EPA
United States
Environmental Protection
Agency

DNA Metabarcoding-Based Identification of Fish Larvae for Invasive Species Early Detection

Joel Hoffman*, Christy Meredith, Erik Pilgrim, Anett Trebitz, Chelsea Hatzenbuhler, John Russell Kelly, Gregory Peterson, Julie Lietz, Sara Okum, John Martinson

Hoffman.Joel@epa.gov

Office of Research and Development
Great Lakes Toxicology and Ecology Division
Duluth, MN

UMISC 2020
NOV 2-6

Great Lakes Early Detection Monitoring

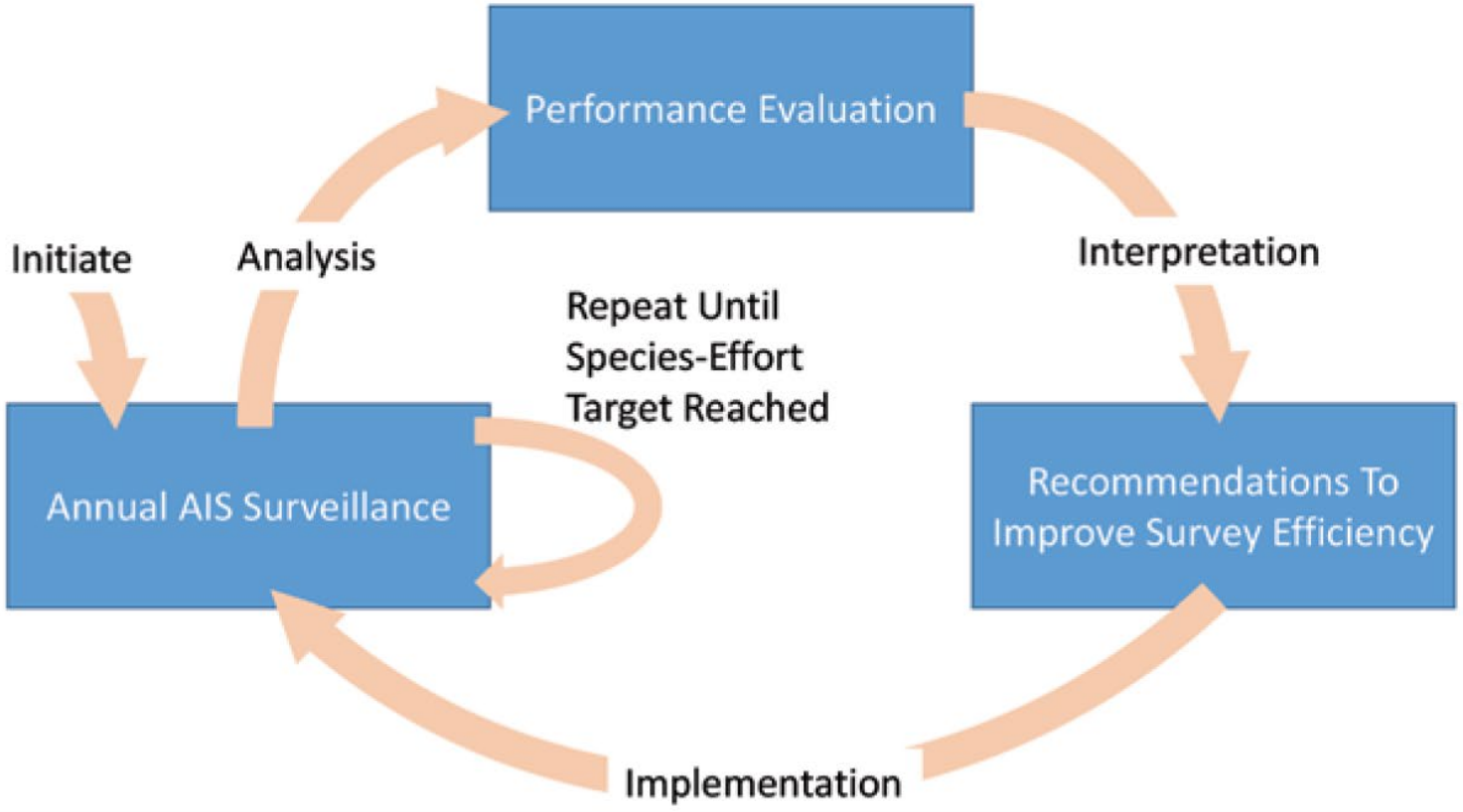


Figure 1. The adaptive monitoring framework used for the pilot Lake Superior aquatic invasive species early detection monitoring program.

Hoffman et al., 2016

Priority Surveillance Locations

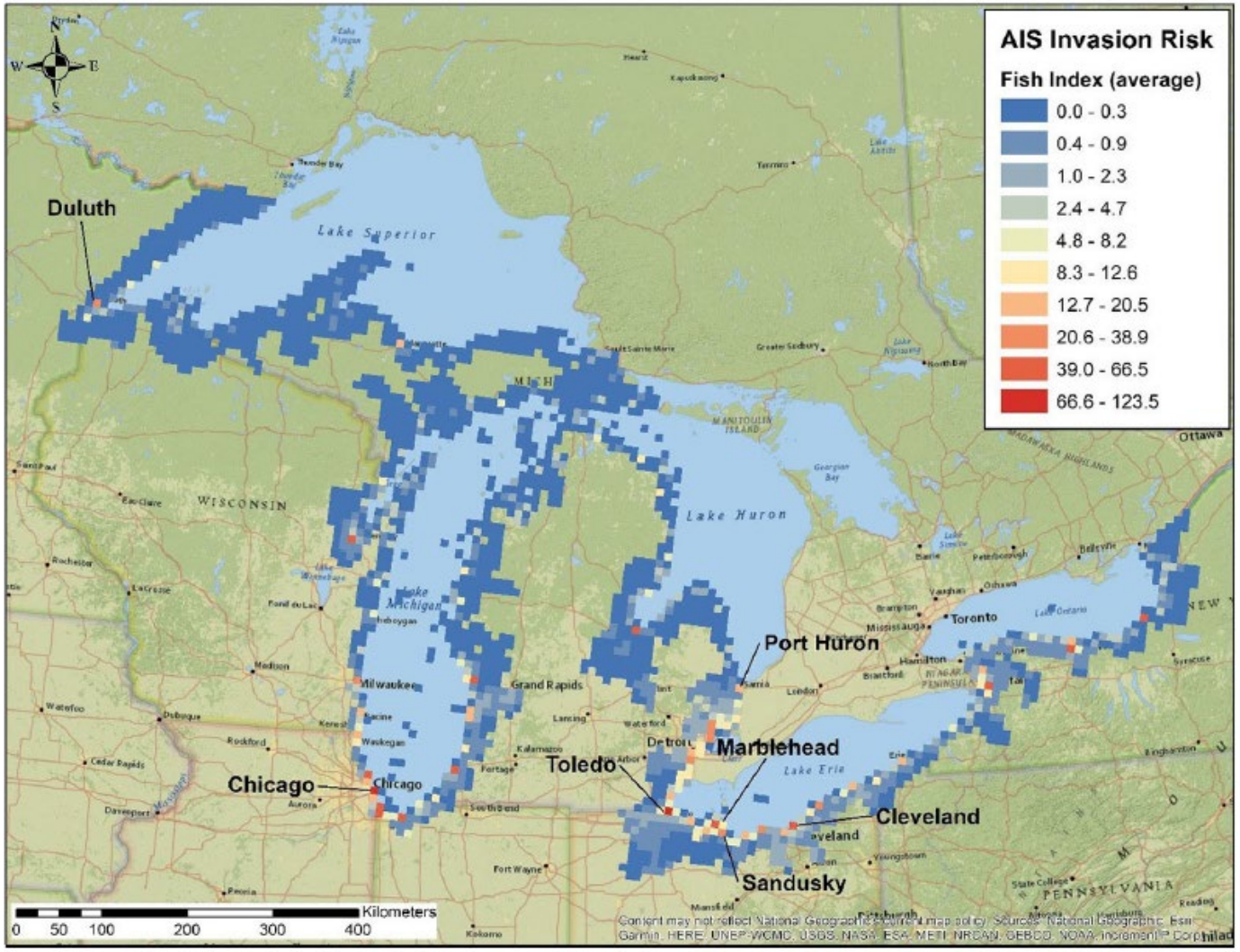


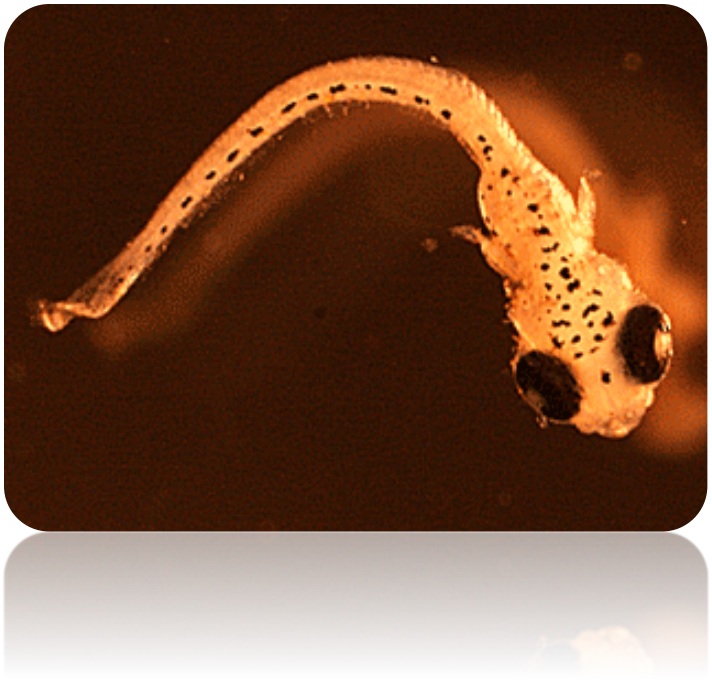
Figure 3. AIS risk scores by grid square for Fish.

Tucker et al., 2020

EPA
United States
Environmental Protection
Agency

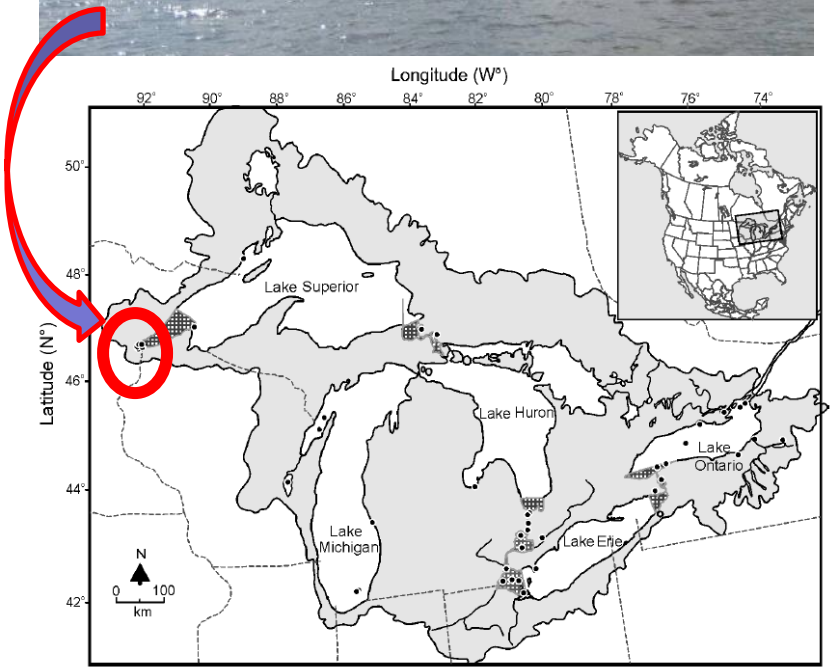
Why use fish larvae for early detection?

- Susceptible to ballast intake and transfer
- Presence of larvae may indicate a reproducing population (larvae more prevalent than adults)
- Relatively easy to sample and process in the field…
  but challenging to identify
  - –Takes time in the lab
  - –Quality of descriptions and keys
  - –Specimen condition
  - –Ability to recognize a newly introduced species



Research Objectives

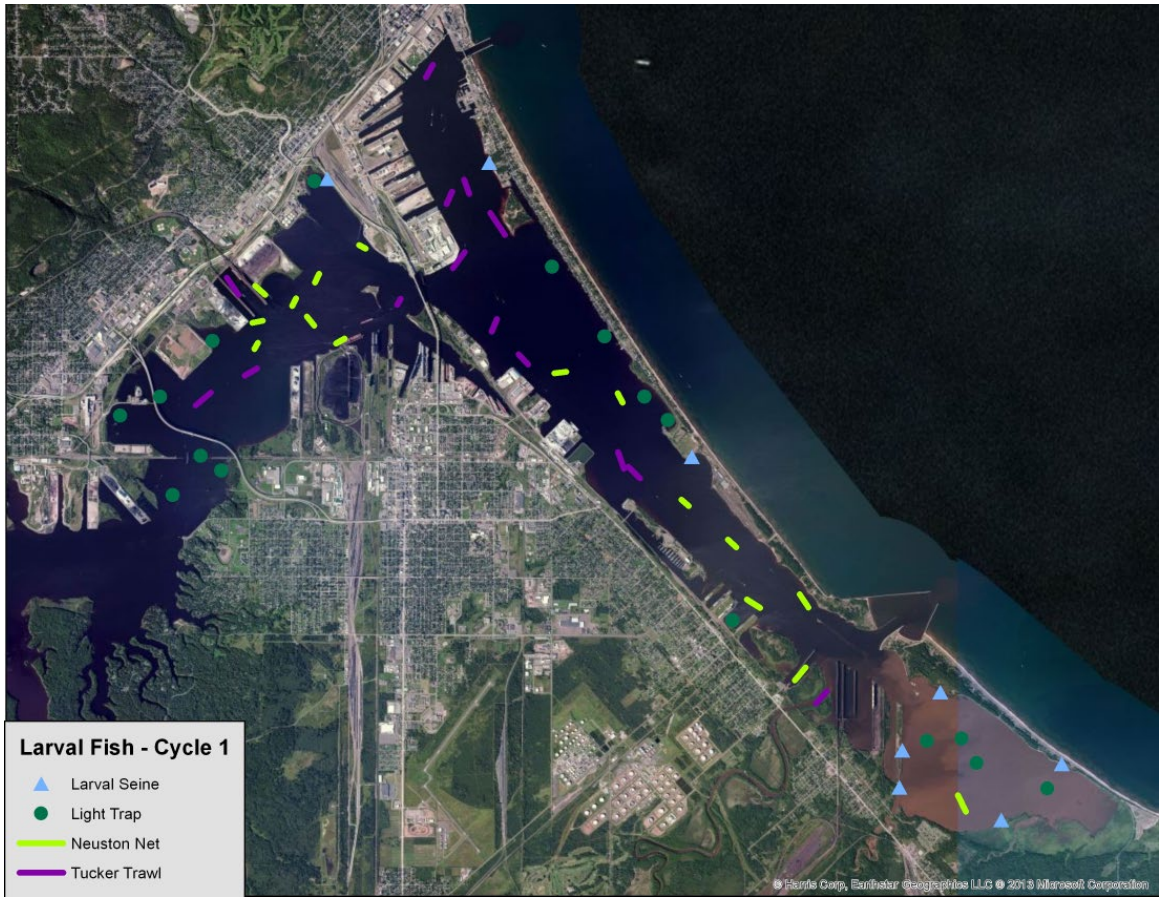
- Determine the taxonomic agreement between ichthyoplankton identified by DNA metabarcoding (i.e., HTS) versus morphological identification,
- Identify the sources of discrepancy between the two methods and quantify errors,
- Compare non-native fish detection based on HTS-based taxonomy of ichthyoplankton collections to ongoing optimized early detection surveys



Grigorovich et al. CJFAS 2003

Methods

- Field collection
- Morphology-based ID
- HTS-based ID
- Comparison
  - Taxonomic synonyms
  - Detection limits
  - Taxonomic Agreement
  - Error classification



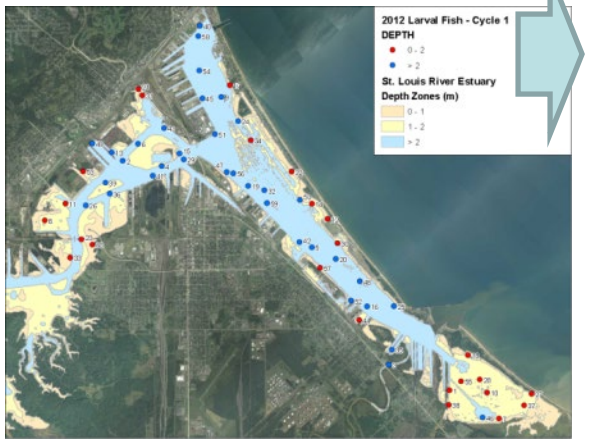
Total of 135 samples
- 54 late April
- 57 mid-May
- 24 late June

Total of 2655 eggs, larvae
- 303 eggs
- 166 unknown larvae (7%)
- 2186 identified larvae

Design and Methods Summary

United States Environmental Protection Agency

I. Probabilistic design

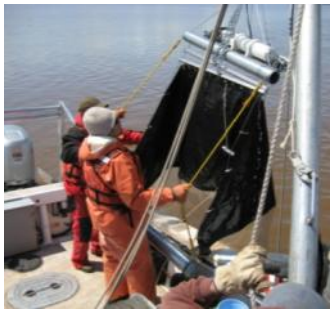


II. Field sampling

larval beach seine

light trap

Tucker trawl



neuston net

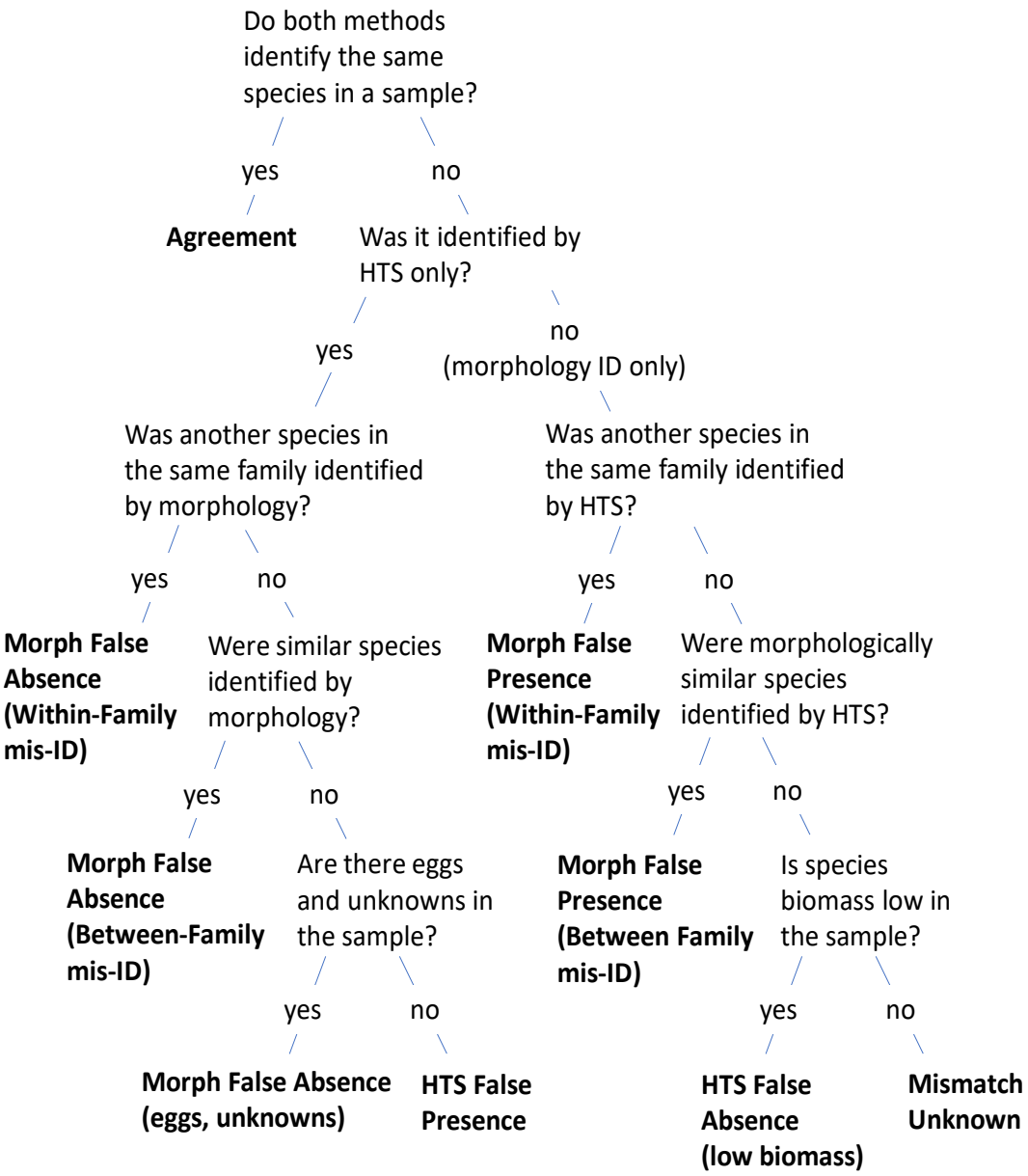
III. Morphological taxonomy

Reconstitute sample (with eggs and unknowns)

IV. Molecular taxonomy (HTS)

- CO1 (barcode region – BOCL)
- 1st run: general barcoding primers (general, for vertebrates and invertebrates)
- 2nd run: cocktail of barcoding primers specifically designed for use with fish

Error Classification



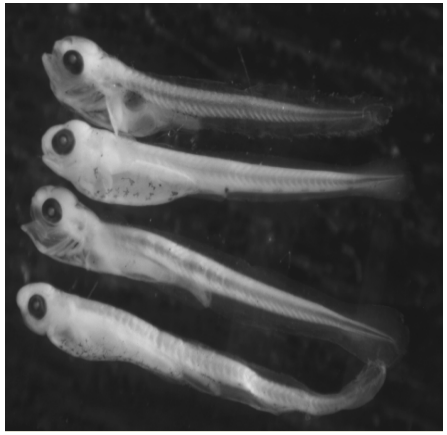
Side Note: Ruffe Mis-Identity

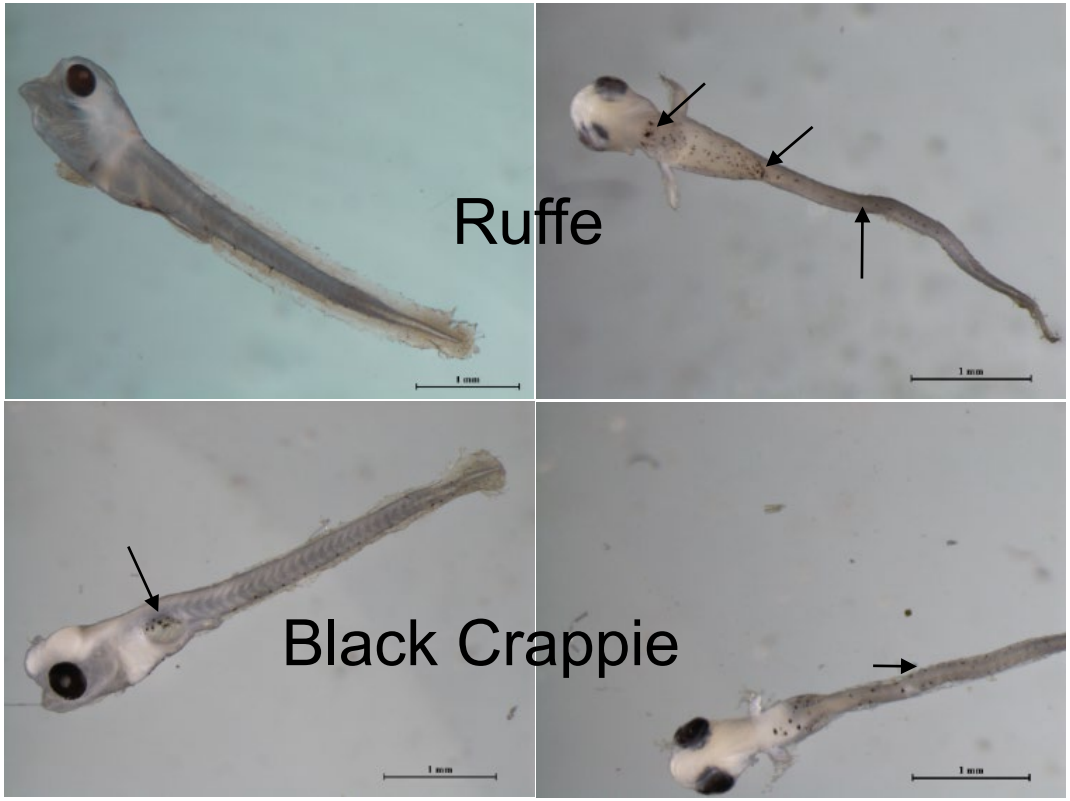
Ruffe

Johnny Darter

Yellow Perch

Logperch





Peterson, G.S., and J.E. Lietz. 2017. *J Great Lakes Research*

Results: Taxonomic Synonyms I

- Compared fishes in the Great Lakes (native and non-native) with existing introduction threat lists; neighbor-joining tree based on COI sequences from BOLD data system
- Morphological cryptic species pairs (as larvae)
  - *Lepomis macrochirus* - *Lepomis gibbosus* (Bluegill - Pumpkinseed)
  - *Catostomus commersonii* - *Catostomus catostomus* (White Sucker - Longnose Sucker)
- COI-cryptic species pairs (GenBank)
  - *Coregonus artedi* - *Coregonus hoyi* (Cisco - Bloater)
  - *Cottus bairdi* - *Cottus cognatus* (Mottled Sculpin - Slimy Sculpin)
  - *Notropis volucellus* - *Notropis buchanani* (Mimic Shiner - Ghost Shiner)

Results: Taxonomic Synonyms II

Species pairs with <3% difference (COI) – potential for introduction confusion

Genus	Extant native or introduced (*)	Threat list (*)	Primary concern
Alosa (shads, herrings)	*A. sapidissima	*A. immaculata	New invasion not recognized
Benthophilus (tadpole gobies)		*B. stellatus, *B. mahmudbejovi	Confused invader identity
Carpiodes (carpsuckers)	C. cyprinus, *C. carpio	*C. velifer	New invasion not recognized
Cottus (sculpins)	C. ricei	*C. gobio	New invasion not recognized
Enneacanthus (sunfishes)		*E. chaetodon, *E. obesus	Confused invader identity
Lepomis (sunfishes)	L. cyanellus	*L. symmetricus	New invasion not recognized
Pterygoplichthys (armored catfishes)		*P. disjunctus, *P. pardalis	Confused invader identity

Results: Detection Limits

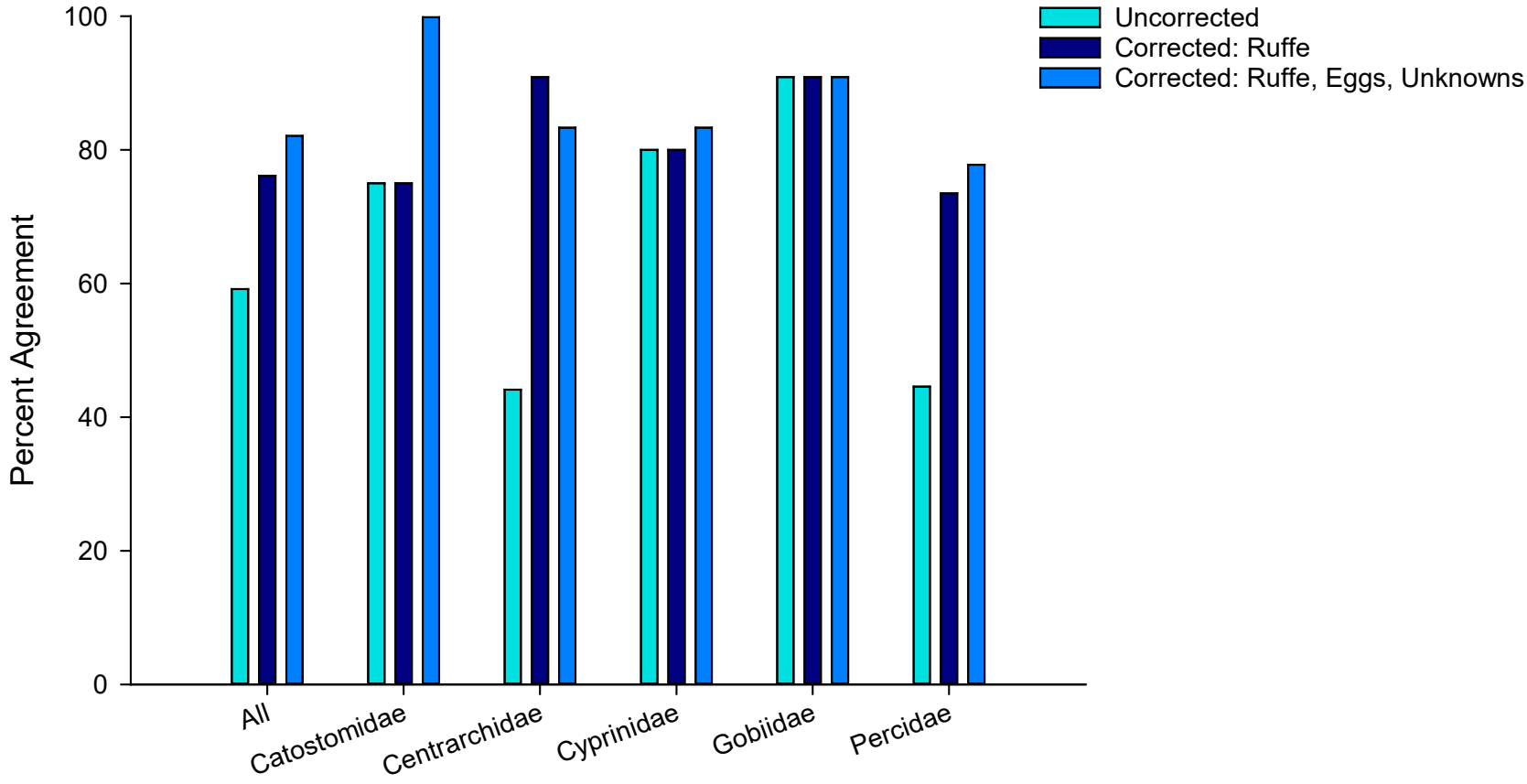
Estimated* biomass (mg wet wt) detection limits for HTS-based detection from field samples

Family	n	Primer Set	Mean biomass detected (±95% CI)	Minimum biomass detected	Maximum biomass not detected
Cyprinidae	88	1	5.37 (10.06)	0.22	37.07
	84	2	0.44 (0.11)	0.22	0.85
	76	1 & 2	0.30 (0.07)	0.22	0.43
	93	1 or 2	3.96 (3.99)	0.22	12.46
Gobiidae	90	1	NA	0.21	NA
	91	2	0.74 (0.33)	0.21	1.67
	78	1 & 2	NA	0.21	NA
	96	1 or 2	NA	0.21	NA

* based on published length-weight regressions

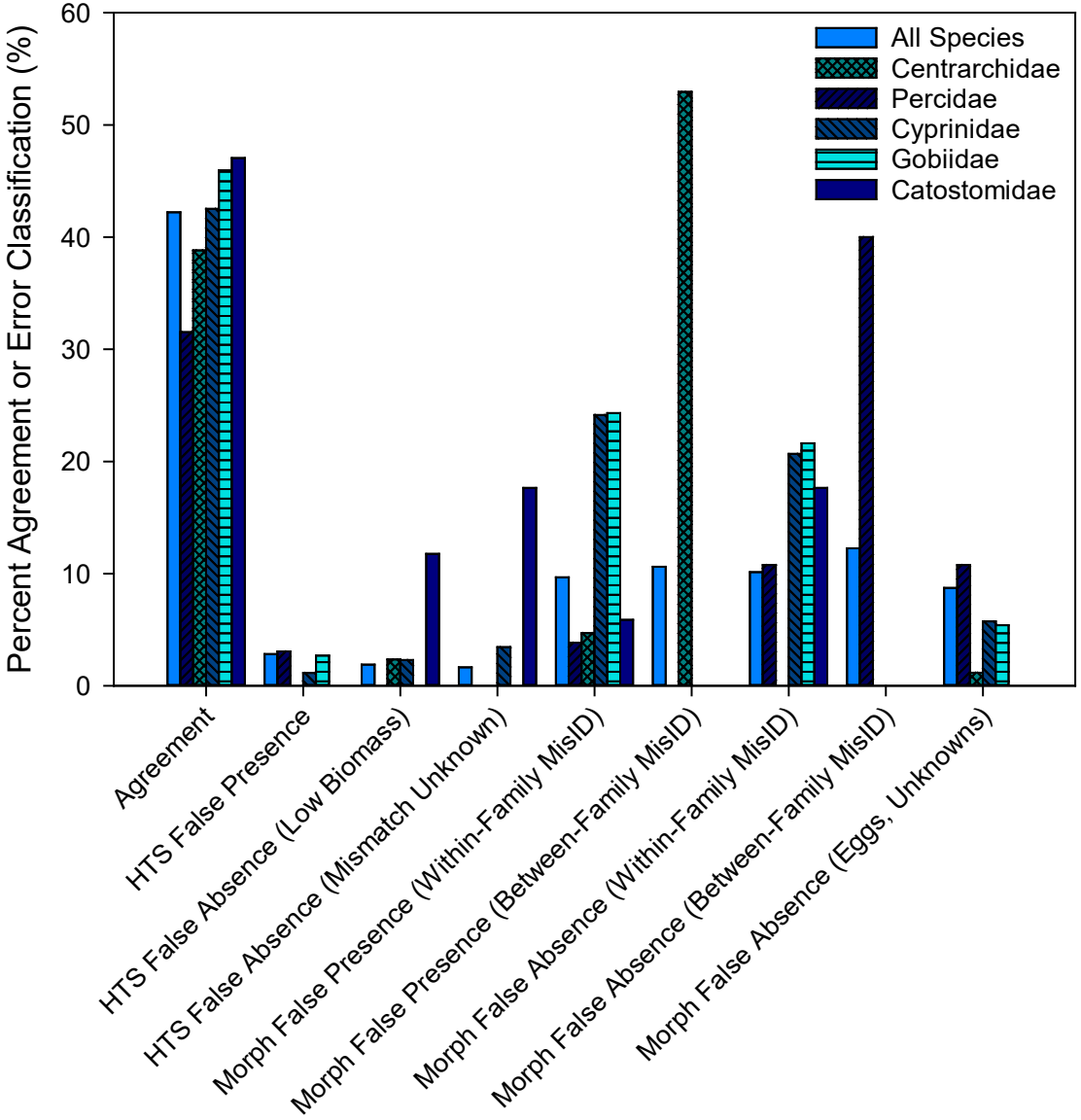
Results: Family-level Agreement

Agreement between HTS vs Morph-ID based on family presence in field samples

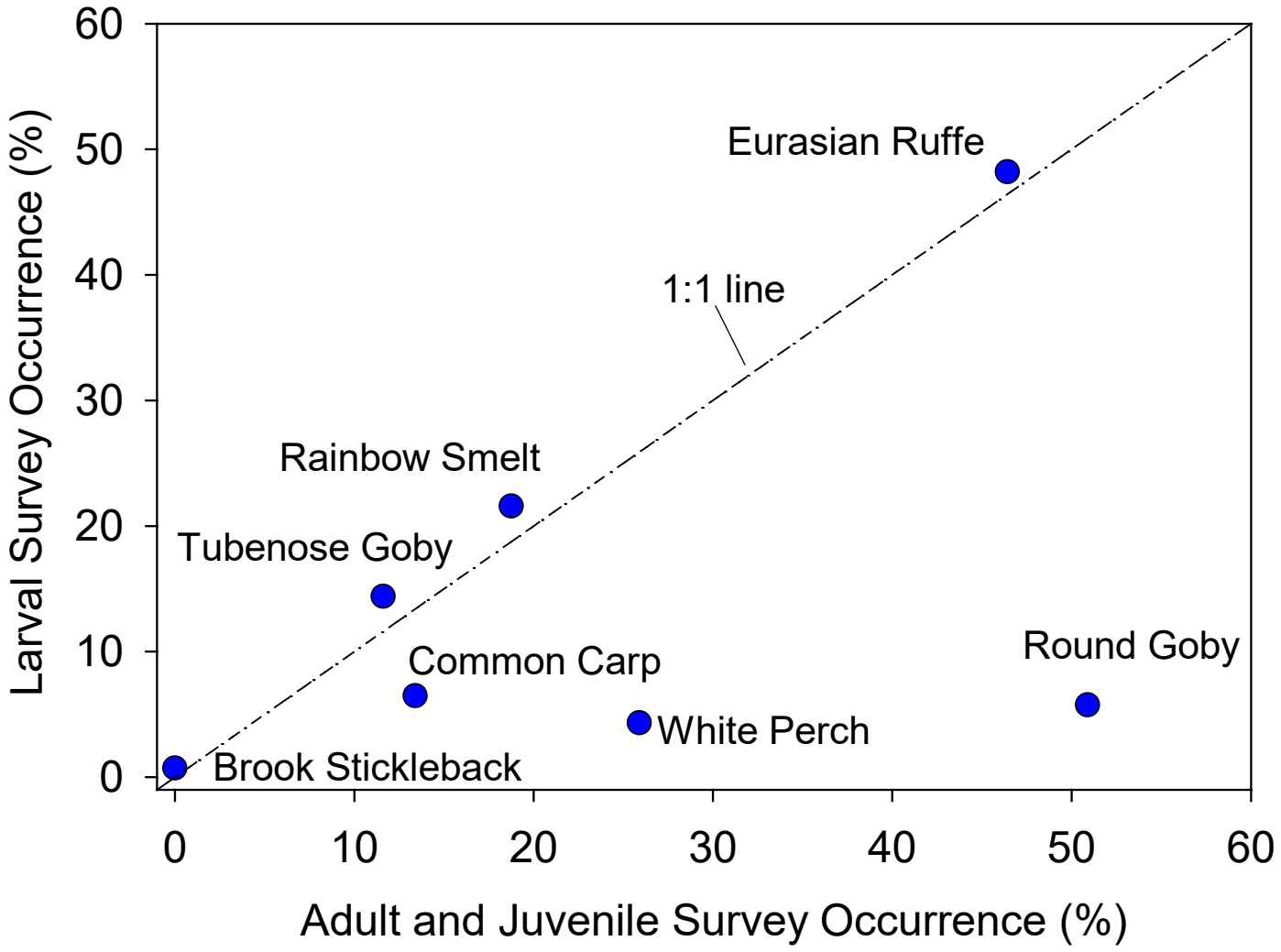


Results: Species-level Agreement

Species agreement between HTS vs Morph-ID based on species presence in field samples and using the error classification tree



Results: Comparison of Survey Performance for Non-Native Species



Summary

- Substantial differences between methods
  - Largely morphology-based taxonomy failures: morphologically similar species (e.g., Ruffe), specimen stage (eggs), specimen condition (larvae)
  - Few HTS-based errors: sequencing error, low biomass
- HTS-based detection required more mass than generally previously reported
- HTS-based taxonomy found more species (after accounting for false positives)
- HTS-based taxonomy revealed errors in existing non-native species descriptions (e.g., Ruffe)
- Novel ichthyoplankton survey had good agreement with an ongoing survey optimized for non-native species early detection

EPA United States Environmental Protection Agency

Conclusions

- Overall, HTS-based taxonomy was more accurate than morphology-based taxonomy
- Routine error classification is needed to refine both survey design and DNA sequencing and bioinformatic processing methods (especially, rare sequences)
- Both taxonomic methods yielded false positives and false negatives, and the error rate for HTS-based taxonomy was substantially lower than for morphology-based taxonomy
- Both taxonomic methods have limits to species-level resolution (i.e., cryptic species groups)
- Recommend a tandem approach to potentially confer benefit on both methods

Contact

Joel Hoffman, PhD
US EPA Office of Research and Development
Center for Computational Toxicology and Exposure
Great Lakes Toxicology and Ecology Division
hoffman.joel@epa.gov
218-529-5420

Acknowledgements

Funds from the Great Lakes Restoration Initiative supported the position of author C. Hatzenbuhler, J. Lietz, C. Meredith, and S. Okum. Will Bartsch, Katherine Bentley, Tyler Billehus, Hannah Coe, Tim Corry, and Jill Scharold provided field and laboratory assistance.