Estimating Sampling Effort for Early Detection of Non-Indigenous Benthic Species in the Toledo Harbor Region of Lake Erie

Jeffrey L. Ram, Fady Banno, Richard Gala, Jason P. Gizicki, Donna Kashian

Wayne State University
Many non-native species have been transported into the Great Lakes in ballast water.
Nearby ports threaten our waters

For example, Toledo, a major port of the Great Lakes, and Near the “fishiest waters of the Great Lakes,” western Lake Erie.
Nearby ports threaten our waters

Toledo is a major recipient of ballast water from outside the Great Lakes

Table E-2. Ballast water discharges at U. S. Great Lakes during 2006-2007. Vessels whose original source of ballast water (prior to ballast water exchange) came from outside the Great Lakes.

<table>
<thead>
<tr>
<th>US Great Lake Port</th>
<th>Tanks Discharged</th>
<th>Volume Discharged (metric tons)</th>
<th>Vessels Discharging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duluth</td>
<td>407</td>
<td>184,844</td>
<td>58</td>
</tr>
<tr>
<td>Toledo</td>
<td>85</td>
<td>65,335</td>
<td>13</td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green Bay</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milwaukee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oswego</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ludington</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erie</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...the port of greatest concern for receiving sufficient propagules and providing the most suitable habitat is Toledo (EPA report EPA/600/R-08/066F, 2009)
RamLab early detection studies in Toledo Harbor

HOW: 2010 Pilot study  
EPA: GLRI 2011-2014 Project: 
Title: Invasives Early Warning Systems Validation in Toledo Harbor

Main objectives:
• To demonstrate sites and methods for surveying species in Toledo Harbor to search for rare or new non-native species
• To identify taxonomic services that any agency might use for accurate non-native species identification
Design based on Trebitz et al., 2009

- Duluth Superior Harbor

![Diagram showing benthic sampling sites and non-native richness, with labels for mixed character and primarily riverine, and a scale of 5 km.]
2010 Pilot Study

25 lb bottom dredge

0.5 mm mesh sieve

27 samples on 3 dates in late Sept/early Oct 2010
All locations recorded for depth, GPS, vegetation, temperature, and time collected

Maumee River

5 km
RamLab early detection studies in Toledo Harbor

Benthic samples
Classical (Morphological) taxonomic identification by EcoAnalysts, Inc.
Molecular taxonomic identifications by the Canadian Center for DNA Barcoding (CCDB)

Questions
What species do we detect? What non-natives?

Molecular taxonomy: Can molecular taxonomy enhance the reliability and specificity of classical identifications? What problems arise?

How close are we to “over-sampling”?
• All 27 samples were scanned for representative organisms for each sample: each site was represented by at least 5 – 10 morphologically different specimens.

• 142 specimens were sent to Ecoanalysts for identification.

• Ecoanalysts identified many different taxa: 80% to species level; remainder only to family or genus.
Pilot study

• All 27 samples were scanned for representative organisms for each sample: each site represented by at least 5 – 10 different species.
• 142 specimens were sent to Ecoanalysts for identification.
• Ecoanalysts identified many different taxa: 80% to species level; remainder only to family or genus.
• Non-native species include *Branchiura sowerbyi, Bithynia tentaculata, Corbicula fluminea, Dreissena polymorpha, Dreissena bugensis, & Lipiniella sp.*
This slide showed a complete species list with unique and duplicate sightings identified. To see the list, write to the authors at jeffram@med.wayne.edu
DNA barcode sequencing

3 major goals:

1) **Quality control of morphology-based taxonomic identifications.** Use DNA sequencing to verify the accuracy and consistency of classical *morphological* identifications.

2) **Identification of rare (possibly non-native) species in the samples.** Verify the identity of rarely seen species.

3) **Identification of immature or ambiguous organisms.** Chironomids (midges) are extremely difficult to identify to species using morphology-based taxonomy methods. Many oligochaetes are identified only to family and then “aggregated”
Canadian Centre for DNA Barcoding

- The mitochondrial gene cytochrome oxidase 1 (CO1) is used as a “DNA barcode”.

Why is the CO1 gene chosen as the barcode?
- CO1 gene is present in a broad range of species.
- The CO1 gene has a large variation between species yet a relatively small amount of variation within species.
- CO1 also has short very highly conserved primer regions enabling the barcode to be amplified and sequenced from virtually any animal.
Pilot Study: CO1 barcode sequencing results

Method:
1) Tissue dissected from 98 specimens (+various controls) sent to CCDB in special submission plates.

37 samples Ecoanalysts’ identified and sent back; 61 other specimens (worms, midges, mollusks) from the 27 sampling sites.
Pilot Study: CO1 barcode sequencing results

Summary of results:
1) Confirmed some Ecoanalysts’ taxa IDs (= >97% match to reference sequence database for same species):
   4 species listed here

2) Matches to reference sequences among other specimens:
   4 species listed here

3) Unmatched (<90% match) to genus or species level identified reference barcodes:
   4 taxa listed here (some only to genus level); also many oligochaetes

For more information about these taxa, contact jeffram@med.wayne.edu
Pilot Study: CO1 barcode sequencing results

Oligochaetes: A special case and the uniqueness of sequence data from Lake Erie specimens

Special: Cannot be identified by Ecoanalysts except with permanent mount [cannot retrieve tissue DNA after mounting]

Pilot study: Ecoanalysts ID’d 3 species (whole worms mounted and couldn’t be sequenced)

CCDB sequencing of pilot study specimens: at least 5 “identifiable taxonomic units”—NO matches to reference sequences EXCEPT for one non-native species listed here.

GLRI study:

Ecoanalysts cut worms in half, mounted head end, and preserved tail end with matching specimen ID #. We now have Ecoanalysts ID and reliable matched sequences from tails but STILL no match to reference sequences.
Oligochaetes (worms) classical taxonomy vs. CO1 sequence:

1. More “identifiable taxonomic units” (differing >10% in sequence) are present than the number of taxa identified by Ecoanalysts.

2. Same color = same species, as identified by Ecoanalysts.

3. No Color = Genbank previous COI identifications.

Neighbor joining tree of oligochaetes, rooted with leech outgroup, is shown here and briefly described on this slide. For more information, contact the authors at jeffram@med.wayne.edu.
Pilot study

• Taxa accumulation was graphed to see if more taxa were likely to be found with more intense collecting and identification effort.
• A still-rising curve of taxa accumulation indicates that additional taxa are likely to be present.
What % of species have been detected?
How many more samples would be needed to detect 90% of those present (at those sites, at those collecting times, with that gear and methods)?

The Chao asymptotic species richness estimator (Chao et al., 2009; Colwell, 2009) was used to make those estimates.

Input to EstimateS Excel sheet calculator:

<table>
<thead>
<tr>
<th></th>
<th>Q1 =</th>
<th>Q2 =</th>
<th>t =</th>
<th>S_{obs} =</th>
<th>S_{est} =</th>
<th>f_0 =</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td># of species in exactly one sample</td>
<td># of species in exactly two samples</td>
<td># of samples</td>
<td>number of species observed</td>
<td>estimated total number of species</td>
<td>estimated # species not detected</td>
</tr>
</tbody>
</table>
Pilot study

Ecoanalysts IDs

Chao Asymptotic Richness Estimator:  Total # of species:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$Q_1=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$Q_2=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$t=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$S_{\text{obs}}=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>$S_{\text{est}}=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$f_0=$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$Q_1 =$ # of species in exactly one sample  
$Q_2 =$ # of species in exactly two samples  
$t =$ # of samples  
$S_{\text{obs}} =$ number of species observed  
$S_{\text{est}} =$ estimated total number of species  
$f_0 =$ estimated # species not detected

Estimate how many more samples to achieve >90% detection:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>$g=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>$x^*=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>$m_g=tx^*=$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$g =$ proportion of species wanted to see  
($g$ must be greater than 0.809 )  
$t x^* =$ # of additional samples needed

For more information, contact the authors at jeffram@med.wayne.edu
Pilot study
Ecoanalysts + molecular IDs

Estimated total # of species, with additional molecular identifications:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Q₁</td>
<td># of species in exactly one sample</td>
</tr>
<tr>
<td>6</td>
<td>Q₂</td>
<td># of species in exactly two samples</td>
</tr>
<tr>
<td>7</td>
<td>t</td>
<td># of samples</td>
</tr>
<tr>
<td>8</td>
<td>Sₒbs</td>
<td>number of species observed</td>
</tr>
<tr>
<td>9</td>
<td>Sₑₚₗ</td>
<td>estimated total number of species</td>
</tr>
<tr>
<td>10</td>
<td>f₀</td>
<td>estimated # species not detected</td>
</tr>
</tbody>
</table>

Estimate how many more samples to achieve >90% detection:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>g</td>
<td>proportion of species wanted to see</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(g must be greater than 0.817)</td>
</tr>
<tr>
<td>33</td>
<td>x*</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>m₉ = tx*</td>
<td># of additional samples needed</td>
</tr>
</tbody>
</table>

For more information, contact the authors at jeffram@med.wayne.edu
1) Ecoanalysts and CCDB provide reliable, consistent data for identifying taxa of rare & non-native species

2) We are establishing a “Lake Erie reference DNA barcode database” of “identifiable taxonomic units” to which future specimens can be compared

3) In the pilot study, no new non-native species were found

4) In the immediate port area (lower Maumee River and inner Maumee Bay) a modest additional effort may achieve “over-sampling,” i.e. detection of >90% of all taxa present in the area.
Thanks to:

Organizations: Wayne State, EPA, Great Lakes Protection Fund, Healing Our Waters, and Lake Erie Waterkeepers,

Excellent collaborators, assistants, and students: Donna Kashian (Biology, WSU), Sarah Bailey (DFO, Canada), Phyllis Green (Isle Royale NP), Dick Gala (WSU emeritus), RamLab members (past and present): Masanori Fujimoto, Jordan Nechvatel, Jason Gizicki, Greg Moyerbrailean, Eboni Reed, Aos Karim, Fady Banno, Sanjay Rama, Joshua Southern, Ahsan Akram, Payel Acharya, and many others.

Excellent family support: Michal Ram, Natalie Ram, and Tammy Ram.
Any questions?
Contact jeffram@med.wayne.edu