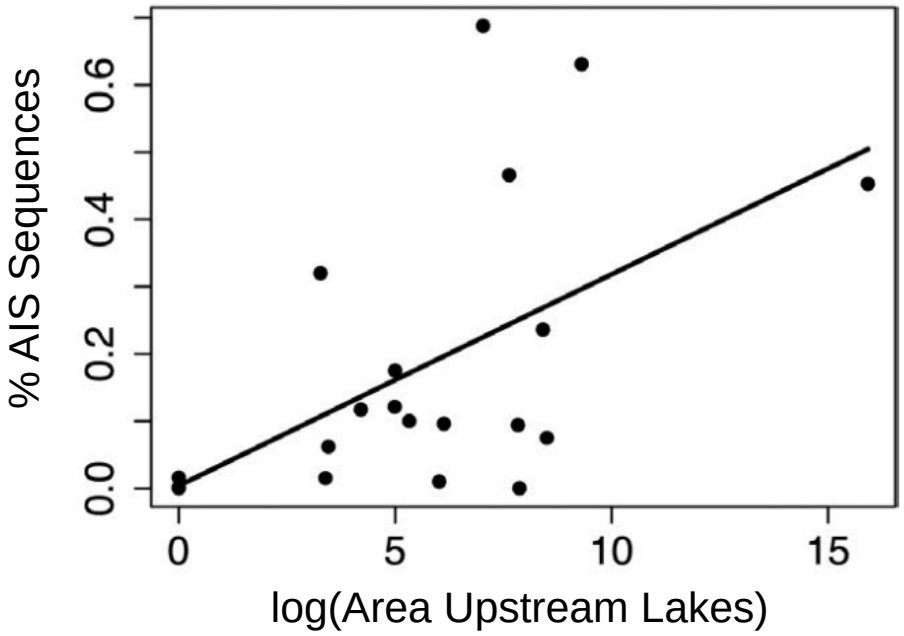



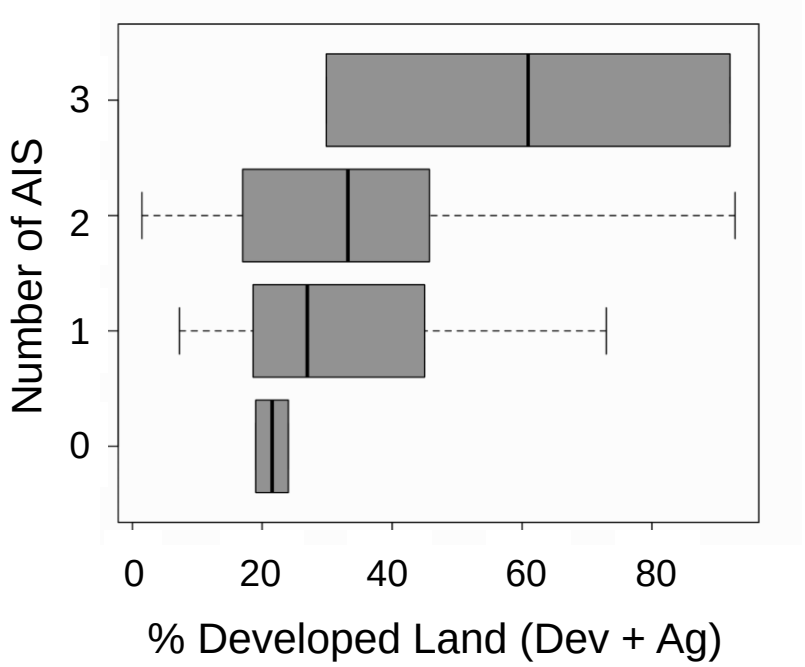
# AIS Prevalence

Evidence for positive relationships between…

- AIS prevalence and area of upstream lakes
- Number of AIS fish and % developed or agricultural land (disturbance / development)



# Objective 4 - Spatial Information in the Data

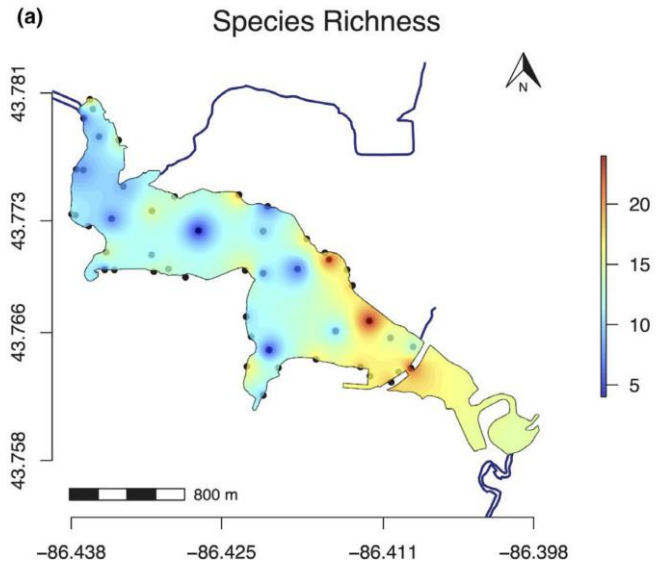
Visualize patterns of species richness and sequence abundance

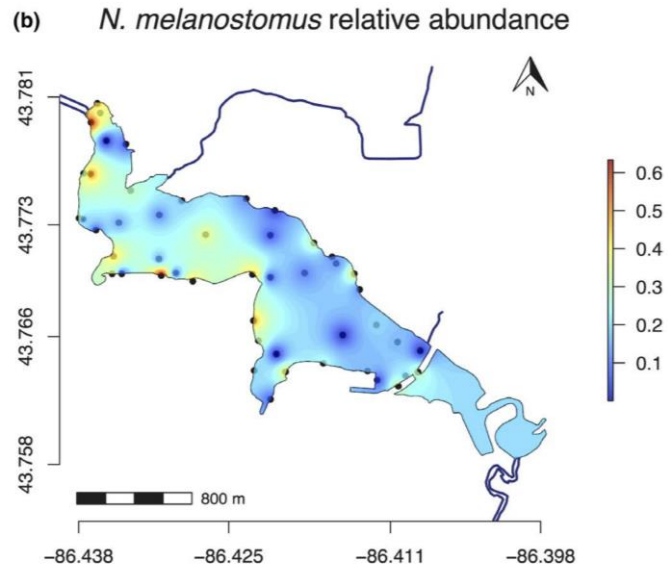
Identify associations between physical lake features and…

- Species richness?
- AIS sequence abundance?

Prioritizing control efforts

- Targeted removal of invasive species
- Specific areas for follow-up sampling?




RESEARCH ARTICLE

Diversity and Distributions WILEY

**eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity**

Lilian Pukk[1] | Jeannette Kanefsky[1] | Amanda L. Heathman[1] | Ellen M. Weise[1] | Lucas R. Nathan[2] | Seth J. Herbst[2] | Nicholas M. Sard[3] | Kim T. Scribner[1,4] | John D. Robinson[1]

# Selecting a Metabarcoding Marker for Plants

Sard et al. (unpubl) MSU – tRNA-Leu

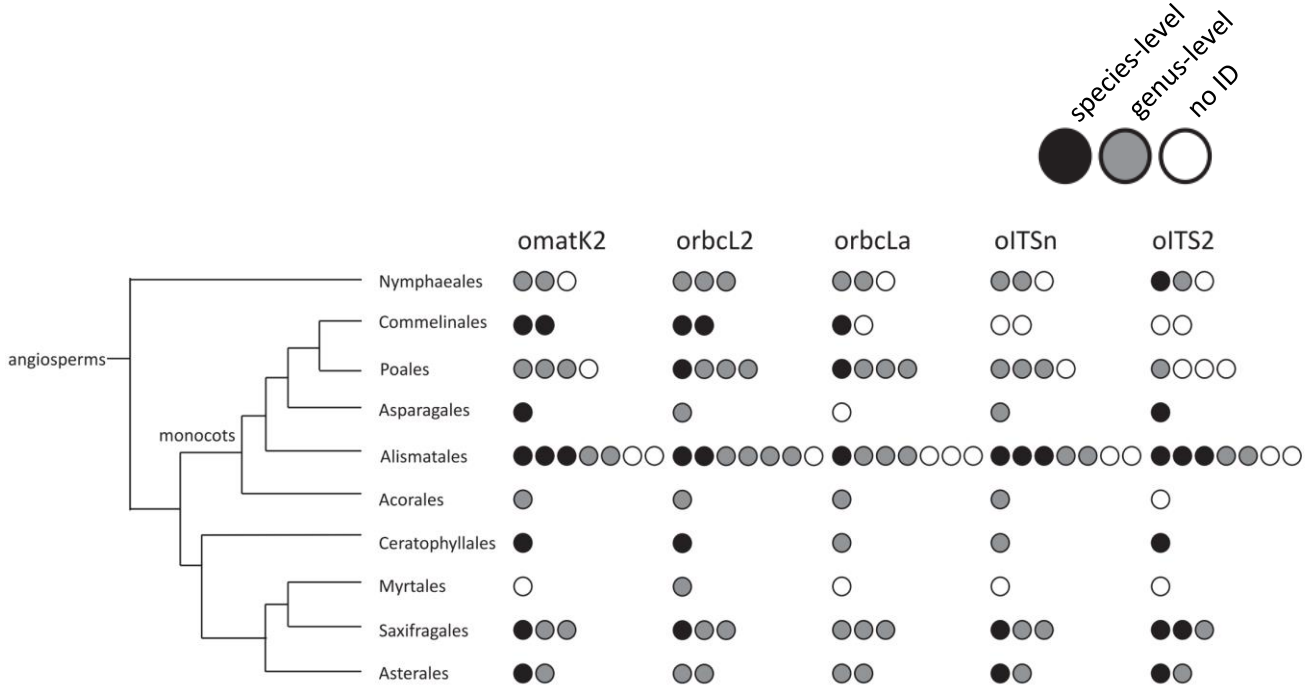
Fahner et al. (2016) – *ITS2* or *rbcL*

Coghlan et al. (2019) assessed both markers

- Improved species-level detection with *rbcL*

Taxonomic database

- 3174 unique sequences from 2212 species
- Habitat and native / non-native classifications




Coghlan et al. (2019)

PLOS ONE

Large-Scale Monitoring of Plants through Environmental DNA Metabarcoding of Soil: Recovery, Resolution, and Annotation of Four DNA Markers

Nicole A. Fahner[1], Shadi Shokralla[1], Donald J. Baird[2], Mehrdad Hajibabaei[1]*

Environmental DNA

Open Access

Dedicated to the study and use of environmental DNA for basic and applied sciences

**Development of an environmental DNA metabarcoding assay for aquatic vascular plant communities**

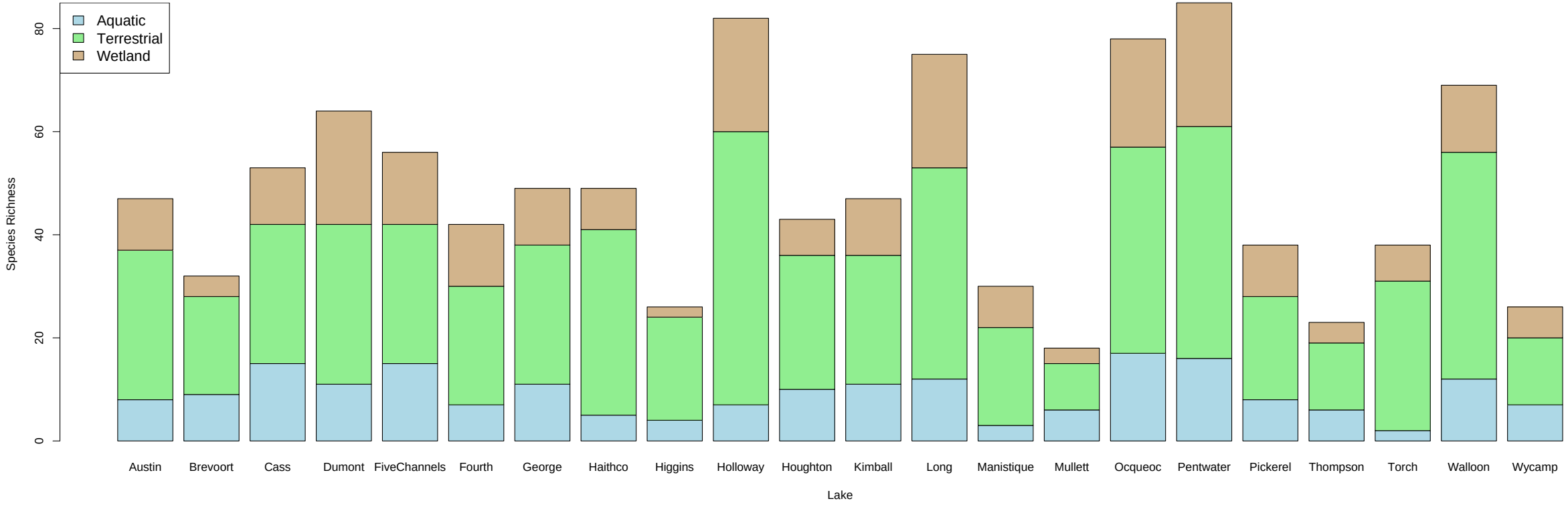
Stephanie A. Coghlan[1] | Aaron B. A. Shafer[1,2] | Joanna R. Freeland[1,3]

# Species Richness by Group

## Species richness varied across lakes

- 18 species detected in Mullett Lake
- 85 species detected in Pentwater Lake

## 48-77% of detected species were from terrestrial habitats

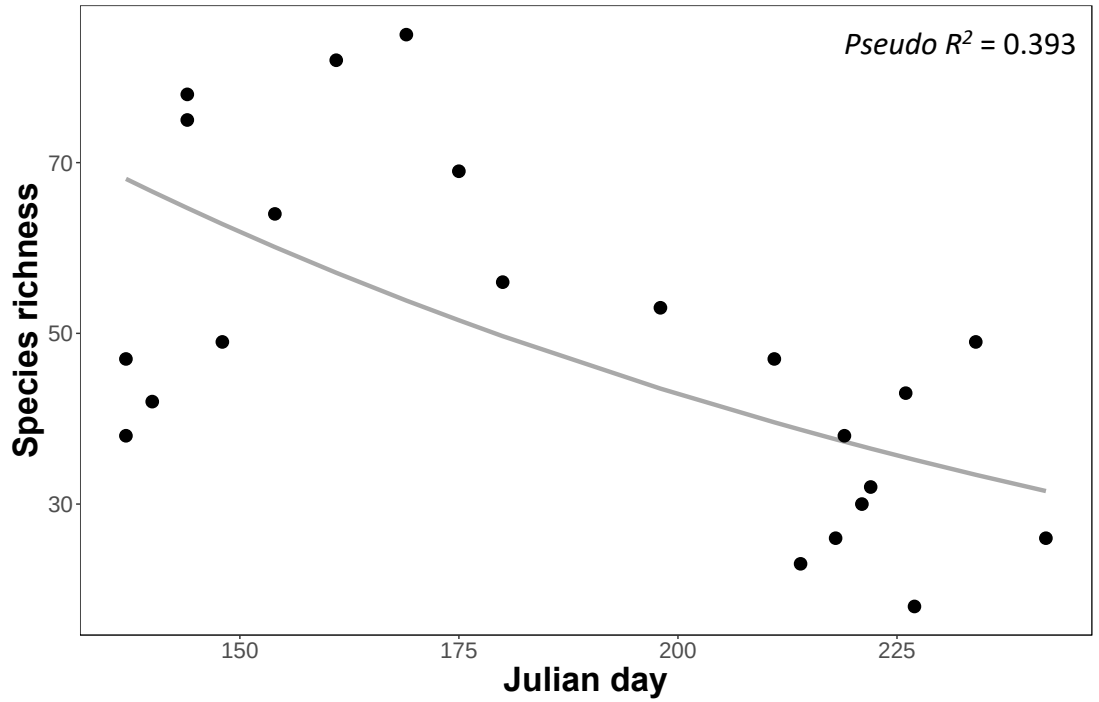


# GLMs: Species Richness of the Plant Community

Plant species richness best predicted by Julian Day (sampling)

Fewer plant species detected in samples collected *later* in the summer

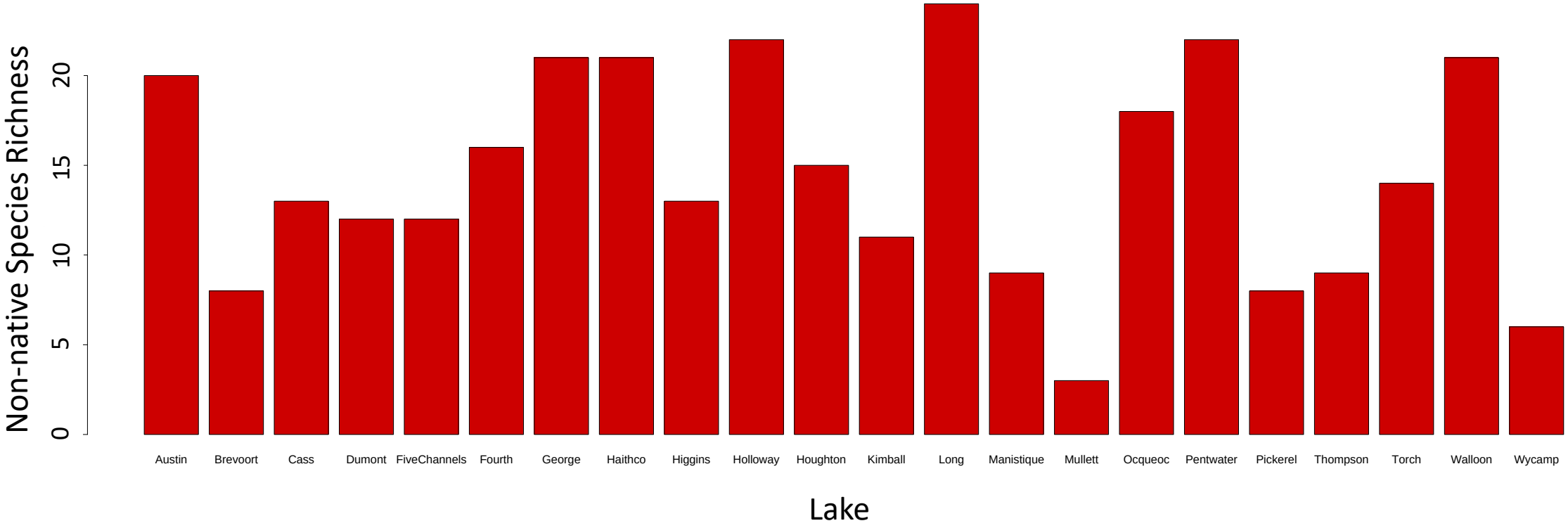
| Model | df | logLik | AICc | ΔAIC | ωi |
|---|---|---|---|---|---|
| Julian day | 3 | -90.184 | 187.7 | 0.00 | 0.940 |
| log.lake.area | 3 | -94.002 | 195.3 | 7.64 | 0.021 |
| Null | 2 | -95.682 | 196.0 | 8.30 | 0.015 |
| IWS.stream.density | 3 | -94.809 | 197.0 | 9.25 | 0.009 |
| log.area.upstream.lakes | 3 | -95.194 | 197.7 | 10.02 | 0.006 |
| max.depth | 3 | -95.484 | 198.3 | 10.60 | 0.005 |
| perc.agric.dev.land | 3 | -95.682 | 198.7 | 11.00 | 0.004 |
| Full | 8 | -89.274 | 205.6 | 17.92 | 0.000 |



# Non-native Plant Species Richness

Across habitats (aquatic, terrestrial, wetland), 3 to 24 non-native species detected per lake

- Mullett Lake – 3 non-native species
- Long Lake – 24 non-native species



# Non-native Aquatic Plant Detections*

White Water Lily
*Nymphaea alba*
20 lakes

Purple Loosestrife
*Lythrum salicaria*
8 lakes

Flowering Rush
*Butomus umbellatus*
Cass & Five Channels

Spiny Water-Nymph
*Najas marina*
Austin

Brittle Water-Nymph
*Najas minor*
Thompson

Common reed
*Phragmites australis**
Holloway & Cass

Curly-Leaf Pondweed?
*Potamogeton* sp.*
21 lakes (many samples)

Eurasian Watermilfoil?
*Myriophyllum* sp.*
16 lakes (1-26 samples)

# Comparison with DEQ/EGLE surveys

11 lakes with DEQ/EGLE survey data (2016-2018): 13 non-native plants

- 5 of 13 identified at species level with *rbcL*

*Lythrum salicaria*
eDNA/DEQ + Extra eDNA
4/8

*Phragmites australis*
1/3 + 1

*Fallopia japonica*
0/1 + 1

*Butomus umbellatus*
1/2

*Frangula alnus*
1/1 + 2

- 4 of 13 with *possible* genus-level detections

*Myriophyllum* sp.
eDNA/DEQ + Extra eDNA
6/9 + 1

*Iris* sp.
2/4

*Typha* sp.
2/4 + 2

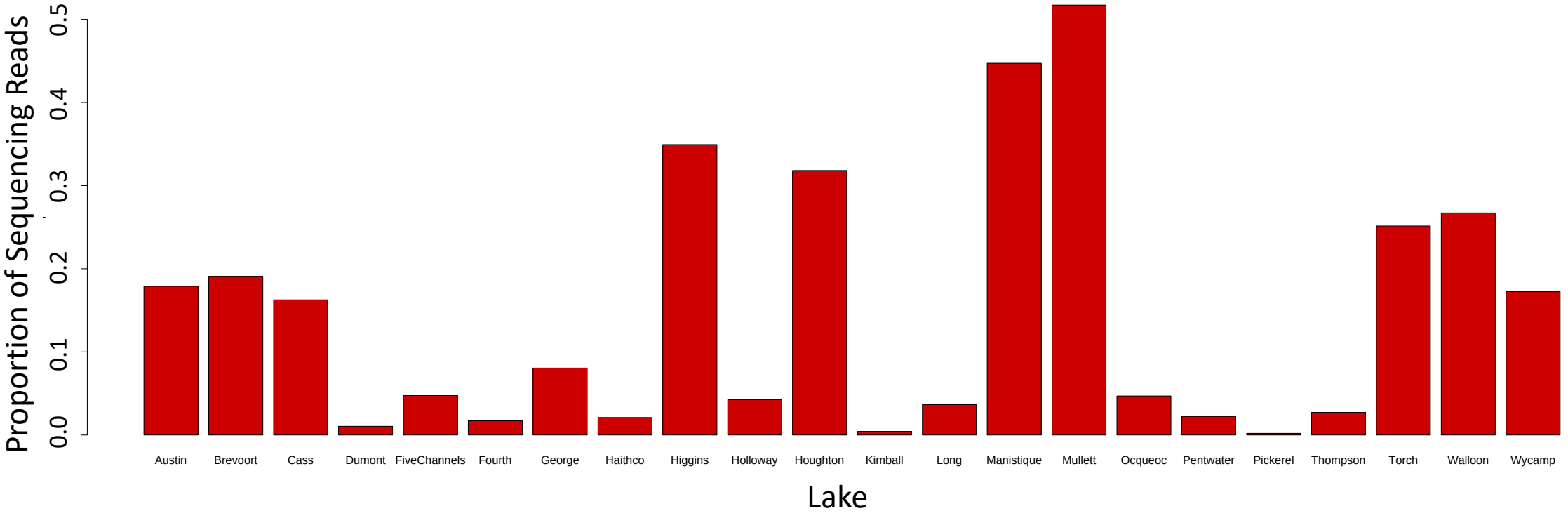
*Potamogeton* sp.
5/5 + 5

# Proportional Representation of Non-native Species

Non-native representation varied widely across lakes

- Kimball and Pickerel Lakes  < 0.5% non-native
- Mullett Lake >50% non-native

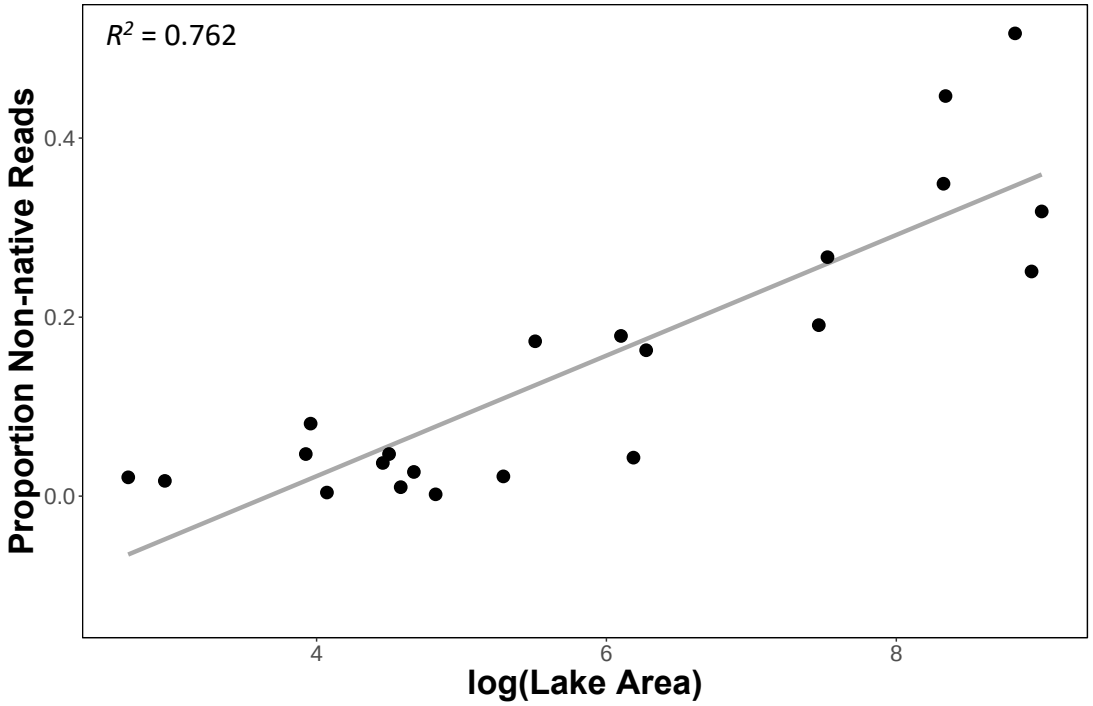
On average, 14.6% of reads were attributed to non-native plants



# GLMs: Proportional Representation of Non-native Plants

Proportion of reads attributed to non-native plants was related to lake area

- Torch Lake – 25%
- Houghton Lake – 32%
- Mullett Lake – 52%




| Model | df | logLik | AICc | ΔAIC | ωi |
|---|---|---|---|---|---|
| log.lake.area | 3 | 26.291 | -45.2 | 0.00 | 0.975 |
| Natural | 6 | 27.739 | -37.9 | 7.37 | 0.024 |
| Full | 9 | 28.900 | -24.8 | 20.45 | 0.000 |
| sqrt.boat.ramp.density | 3 | 13.574 | -19.8 | 25.43 | 0.000 |
| Anthropogenic | 5 | 15.225 | -16.7 | 28.55 | 0.000 |
| Null | 2 | 10.490 | -16.3 | 28.90 | 0.000 |
| dist.weig.pop | 3 | 11.807 | -16.3 | 28.97 | 0.000 |
| max.depth | 3 | 11.470 | -15.6 | 29.64 | 0.000 |
| IWS.stream.density | 3 | 11.211 | -15.1 | 30.16 | 0.000 |
| perc.agric.dev.land | 3 | 10.829 | -14.3 | 30.93 | 0.000 |
| log.area.upstream.density | 3 | 10.581 | -13.8 | 31.42 | 0.000 |

# Summary

## Surprising amount of terrestrial signal

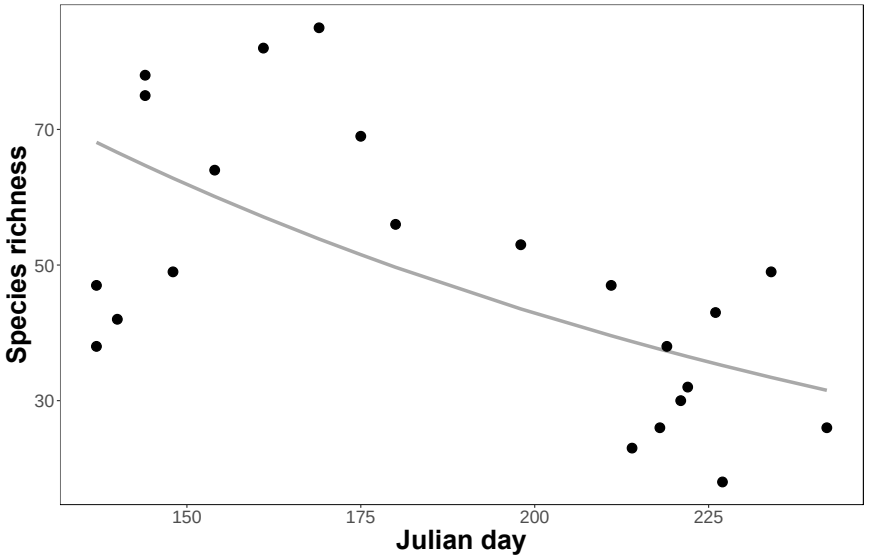
- Highest in areas with low human impact

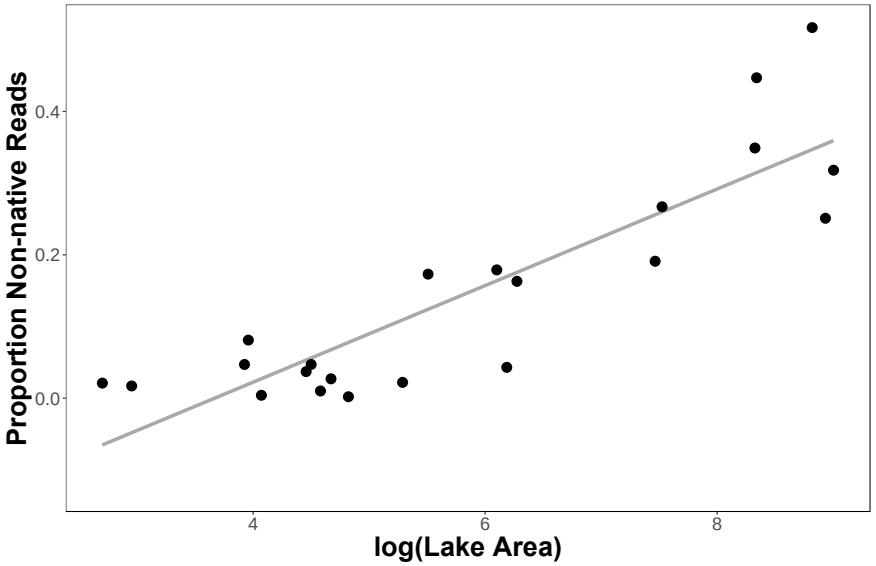
## Species richness highest earlier in the year

- Partially driven by terrestrial taxa
- Inconsistent with periods of higher growth (aquatics)

## Several common non-native spp. detected

- Non-native representation increased with lake size
- Limited resolution to species level with *rbcL*



# Considerations for eDNA Metabarcoding

## Marker choice is critical

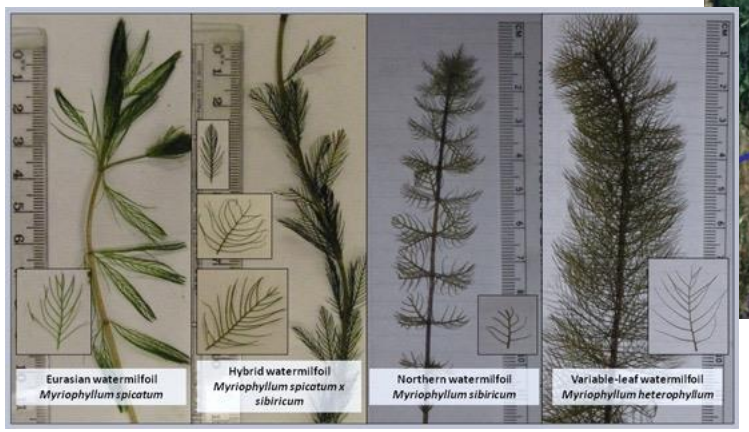
- Short gene fragments likely persist longer in the environment
- ***Long fragments allow for more species-level identification***
- Need a database of sequences at the chosen marker (BOLD, GenBank)

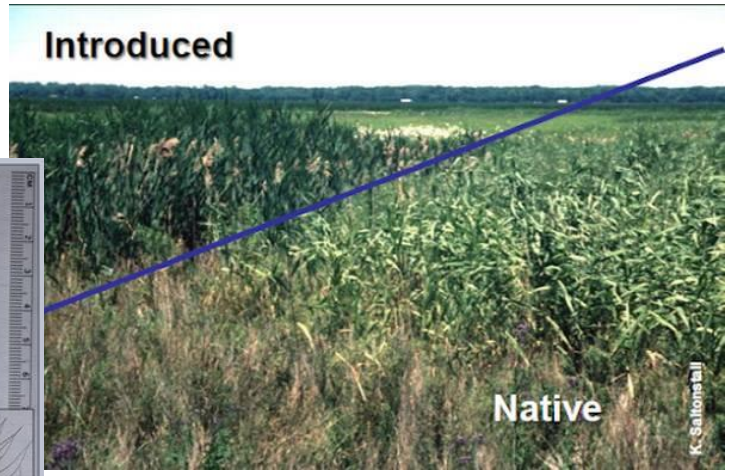
## Connections with abundance / biomass

- Relationships are influenced by a wide variety of biotic and abiotic factors
- Biotic – density, other community members, deposition, reproductive status
- Abiotic – pH, temperature, flow, sampling season

## Potential for false-positive detections

- Species-specific eDNA analysis
- May require follow-up sampling



# Recommendations

## Plant eDNA metabarcoding has potential…

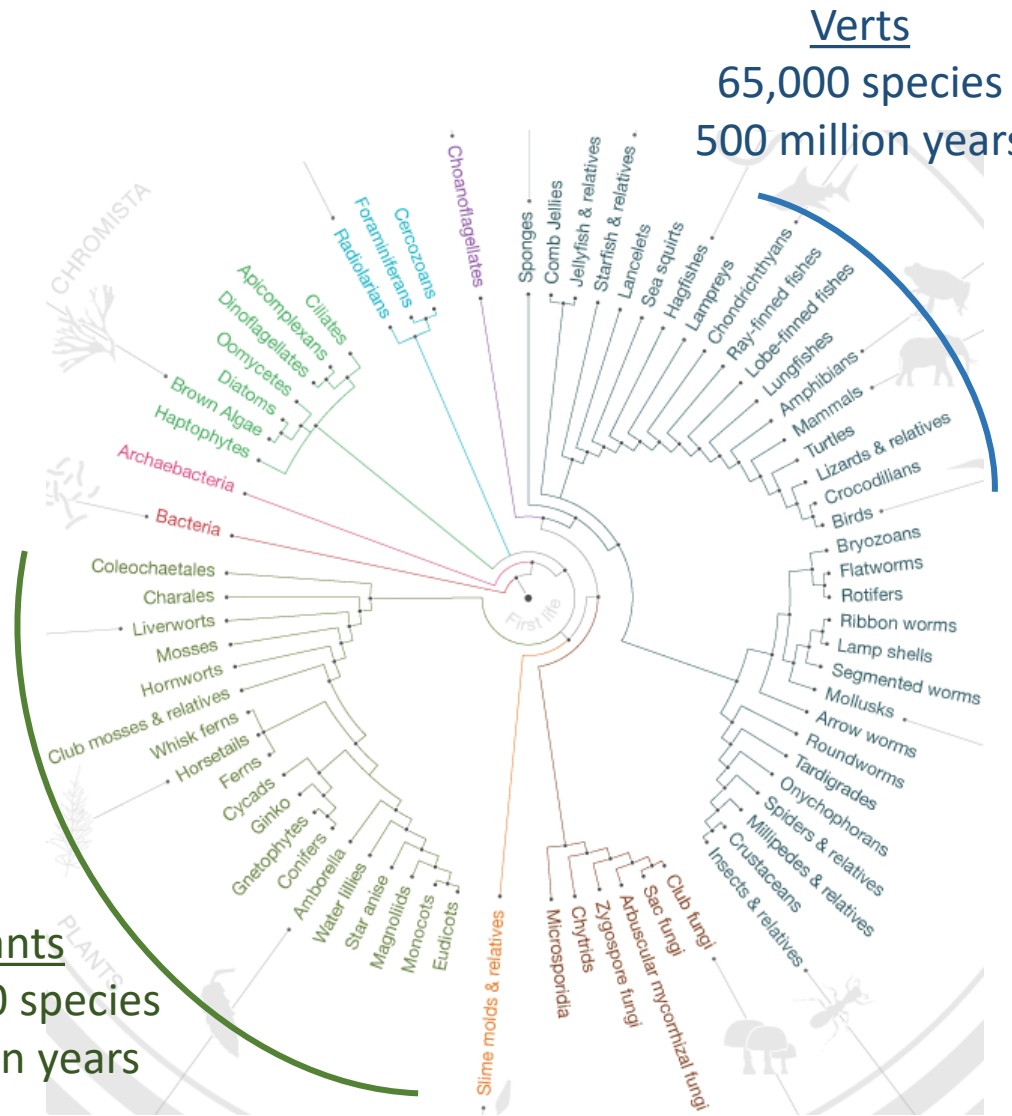
- Need markers with species-level resolution
- Timing of sampling is important

## Likely best if paired with other approaches

- Traditional surveys
- Species-specific eDNA assays (potentially using the same samples)

## Develop new markers to improve resolution

- Longer sequences?
- Plant group-specific primers?
- Expand on existing sequence repositories (ITS2, trnL, matK)



# Acknowledgements


## Michigan DNR / EGLE

Seth Herbst

Lucas Nathan

Billy Keiper

Sarah LeSage



## MSU Fisheries & Wildlife

Jeannette Kanefsky

Nick Sard (SUNY Oswego)

Lilian Pukk (EMU)

Antonio Castilla (MSU Texas)

Amanda Heathman

Ellie Weise



## Funding

Great Lakes Restoration Initiative

Michigan DNR

## MSU Research Support

RTSF Genomics Core

ICER High Performance Computing Cluster

# Proportion of rbcL plant sequences per taxonomic group per lake and total across all 22 sampled lakes