Environmental DNA surveys for early detection of invasive species in the Lower Hudson

Lower Hudson PRISM Final Report

Michael Tessler

Overview

The research presented here used environmental DNA (eDNA) and ecological niche modeling to provide insights into aquatic invasive species in the Lower Hudson PRISM (LHPRISM). Plants were the primary taxonomic focus, but animals were also studied. Students were trained on molecular methods. Outreach materials — a video and photos — were produced. In short, all deliverables have been fulfilled for the LHPRISM 4rd quarter, and they are individually listed and detailed below.

1. Field videography and photography

The video produced focuses on the Hudson River and concisely presents information on environmental DNA, invasive species, the research conducted for the LH PRISM, the Invasive Species Strike Force, and how to avoid spreading invasives ("LH_video_final.mp4"). Additionally, for this deliverable, I have provided 25 edited photos. The photos portray the molecular fieldwork, physical surveys conducted by the Invasive Species Strike Force, aquatic plants, and the Lower Hudson generally.

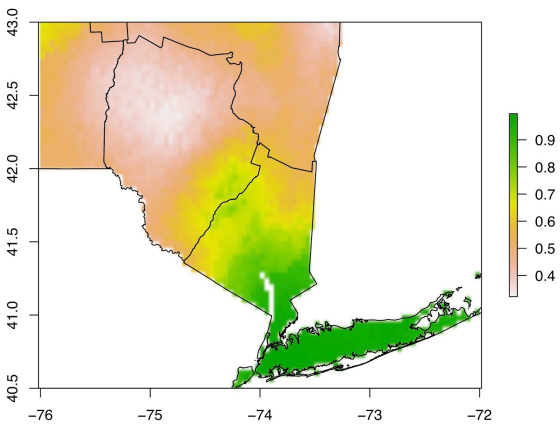
2. Ecological niche models for LH PRISM Tier 1, 2, and 5 species

Background: Ecological niche models (ENMs) provide a way to predict where a species may have suitable habitat, based on known distributions of the species and climatic data for a region. Overall, the goal of this research is using modern techniques to help find invasive species. ENMs make this possible by pointing out where species are likely to be found. Based on the models described below, the LH PRISM and Hudson River appear to be an area where many species may have suitable habitat.

Methods: Point records for all 153 species identified in the invasive species categorization for the Lower Hudson (https://www.lhprism.org/document/lower-hudson-species-categorization; Tier 1, 2, and 5) were downloaded from the Global Biodiversity Information Facility (GBIF; gbif.org), an open source database of species global occurrence records. With these data, ENMs were created using point records from North America. The other input data for ENMs consisted of 19 bioclimatic variables downloaded from Worldclim.org at 30 arc-seconds resolution (Hijmans et al., 2005). Independent ENMs were constructed using maximum entropy [Maxent (Phillips, Anderson & Schapire, 2006)] in the R package ENMeval (Muscarella et al., 2014). In order to combat spatial autocorrelation of large datasets, point records that contained more than 500 points were spatially thinned to that amount using the R function thin.max.R (gist.github.com/danlwarren/271288d5bab45d2da549) to maximize the differences between all points. This insures models incorporate the full environmental suite of characters that the species is known to persist in. Furthermore, background data were obtained for models using a minimum convex polygon of pre-thinned occurrence points. Multiple linear models were then constructed for each species with best models chosen based on high AUC scores and model feature class complexity. Maps have been trimmed to cover lower New York State and include PRISM boundaries, centering on the LH PRISM boundary. Green areas on the ENM maps indicate areas that have environmental characteristics best suited for a given species, while grey indicates that an area is likely inhospitable for the species.

Results and discussion: Out of the PRISM Tier 1, 2, and 5 invasive priority species, well over 100 had sufficient data to produce predictions in the LH PRISM region. The most important

overall finding was that 70 species have ENM models that indicate the entirety of the LH PRISM likely provide reasonably suitable habitat (e.g., the map shown for *Cabomba caroliniana*).



Furthermore, a number of species are predicted to extend furthest north along the Hudson. A total of 19 produced ENMs predicted ranges that barely extend into the extreme southern range of the LH PRISM area, suggesting this is the northern extent of the species' suitable habitat; however, this can be hard to say for new invasive species. Notably, this includes six out of the 30 aquatic species (all plants) targeted. Of the species without enough data to process, eight had no GBIF records in North America. An additional 12 contained predicted distributions that fell outside of the LH PRISM boundary. A disportionate number of these 20 species that could not have ENM maps produced were species that are either aquatic plants or terrestrial invertebrates.

3. Fieldwork and DNA quantifications

Fieldwork was conducted in collaboration with Lindsay Yoder (New York-New Jersey Trail Conference) and her AmeriCorps members (Invasive Species Strike Force - Aquatic). During this fieldwork, Lindsay and the AmeriCorps members conducted rake-toss surveys (a form of physical survey) to determine the native and invasive aquatic plants at a given site, while my focus was on collecting water samples for eDNA determination of invasive plants and animals in the area. In total 70 samples were collected from 37 surveyed sites along the Hudson River. The majority of sites (33) were sampled on two separate occasions — the beginning of July (7/1/2019-7/2/2019) and the end of August (8/21/2019-8/22/2019) — to determine seasonal differences and ensure thorough sampling.

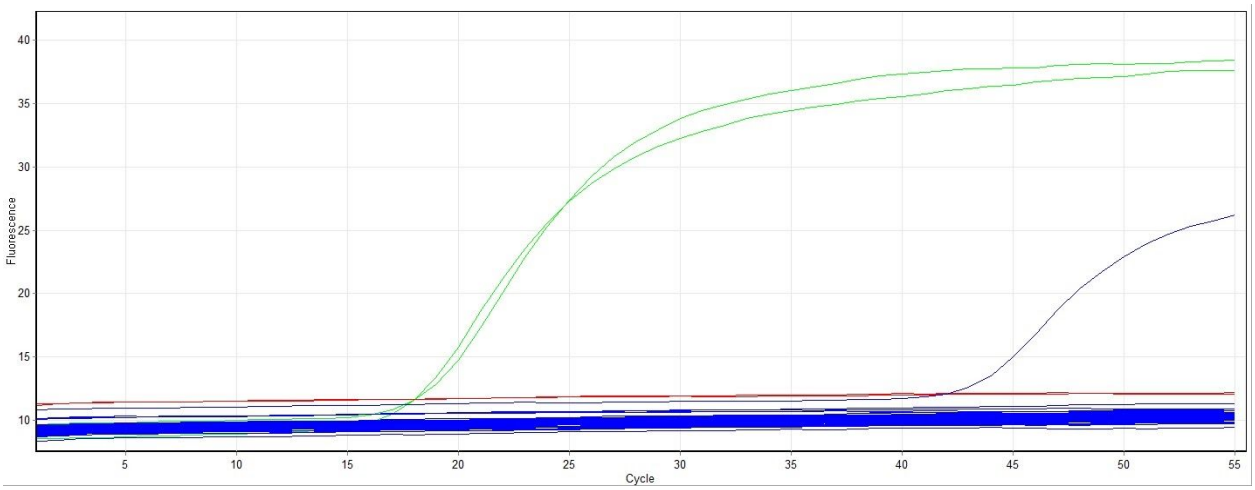
Trapa natans (the image above shows a large area of this species growing on the Hudson) and *Vallisneria americana* were the most common species found in physical surveys. An additional 9 plant species were documented with the physical surveys: *Ceratophyllum demersum*, *Elodea canadensis*, *Lemnas minor*, *Myriophyllum spicatum*, *Najas minor*, *Najas* sp., *Potamogeton crispus*, *Potamogeton perfoliatus*, and *Wolffia* sp. *Dreissena polymorpha*, the zebra mussel, was also recorded. Quantifications of eDNA varied from 10-33 ng/µL. I have found these to be typical levels for freshwater eDNA extractions, and this concentration typically works well for downstream sequencing and analysis. Furthermore, the level of variation in quantities is typical, and can depend on factors such as water clarity, organismal density at a site, or if an organism recently shed a lot of DNA in the site.

4. eDNA analysis

Background: Metabarcoding is an emerging tool for studying plant, animal, and microbial communities. The technique sequences a single locus for lineage (group) of organisms with DNA present in an environmental sample. The sequences are then used to identify the species within that lineage by comparing them to reference databases. Plants, and particularly aquatic plants, are understudied in regards to this method. The principal focus of this study was accordingly to compare metabarcoding and physical surveys for aquatic plants in the Lower Hudson. To do this, 70 water samples were collected from the Hudson River. The full methods and results from this are presented at the bottom of this document in the draft manuscript titled "Comparing metabarcoding and rake-toss surveys of aquatic plants in the lower Hudson River, NY." Overall, the results indicate that metabarcoding and physical surveys are highly complementary and best used in concert. Additional results for an aquatic plant using qPCR as well as metabarcoding results for other organisms are presented below.

qPCR: In addition to the metabarcoding work on aquatic plants, I used a quantitative PCR (qPCR) analysis in an attempt to survey for hydrilla, given this species' potential for spread in the LH PRISM boundary. This technology is more established for eDNA work than

metabarcoding, and has been used in a few previous studies on hydrilla (Matsuhashi et al., 2016; Gantz et al., 2018). The analysis was conducted with a Qiagen Rotor Gene Q using hydrilla-specific primers and protocols developed by prior researchers (Matsuhashi et al., 2016). With the 70 samples, 3 positive controls were used [hydrilla tissue extract, Croton Reservoir eDNA (which, has hydrilla present), and Hudson/Croton river junction eDNA], as well as a negative (lab water). All samples were run in duplicate to ensure quality.



The above plot shows the qPCR analysis looking for hydrilla in the 70 samples. Green lines indicate positive controls (i.e., hydrilla DNA), showing amplification peaks. Blue lines represent eDNA samples from the Lower Hudson. The results appear to be negative for all samples. One duplicate had a modest peak at a late cycle. While this could be a positive result, long cycling times can produce false-positives. This sample was run again twice and did not result in amplification peaks; accordingly, it seems most convincingly to be negative for hydrilla like the other samples.

Metabarcoding other species: Plant metabarcoding produced numerous sequences for terrestrial species. This is both a pro and potentially a con of the method, as discussed along with the methods used in the draft manuscript. However, sequences matched terrestrial species in Tiers 3-5. Two Tier 3 invasive plants were identified to species: *Aralia elata* (ITS2) and *Humulus japonicus* (*trnL2*). A total of 13 of additional Tier 4 invasive plants were identified to the species level: *Acer platanoides* (ITS2), *Ailanthus altissima* (ITS2, *rbcL*, and *trnL*), *Lonicera japonica* (*trnL2*), *Lonicera morrowii* (ITS2), *Lythrum salicaria* (ITS2), *Microstegium vimineum*

(ITS2), *Morus alba* (ITS2), *Myosotis scorpioides* (ITS2), *Phalaris arundinacea* (ITS2), *Phragmites australis* (ITS2), *Polygonum cuspidatum* (*rbcL* and *trnL*), *Robinia pseudoacacia* (ITS2), and *Rubus phoenicolasius* (ITS2). One Tier 5 invasive plant appeared to be identified to species: *Fallopia sachalinensis*.

Animal metabarcoding was conducted using a general eukaryote 18S mini barcode (Amaral-Zettler et al., 2009), as well as one 16S primer for fishes and one 16S primer for arthropods (Stat et al., 2017). The 16S primer set for fishes found a total of 20 native species: *Alosa aestivalis*, *Alosa pseudoharengus*, *Anchoa mitchilli*, *Anguilla rostrata*, *Apeltes quadracus*, *Brevoortia tyrannus*, *Dorosoma cepedianum*, *Enneacanthus gloriosus*, *Fundulus diaphanus*, *Fundulus heteroclitus*, *Lepomis gibbosus*, *Lepomis macrochirus*, *Menidia menidia*, *Micropterus salmoides*, *Morone americana*, *Morone saxatilis*, *Paralichthys dentatus*, *Pomatomus saltatrix*, *Tautoga onitis*, and *Trinectes maculatus*. A few additional species were clearly native, but the locus could not discern between two similar species. *Anguilla rostrata* (the American eel) is of note, as it is a species that has declined greatly and is often monitored for conservation purposes (Morrison & Secor, 2003). Two non-natives were also detected: *Aplodinotus grunniens* and *Salmo trutta*.

The arthropod 16S primers picked up fewer species. A match to the non-native crayfish *Orconectes rusticus* was detected, as were native crabs such as *Callinectes sapidus* and

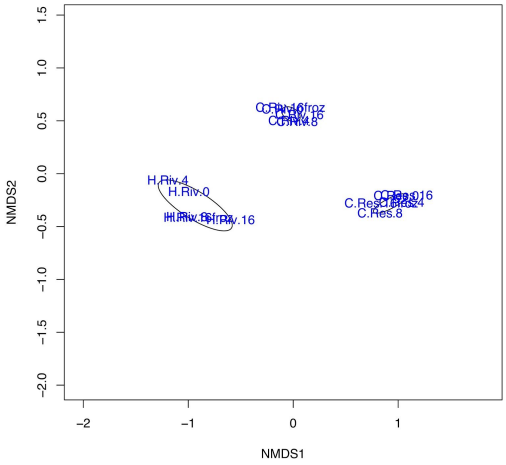
Rhithropanopeus harrisii. Other native arthropods included *Graminella nigrifrons* and *Narceus americanus*. A few copepods (e.g., *Eurytemora affinis*), cladocerans (e.g., *Bosmina longirostris*), and other arthropods (e.g., *Milnesium tardigradum*) were also detected. Unfortunately, these are harder to confirm to the species level, as many of the detected organisms likely have fewer relatives with sequences in the 16S reference database and less geographic information is available for them.

18S is used to identify sequences across all eukaryotes. Given that this is a mini barcode across a large diversity of life, it often cannot produce species-level results. The locus also picks up on a large diversity of protists along with a smaller fraction of multicellular organisms like plants and animals. This is likely because protists are abundant and often will be caught on a filter as whole cells with intact DNA, unlike the eDNA from plants and animals that is likely to be highly fragmented. Still, a number of sequences matching multicellular organisms were found. Here I will highlight a few animals, as little additional information on plants was recovered beyond what is in the draft manuscript. Two sites had *Acipenser* (sturgeon) sequences, which is noteworthy as these fishes are of conservation concern. Four invasive animals were also detected: *Corbicula fluminea*, *Cyprinus carpio*, *Dreissena polymorpha* (equally matched another species that was out of range), and *Tinca tinca*.

5. Short summary of student internships

Overview: Two students were engaged in optimizing loci for plants and animals. These students learned lab procedures (DNA extraction, gel electrophoresis, and PCR) as well as basic analysis of eDNA metabarcoding data. Furthermore, one student was present for the July sampling, and was able to gain fieldwork skills (collecting samples and recording notes), as well as sample filtration.

Lab work: Lab work was conducted to optimize protocols on loci I had not previously used. In particular, a new *trnL2* chloroplast primer was optimized for working on plants. This primer set amplifies a short stretch of DNA compared to the other plant loci used in this study for metabarcoding. Shorter primers tend to work best for amplifying highly degraded genetic material, which can be helpful for eDNA work where DNA is typically degraded. Additionally, animal primers have been optimized, and have been particularly successful at detecting fishes with a 16S primer pair set. The students also worked to determine if our storage method caused any differences



in metabarcoding output. Regardless of time stored in lysis buffer (0-16 days), the species assemblages recovered were similar and nonmetric multidimensional scaling (NMDS) analysis, pictured above, found little difference between samples preserved for different amounts of time. Samples in this visualized NMDS used fish 16S eDNA and were clustered tightly by site (Croton Reservoir, Croton River, and where the Croton River meets the Hudson River) and showed little patterning by preservation time. This result helped provide confidence in our general methodology.

References

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PloS one* 4:e6372.

Gantz CA, Renshaw MA, Erickson D, Lodge DM, Egan SP. 2018. Environmental DNA detection of aquatic invasive plants in lab mesocosm and natural field conditions. *Biological Invasions* 20:2535–2552.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.

Matsuhashi S, Doi H, Fujiwara A, Watanabe S, Minamoto T. 2016. Evaluation of the environmental DNA method for estimating distribution and biomass of submerged aquatic plants. *PloS one* 11:e0156217.

Morrison WE, Secor DH. 2003. Demographic attributes of yellow-phase American eels (*Anguilla rostrata*) in the Hudson River estuary. *Canadian Journal of Fisheries and Aquatic Sciences* 60:1487–1501.

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014. ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity forMaxentecological niche models. *Methods in ecology and evolution / British Ecological Society* 5:1198–1205.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling* 190:231–259.

Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES, Bunce M. 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports* 7.