NOAA Technical Memorandum GLERL-176
DOI: 10.25923/96kg-kj92

A Gap Analysis of the State of Knowledge of Great Lakes Nonindigenous Species

GLANSIS 2020

Rochelle Sturtevant[1], El Lower[1], Nick Boucher[1], Austin D. Bartos[1], Doran M. Mason[2], Ed Rutherford[2], Felix Martinez[3], and Ashley K. Elgin[2]

[1] Michigan Sea Grant
[2] NOAA Great Lakes Environmental Research Laboratory
[3] NOAA National Ocean Service National Centers for Coastal Ocean Service

NOAA Great Lakes Environmental Research Laboratory
4840 S. State Road, Ann Arbor, Michigan

Wednesday, May 25, 2022

UNITED STATES
DEPARTMENT OF COMMERCE

Gina Raimondo
Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

Dr. Richard W. Spinrad
Administrator

NOTICE

 This is GLERL Contribution No. 1995. This publication is available as a PDF file and can be downloaded from GLERL's web site: www.glerl.noaa.gov or by emailing GLERL Information Services at oar.pubs.glerl@noaa.gov.

Table of Contents

Table of Figures

Introduction

> *"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."*
> Donald Rumsfeld, former United States Secretary of Defense

The purpose of a gap analysis is to discover *unknown unknowns*, or more succinctly, identify gaps in a body of knowledge. We acknowledge this lack of understanding in our natural ecosystems and the response of our natural ecosystems to human-induced stressors, but few studies take a broad picture view of where these knowledge gaps occur. Identifying them can help direct research, inform management, and enlighten stakeholders and the public. Research on nonindigenous species is a field in which the identification of knowledge gaps is especially critical: while some nonindigenous species are known to be extremely harmful to ecosystems, we know very little about the vast majority and the potential harm (or benefits) they may inflict. In addition, there may be some taxa that are consistently undetected and/or underreported, resulting in unknown introductions and distributions.

The Great Lakes are a microcosm of global concerns and issues surrounding the invasion and establishment of nonindigenous species. As of 2019, more than 180 nonindigenous species have been reported to have reproducing populations in the Great Lakes basin, including lakes Superior, Michigan, Huron, St. Clair, Erie, Ontario, their connecting channels, and water bodies within their respective drainages. While some aquatic nonindigenous species (ANS) have had benign or beneficial impacts, others threaten the economy, the environment, or human health, and are thus referred to as "invasive." Protecting the Great Lakes ecosystem, regional economy, and community well-being from ANS requires effective prevention, early detection, and rapid response, management, and control efforts, which in turn depend upon the quantity and quality of accessible information. The Great Lakes Nonindigenous Species Information System (GLANSIS) is a NOAA-led, inter-agency, Great Lakes-specific database for aquatic nonindigenous species. GLANSIS serves as the Great Lakes' "one-stop shop" that centralizes and synthesizes the best available information to support effective management and control strategies that limit the introduction, spread, and impact of ANS in the Great Lakes.

This report examines gaps in the baseline information compiled by GLANSIS both to guide GLANSIS developers in future information searches and to inform the Great Lake research community about areas where additional research is needed. The Great Lakes Panel on Aquatic Nuisance Species has called for several types of gap analyses in their priorities for aquatic invasive species prevention and control including gaps in life history information needed for development of control strategies, and development of a prioritized list of established species or predicted imminent invaders where data on impacts is lacking or inconclusive which may be partially addressed by the gap analysis conducted for this report.

Methods

GLANSIS Listing Criteria

Nonindigenous List: This list includes only species that are present in the Great Lakes basin below the ordinary high-water mark – including connecting channels, wetlands and waters ordinarily attached to the Great Lakes. GLANSIS includes only aquatic species. USDA wetland indicator status is used as a guideline for determining whether wetland plants should be included in the list—obligate wetland, facultative wetland, and facultative plants are included in this list as aquatic; facultative upland and upland plants are not. Waterfowl, amphibians, reptiles, and mammals spending significant time in and dependent on the water are not currently included. To be included, a species must be reproducing and overwintering within the basin, as inferred from multiple discoveries of adult and juvenile life stages over at least two consecutive years. Species are excluded from the core list (but may be included on the watchlist) if their records of discoveries are based on only one or a few non-reproducing individuals whose occurrence may reflect merely transient species or unsuccessful invasions. The species included in this list are those that are considered nonindigenous within the Great Lakes basin according to the following definitions and criteria (based on Ricciardi 2006): (1) appeared suddenly and had not been recorded in the basin previously; (2) subsequently spreads within the basin; (3) its distribution in the basin is restricted compared to native species; (4) its global distribution is anomalously disjunct (i.e. contains widely scattered and isolated populations); (5) its global distribution is associated with human vectors of dispersal; and (6) the basin is isolated from regions possessing the most genetically and morphologically similar species. Cryptogenic species that cannot be verified as either native or introduced (after Carlton, 1996) are excluded. Range expander species that are native to part of the Great Lakes basin but which are expanding their range into other parts of the basin (e.g., species native to Ontario but non-native above Niagara Falls) are excluded.

Watchlist: This list includes only aquatic species (excluding facultative upland and upland plants, birds, mammals, reptiles and amphibians, as discussed above). These species are not already reproducing and/or overwintering in the Great Lakes, but have been assessed as 'likely' to successfully do so in the Great Lakes based on peer-reviewed scientific literature or other federal/state agency risk assessment literature. These species are thriving in a known donor region, such as rivers adjacent to Great Lakes, inland lakes in the Great Lakes region, western Europe, the Ponto-Caspian region, or in a zone with high specialization, species pool, or climate conditions that match the Great Lakes. Watchlist species must meet at least 3 of the 5 following criteria: (1) A transport vector currently exists that could move the species into the Great Lakes; (2) The species is likely to tolerate/survive transport (including in resting stages) in the identified vector; (3) The species has a probability of being introduced multiple times or in large numbers; (4) The species is likely to be able to successfully reproduce and overwinter in the Great Lakes (based on known tolerances or climate matching); (5) The species has been known to impact other areas or is assessed as likely to impact the Great Lakes system or have been officially listed as a potential invasive species of concern by federal, state or provincial authorities with jurisdiction in the Great Lakes basin.

Vector Definitions for Nonindigenous Species

Where available, for nonindigenous species GLANSIS elucidates the pathway of introduction from 1st introduction to the U.S., 1st Introduction to the Great Lakes basin, and spread throughout the basin to individual watersheds. This gap analysis focuses solely at the level of 1st introduction to the basin. For purpose of analysis, the refined description of the vector of introduction to the basin included in the database is here aggregated into broad categories as follows:

1. Stocked - these are authorized introductions to natural waters carried out or approved by government agencies (federal, state or provincial). Most stocked species are fish.
2. Planted/Escaped Cultivation - These include species that were deliberately planted in natural environments (e.g., for erosion control, site remediation, etc.) as well as species that were brought in for cultivation (e.g., agriculture, ornamental). These introductions were not 'authorized' in the same sense as stocked species, but were deliberate. Most species attributed to this vector are plants.
3. Aquaculture/aquarium release - This category includes species deliberately cultured in the Great Lakes region, which were either released (deliberately) or escaped (accidentally) to natural waters. Release/escape may have been either by culture operations or by individuals, in most cases that distinction cannot be consistently determined. This vector includes a mixture of plants, fish and invertebrates.
4. Hitchhiker - This category is composed of species that were carried in accidentally with another species or with recreational gear, etc. This category includes parasites of stocked/planted/cultured/escaped species as well as species that were introduced accidentally with approved species (e.g., misidentified bait species, weed seed contaminants, invertebrates attached to plants, etc..) as well as species that were accidentally introduced with non-living products deliberately brought into the region (gear, packing materials, etc.)
5. Shipping - This category includes organisms brought in with ballast water, solid ballast, residual ballast and (rarely) hull fouling.
6. Canal - This category includes species native to adjacent regions which entered the Great Lakes through canals shortly after construction as well as species that had previously been introduced to those adjacent waterways which were later able to take advantage of the opportunity to disperse across the basin boundary through such man-made corridors.
7. Unauthorized deliberate release and Accidental release -- For a handful of species a specific vector could not be identified sufficiently to fit the foregoing categories (or multiple possible vectors were identified), but were sufficiently identified to be categorized as to likely deliberate versus accidental. These species are not quite 'unknown' vectors, but neither do they rise to the level of certainty needed for inclusion in the foregoing categories. These are displayed separately for completeness, but due to low numbers not explicitly included in the comparison.
8. Unknown - Several species in the database are not assigned to a vector at all -- or are assigned to multiple possible vectors not obviously deliberate or accidental. These include several of the most recent invaders, but also include several species that have been here >50 years.

Within visualizations, the vector categories are always presented in this order left to right, which reflects the degree of 'intentional human facilitation' in each vector category for at least the first 5 categories.

For Watchlist species, most species are assigned multiple potential vectors in GLANSIS, here we assign species to only their most likely (highest scoring) potential vector. No watchlist species are in the 'stocked' category. All 'planted and escaped cultivation' vectors are included as unauthorized deliberate release (the more general term used in risk assessments). Three species assigned as aquarium/aquaculture escape and 1 hitchhiker are included in the 'Other category'. The shipping vector here consists solely of species assigned primarily to the specific ballast vector. The 'Dispersed' category includes those species native to or already established in adjacent basins (Mississippi, Ohio, Atlantic) which may disperse into the region via canals OR other manmade or natural connections between those watersheds and the Great Lakes. These categories are also ordered left to right in reflection of the degree of intentional human facilitation.

Literature

A thorough literature review is conducted for each listed species at least every 5 years. Data analyzed for this report reflects the GLANSIS holdings as of June 2019. Peer-reviewed publications (including risk assessments and ecosystem model forecasts) are sought first, followed by federal and state agency reports and other online fact sheets, books, professional communications with experts around the Great Lakes basin, and news stories, respectively. Bibliographies, profile information and maps are updated based on this review. All updates are reviewed by the Research Associate(s), GLANSIS Program Manager, and at least one additional member of the Executive Committee prior to posting. External review is solicited where particular additional expertise is needed. Reviewers are also solicited from the membership of the Great Lakes Panel on ANS. All map data, references, and information in shared fields (e.g., Identification, Nonindigenous Occurrences, Ecology) is also reviewed by USGS NAS Biologists. The Great Lakes Panel on ANS is apprised of all substantial updates and provided opportunity to comment. In order to understand the breadth of available literature and potential gaps in scientific knowledge, this report analyzes the literature held by GLANSIS in 3 distinct categories:

1. Bibliographies compiled for the nonindigenous list species and provided to the USGS NAS reference database (cross-linked from GLANSIS - direct at https://nas.er.usgs.gov/queries/references/default.aspx). Each bibliographic entry is tagged with the relevant species names up to a maximum of about 20 species per reference (to the capacity of the text field). Overview papers that include information for more than 20 species may not be tagged to all of the species mentioned in the paper. Non-standardized bibliographic entries – e.g., personal communications, news articles, press releases, etc. are often excluded from this database.
2. Profile-specific literature cited sections, which generally include only the most relevant subset of the entries from the bibliographic database, but may in some cases include particular sources (especially personal communications) that were excluded in the bibliographic database. These are available through the literature cited sections of the

individual species profiles (available at https://www.glerl.noaa.gov/glansis/nisListGen.php).

3. GLANSIS Organism Impact Assessments (nonindigenous list species) and Risk Assessments (watchlist species), which have directly relevant literature that are separately tracked, generally forming a subset of the full bibliography, but occasionally including sources that were excluded from that database. Each assessment seeks literature to specifically address a common set of questions (as outlined in NOAA TM-161 and 169). Literature cited within this category is NOT exhaustive in the same sense as the bibliography, but includes only the subset of the literature used to actually answer the particular question(s). The entire bibliography is reviewed on the regular 5-year cycle, but literature additions are made to this category ONLY if they alter the conclusion of the assessment. All assessments are available through NOAA GLERL Technical Memoranda (TM-161 and TM-169 et seq available at https://www.glerl.noaa.gov/pubs/#techRep).

Collection Records

GLANSIS tracks verified reports within the US Great Lakes basin for all species on the nonindigenous list. All data collected and curated by NOAA-GLANSIS staff are entered into the USGS NAS specimen database, quality-controlled by USGS, and merged with USGS data subject to identical controls. To the greatest extent possible, distribution data are point-mapped to specific latitude-longitude coordinates provided by the original report – in some cases GIS coordinates are assigned as centroids based on provided description. Each report is also georeferenced to a particular watershed hydrologic unit (USGS HUC8). Most data are verified by local agencies, universities or other experts; in some cases, NOAA-GLANSIS or USGS-NAS staff verify reports from submitted photos. USGS-NAS has data-sharing arrangements in place for direct sharing of verified reports from other databases collecting similar data in the Great Lakes region (such as MISIN, iMapInvasives, EDDMaps) – the quality control process includes procedures for removing duplicate entries. We avoid inclusion of multiple reports for a single site (e.g., hundreds of 'separate' reports from a recurring survey of a single wetland are normally condensed to a single report with potential updates to status at that location). We analyze this data in an attempt to understand geographic gaps in the knowledge base. All collection records and associated references are available through GLANSIS (link below the map in each profile or through map explorer at https://www.glerl.noaa.gov/glansis/mapExplorer.php - most species also available through USGS NAS at https://nas.er.usgs.gov/queries/default.aspx).

Risk and Impact Assessments

In 2014, NOAA researchers formulated and implemented a baseline assessment tool to quantify the realized, potential, and unknown impacts of established nonindigenous species in the Great Lakes (Sturtevant et al., 2014). This organism impact assessment (OIA) tool was developed and used to assess the magnitude of each species' impact in a standardized manner across all ANS established in the Great

Lakes basin by reviewing and synthesizing information from peer-reviewed research publications as well as gray literature.

A variation on the OIA forms Part C (Risk of Impact) of the NOAA-GLANSIS Risk Assessment (RA) and was expanded to be accompanied by Part A (Risk of Introduction) and Part B (Risk of Establishment). The RA was developed as part of a larger project funded in part by the US Environmental Protection Agency's Great Lakes Restoration Initiative that assessed the potential relative risk of species predicted to invade the Great Lakes (Fusaro et al., 2016, Davidson et al., 2017).

Both the NOAA-GLANSIS OIA (nonindigenous list species) and NOAA-GLANSIS RA Part C (watchlist species) considered two types of negative ANS impacts: environmental and socio-economic, along with one category for beneficial impacts. We defined *environmental impacts* as the effects of nonindigenous species on the biotic and/or abiotic components of the ecosystem relative to pre-invasion conditions. *Socio-economic impacts* include those that directly affect individual or societal values relative to pre-invasion conditions.

Environmental impacts were divided into six categories, each posed as a specific question as follows:

1. *Environmental Health*: Does the species pose some hazard or threat to the health of native species?
2. *Competition*: Does it out-compete native species for resources?
3. *Predator-Prey*: Does it alter predator-prey relationships?
4. *Genetics*: Has it affected any native populations genetically?
5. *Environmental Water Quality*: Does it negatively affect environmental water quality?
6. *Physical Ecosystem*: Does it alter the physical ecosystem in some way?

Socioeconomic impacts were also divided into six different categories, defined as the following:

1. *Human Health*: Does this species pose some hazard or threat to human health?
2. *Infrastructure*: Does it cause damage to infrastructure?
3. *Water Quality for Human Use*: Does it negatively affect water quality for human use?
4. *Economy*: Does it harm any markets or economic sectors?
5. *Recreation*: Does it inhibit recreational activities and/or associated tourism?
6. *Aesthetics*: Does it diminish the perceived aesthetic or natural value of the areas in inhabits?

Beneficial impacts were also divided into six categories, defined as the following:

1. *Biocontrol*: Does this species act as a biological control for other harmful nonindigenous organisms?
2. *Commercial Value:* Is it commercially valuable (fisheries, aquaculture, bait, agriculture, ornamental trade)?
3. *Recreational Value:* Is it valuable for recreation (sport fisheries, pet, etc.)?
4. *Medicinal Value:* Does it have medicinal or research value (other than research on its control)?

5. *Bioremediation:* Does it remove toxins or pollutants from the water or otherwise increase water quality?
6. *Ecological Value:* Does it have positive ecological impacts outside of biocontrol (e.g., increases growth or reproduction of native or desirable species, support survival of a threatened or endangered species, or fill a gap in the food web)?

In addition to potential impact, the NOAA-GLANSIS RA evaluated the potential for introduction and establishment (Fusaro et al., 2016; Davidson et al., 2017). These assessment tools were modeled after the United Kingdom Non-Native Organism Risk Assessment scheme (Baker et al., 2007) with adaptations for freshwater species. This model was selected because it was one of the most comprehensive and straightforward models that could be applied broadly across taxa as well as being particularly transparent in that every individual score is justified with a written comment.

The Potential for Introduction component of the risk assessment (Part A) sought to identify all possible vectors of introduction and a species' proximity (or ease of introduction) to pathways into the Great Lakes. This assessment considers paired questions for each of six vectors.

1. *Dispersal* – (a) Does this species occur near waters (natural or artificial) connected to the Great Lakes basin? (b) What is the proximity of this species to the Great Lakes basin?
2. *Hitchhiking/Fouling* - (a) Is this species likely to attach to or be otherwise transported by, or along with, recreational gear, boats, trailers, fauna (e.g., waterfowl, fish, insects), flora (e.g., aquatic plants), or other objects (e.g., packing materials), including as parasites or pathogens, entering the Great Lakes basin? (b) What is the proximity of this species to the Great Lakes basin?
3. *Unauthorized intentional release* – (a) Is this species sold at aquarium/pet/garden stores ("brick & mortar" or online), catalogs, biological supply companies, or live markets (e.g., purchased for human consumption, bait, ornamental, ethical, educational, or cultural reasons) and as a result may be released into the Great Lakes basin? (b) How easily is this species obtained within the Great Lakes region (states/provinces)?
4. *Stocking/Planting/Escape from Recreational Culture* – (a) Is this species being stocked/planted to natural waters or outdoor water gardens around the Great Lakes region? (b) What is the nature and proximity of this activity to the Great Lakes basin?
5. *Escape from Commercial Culture* – (a) Is this species known to be commercially cultured in or transported through the Great Lakes region? (b) What is the nature and proximity of this activity to the Great Lakes basin?
6. *Shipping* – (a) Is this species likely to be taken up in ballast, and capable of surviving adverse environments (i.e. extreme temperatures, absence of light, low oxygen levels) and partial-to complete ballast water exchange/flushing (e.g., is euryhaline, buries in sediment, produces resistant resting stages, has other attributes or behaviors facilitating survival under these conditions)? (b) Does this species occur in waters from which shipping traffic to the Great Lakes originates?

The Potential for Establishment component of the risk assessment (Part B) consisted of 18 questions evaluating characteristics of potential invaders, such as biological and ecological attributes, environmental compatibility, propagule pressure, and history of invasion and spread.

- Ecological Attributes

    1. *How would the physiological tolerance of this species (survival in varying temperature, salinity, oxygen, and nutrient levels) be described?*
    2. *How likely is it that any life stage of this species can overwinter in the Great Lakes (survive extremely low levels of oxygen, light, and temperature)?*
    3. *If this species is a heterotroph, how would the flexibility of its diet be described?*
    4. *How likely is this species to outcompete species in the Great Lakes for available resources?*
    5. *How would the fecundity of this species be described relative to other species in the same taxonomic class?*
    6. *How likely are this species' reproductive strategy and habits to aid establishment in new environments, particularly the Great Lakes (e.g., parthenogenesis/self-crossing, self-fertilization, vegetative fragmentation)?*

- Environmental Compatibility

    1. *How similar are the climatic conditions (e.g., air temperature, precipitation, seasonality) in the native and introduced ranges of this species to those in the Great Lakes region?*
    2. *How similar are other abiotic factors that are relevant to the establishment success of this species (e.g., pollution, water temperature, salinity, pH, nutrient levels, currents) in the native and introduced ranges to those in the Great Lakes?*
    3. *How abundant are habitats suitable for the survival, development, and reproduction of this species in the Great Lakes area (e.g., those with adequate depth, substrate, light, temperature, oxygen)?*
    4. *How likely is this species to adapt to or to benefit from the predicted effects of climate change on the Great Lakes freshwater ecosystems (e.g., warmer water temperatures, shorter duration of ice cover, altered streamflow patterns, increased salinization)?*
    5. *How likely is this species to find an appropriate food source (prey or vegetation in the case of predators and herbivores, or sufficient light or nutrients in the case of autotrophs)?*
    6. *Does this species require another species for critical stages in its life cycle such as growth (e.g., root symbionts), reproduction (e.g., pollinators, egg incubators), spread (e.g., seed dispersers), or transmission (e.g., vectors)?*
    7. *How likely is the establishment of this species to be aided by the establishment and spread of another species already in the Great Lakes?*
    8. *How likely is establishment of this species to be prevented by the herbivory, predation, or parasitism of a natural enemy that is already present in the Great Lakes and may preferentially target this species?*

- Propagule Pressure

    1. *On average, how large and frequent are inoculations (introduction events) from the potential vectors identified in Section A for this species? (What is the total number of individuals introduced?)*

- History of Invasion and Spread

    2. *How extensively has this species established reproducing populations in areas outside its native range as a direct or indirect result of human activities?*
    3. *How rapidly has this species spread by natural means or by human activities once introduced to other locations?*
    4. *Are there any existing control measures in the Great Lakes set to prevent the establishment and/or spread of this species?*

Once this format was developed, trained members of the GLANSIS team conducted individual assessments to a standard template and framework based on comprehensive literature review. All assessments were reviewed by senior members of the team according to specialty as well as external experts where in-house expertise was lacking, and all assessments were reviewed by the program manager for consistency. Updates to the OIAs and RAs are on the same 5-year review cycle as the literature reviews and data used for this assessment and represent the state of the GLANSIS holdings as of October 2019.

Each question in these assessments allows for the possibility of an 'unknown' response, which is reflected in the certainty assigned to the overall assessment. In addition to examining the raw numbers of papers contributing to the knowledge base for these assessments, we also examine the pattern of unknowns in this gap analysis. This analysis sheds light on specific gaps in understanding of particular components of the knowledge base.

Profile Data Examination

GLANSIS profiles are created to a template with specific consistent data types gleaned from the literature used within the profiles. Here we examine specific subsets of this data, categorizing the data as high quality (specific or quantitative data), moderate quality (general and/or qualitative statements) and not available. For example, a habitat description such as "shoals, ponds, shallow lakes, canals; found on the substrate in fall and winter (including gravel, sand, clay, mud or undersides of rocks) and on aquatic macrophytes in summer" would be considered a high-quality habitat description, while "littoral" would be considered a general description. Similarly, "salinity intolerant with mortality at 3ppt" would be considered a quantitative description of salinity tolerance while "salt intolerant" would be considered qualitative. Full data used is available (Appendix 1); data presented here is summarized in pie-chart format for major taxonomic groups with each chart depicting the fraction of species within the group for which the data is high quality, moderate quality or unavailable. A similar approach is taken to examination of the regulatory (regulated/unregulated) and control (in use, in research, not available) information summarized to broad categories (biological, physical and chemical controls) in the profiles. All data is available in the individual GLANSIS species profiles (accessible through https://www.glerl.noaa.gov/glansis/nisListGen.php).

Risk Assessment Clearinghouse

The GLANSIS Risk Assessment Clearinghouse (Clearinghouse) was developed in partnership with the Great Lakes Panel on Aquatic Nuisance Species Risk Assessment Ad Hoc Committee and the Great Lakes Commission. This 3-tiered database provides access to invasive-species risk assessment literature regardless of whether the species meets the specific criteria for listing on the GLANSIS Nonindigenous or Watch lists. Tier 1 provides access to the entire body of bibliographic information that is included in the NAS database and tagged as a risk assessment (116 publications). Based on feedback from the Great Lakes Panel, 14 assessments (NOAA-GLANSIS Organism Impact Assessments; NOAA-GLANSIS Watchlist Risk Assessments; USFWS Ecological Risk Screening Summaries; University of Notre Dame STAIR Assessments for Fish, Crayfish and Mollusks; U.S. Aquatic Weed Risk Assessment; Canadian Aquatic Weed Risk Assessment; Great Lakes Aquatic Weed Risk Assessment; Wisconsin Department of Natural Resources Risk Assessments; and the New York Invasive Species Information Assessments) in 'common use' by managers in the Great Lakes region were selected for inclusion in Tier 2, which includes a more detailed summary of the assessment methods in a format allowing side-by-side comparison of the different methods (Tier 2 including links to original sources available at https://www.glerl.noaa.gov/glansis/raExplorer.html). Each of these 14 assessments have been used to conduct species-specific risk assessments and the summaries of the actual risk assessments for all 14 is now included in Tier 3, allowing side-by-side comparison of the actual assessment results for particular species. Assessments are included regardless of whether the final assessment demonstrates the species to pose a risk to the Great Lakes region, allowing users to see the rationale for particular species not being considered a risk to the Great Lakes if they have been examined by any of these agencies or methods. The Tier 3 Clearinghouse currently contains risk assessment summaries for 2334 species, many of which were assessed by multiple methods (available at https://www.glerl.noaa.gov/glansis/raT2Explorer.html). Here we analyze only the overall composition of the Clearinghouse with comparisons to the Nonindigenous List and Watchlist to provide insight on gaps in the GLANSIS methods that are currently filled by other assessments as well as to understand potential gaps in the overall risk assessment literature.

Data Visualization and Analysis

Quartile analysis of bibliographic and collection record data (taxa and vector)

Quartile analysis provides a concise 5 number summary of a dataset: sample minimum, lower quartile (lowest 25% of data), median, third quartile, and maximum. Quartiles are less susceptible to long-tailed distributions and outliers, which we considered likely in this dataset. We represent our quartile analysis using box and whisker plots in which the red dot indicates the median, the box indicates half the data nearest the median (quartile above and below) and the whiskers indicate the range. Quartile analysis is appropriate for comparing distribution of ordinal information (such as number of publications or number of records for each species) across discrete categories (such as taxa or vector). For bibliographic entries and collection records, we have placed the Y-axis on a logarithmic scale, reflecting the logarithmic distribution of bibliographic entries and collection records across species within each taxa. Impact-relevant studies are normally-distributed, so not presented on a logarithmic scale.

Scatterplot analysis of bibliographic and collection record data (size)

Quartile analysis is inappropriate for comparisons across a continuous variable (such as size or time). Here we employ a scatterplot to look for trends or gaps in bibliographic data as relates to the maximum size of each species. For bibliographic entries and collection records, we have placed the Y-axis on a logarithmic scale, reflecting the logarithmic distribution of bibliographic entries across species (many species with few publications and a handful of species with many publications) and have also placed size on a logarithmic scale reflecting the logarithmic distribution of nonindigenous species across the size spectra (many small species, relatively fewer large species). This minimizes the influence of outliers. Because organism size and taxa are not independent, we overlay taxonomic information on our analysis of size.

GIS analysis

Spatial data for each species and their respective collection records were retrieved from GLANSIS holdings. Resolution was to hydrologic unit code 8 (HUC8) watersheds as defined by the United States Geological Survey Watershed Boundary Dataset. The HUC8 watersheds of Lake Champlain were excluded from this analysis because their collection records are not subject to the same level of quality control. Species and collection records were mapped in ArcGIS Pro (Version 2.5.2) software and summary statistics were calculated in RStudio (Version 1.3.1056) for each HUC8 watershed within the Great Lakes basin.

Pie chart visualization of profile categorical data

Profile data is summarized only into qualitative categories, with each species assigned to a particular category reflective of the quality of data available for that category. Pie charts are used to visualize the breakdown of data availability for each taxa.

Results and Discussion

Gaps in the Scientific Literature for the Nonindigenous Species List

The full bibliographies available in GLANSIS for 188 nonindigenous species average 63 references per species (total not reported because some references were used for multiple species). Numbers for each species vary widely; several microorganisms have just a single reference listed, while *Dreissena polymorpha* has 1536 references. More than 75% of the nonindigenous species have fewer than 100 references (Figure 1). Patterns across taxonomic groups are unclear, but a distinct ‘size bias’ is apparent, in which more than half of the taxa in microscopic species (<1 mm) have fewer than 10 references (Figure 2). A trend is apparent in the medians of the first 6 vector categories in which stocked species - the vector with the greatest human intentional facilitation of introduction have a greater number of bibliographic references (median 121 references) than successively less intentional introductions (canal median = 12 references) (Figure 3).

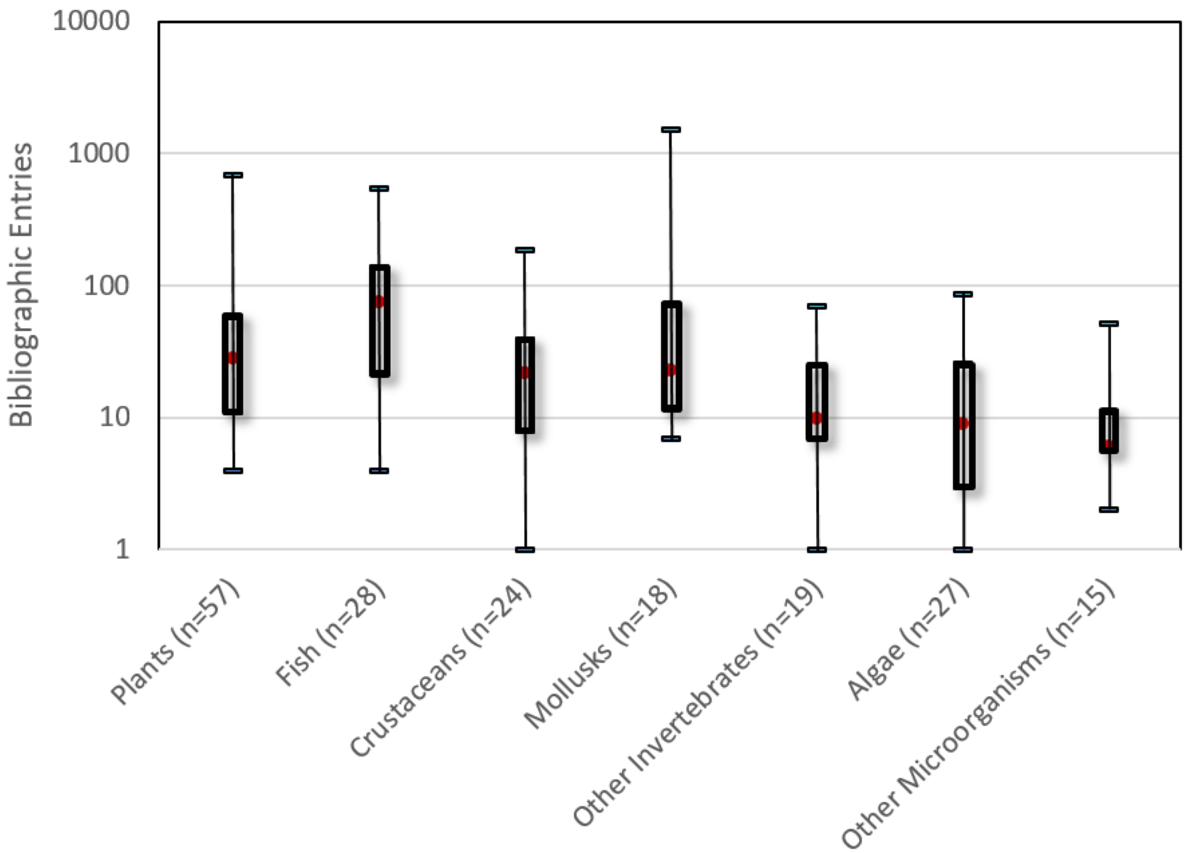


Figure 1. Quartile Analysis of the Total Bibliographic Holdings for Nonindigenous Species in the Great Lakes by Taxa. Number of species within each taxa given in parentheses in the x-axis. Bibliographic entries are logarithmically distributed.

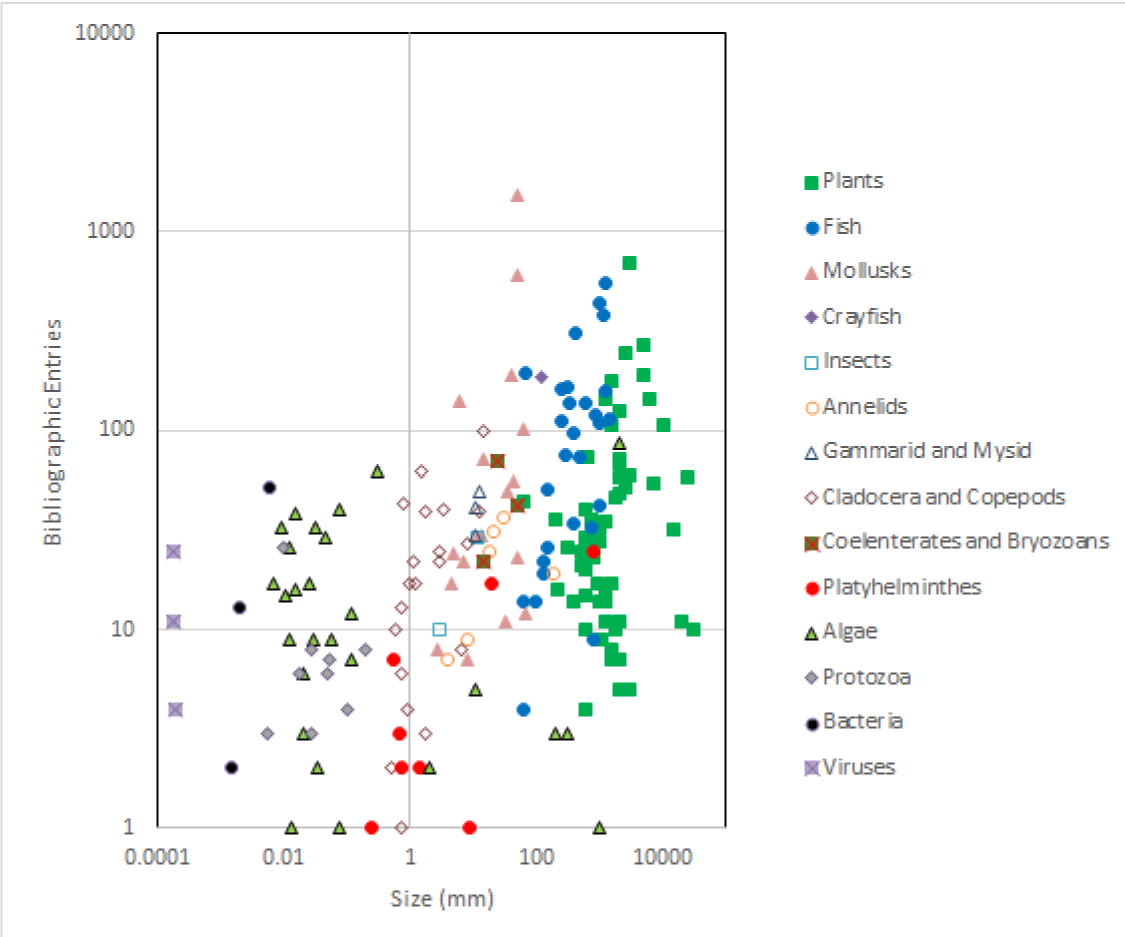


Figure 2. Total Bibliographic Holdings for Nonindigenous Species in the Great Lakes by Size. Bibliographic entries and size are logarithmically distributed.

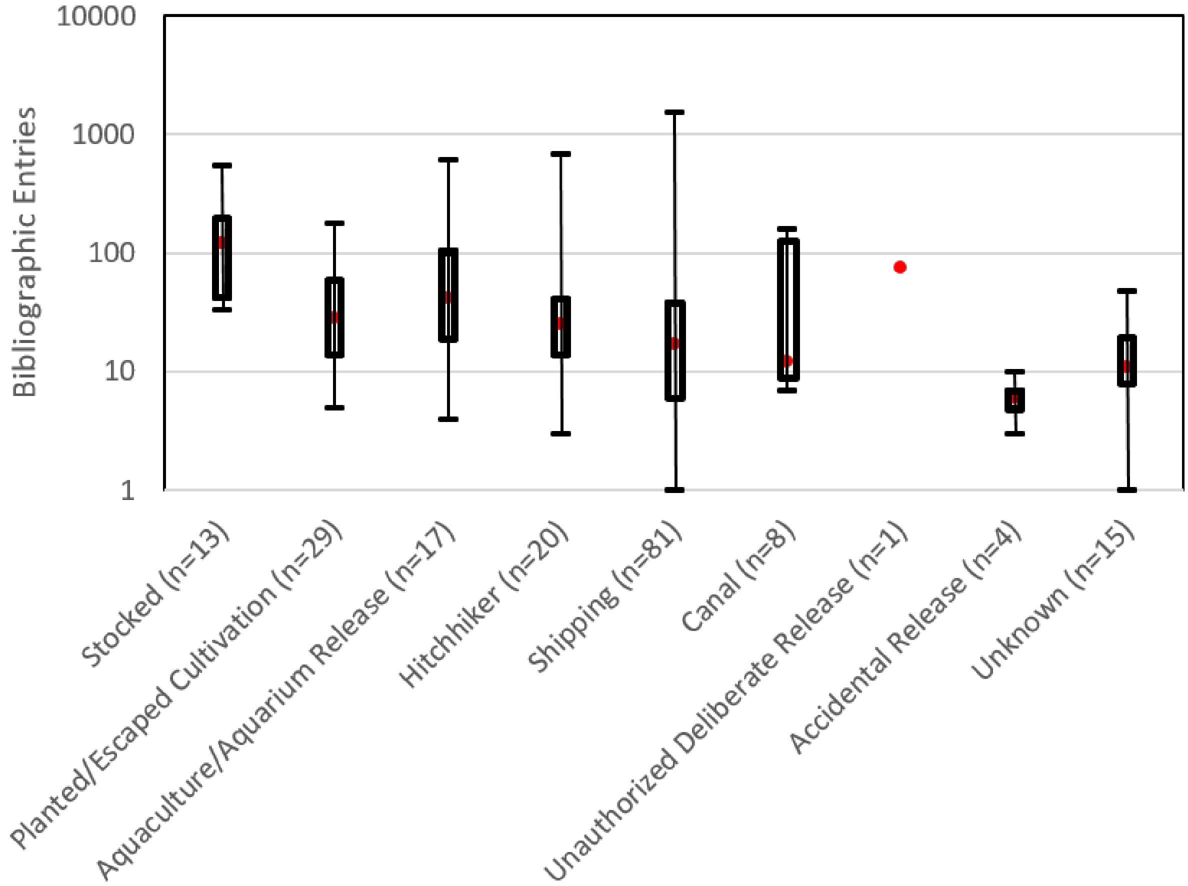


Figure 3. Quartile Analysis of the Total Bibliographic Holdings for Nonindigenous Species in the Great Lakes by Vector. Number of species within each vector given in parentheses in the x-axis. Bibliographic entries are logarithmically distributed.

Gaps in Collection Records for the Nonindigenous Species List

GLANSIS maintains verified georeferenced collection records (reports) for all nonindigenous species. The system currently includes more than 90,000 total reports. This is an average of 471 reports per species, but the numbers of reports range from a single verified report for some species to more than 13,000 reports for *Salmo trutta* (brown trout). 68% of species have fewer than 100 reports. Plants and fish tend to have more reports than do other taxa (Figure 4). Again, a distinct bias against reports for microscopic species is apparent, with more than half of the species under 1 mm in maximum length having fewer than 10 verified reports (Figures 5). The same apparent trend across vector categories as seen in the bibliographic data (in which the database contains more records for categories most strongly associated with human intentional facilitation of introduction) (Figure 6).

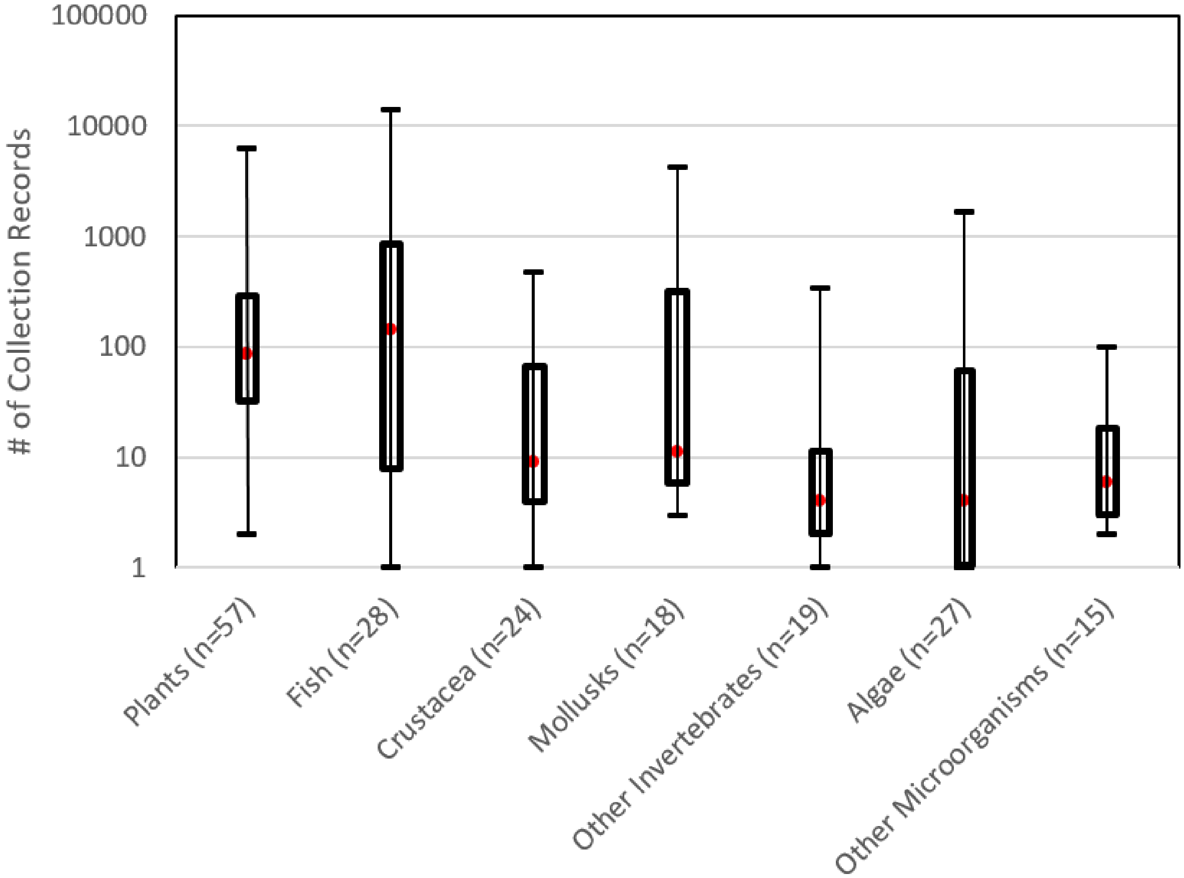


Figure 4. Quartile Analysis of the Collection Record Data for Nonindigenous Species by Taxa. Number of species within each taxa given in parentheses in the x-axis. Collection records are logarithmically distributed.

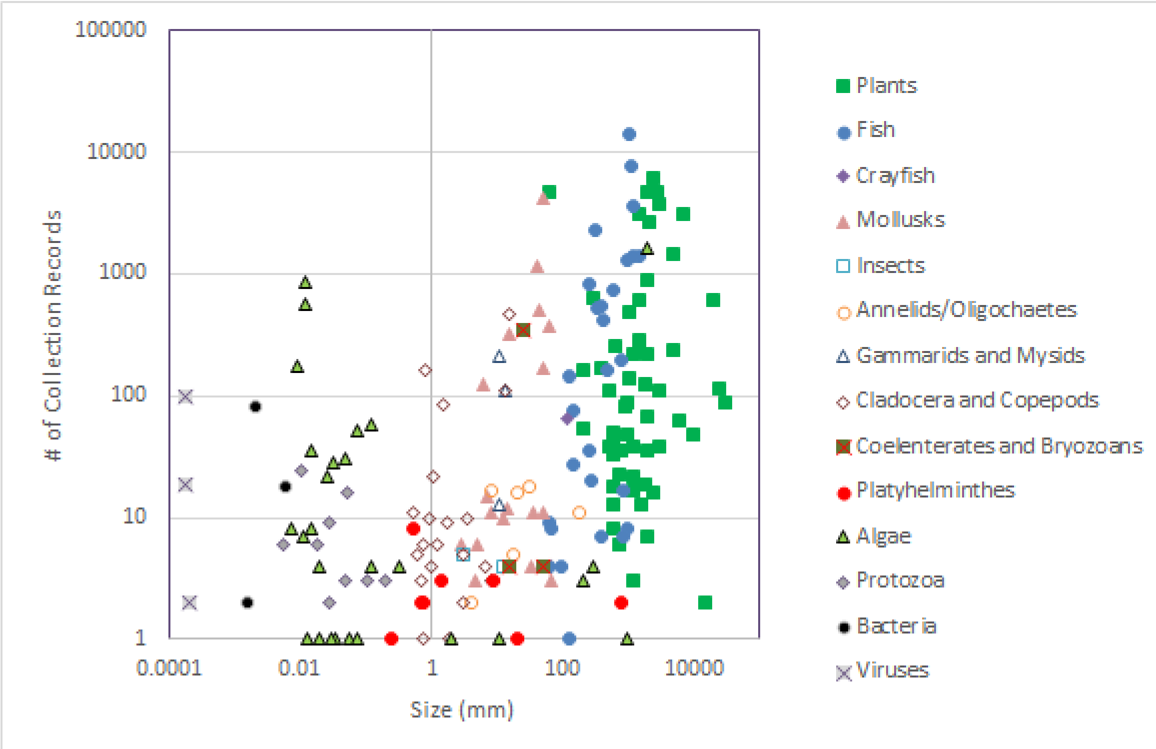


Figure 5. Nonindigenous Species Collection Records in the Great Lakes by Size. Collection records and size are logarithmically distributed.

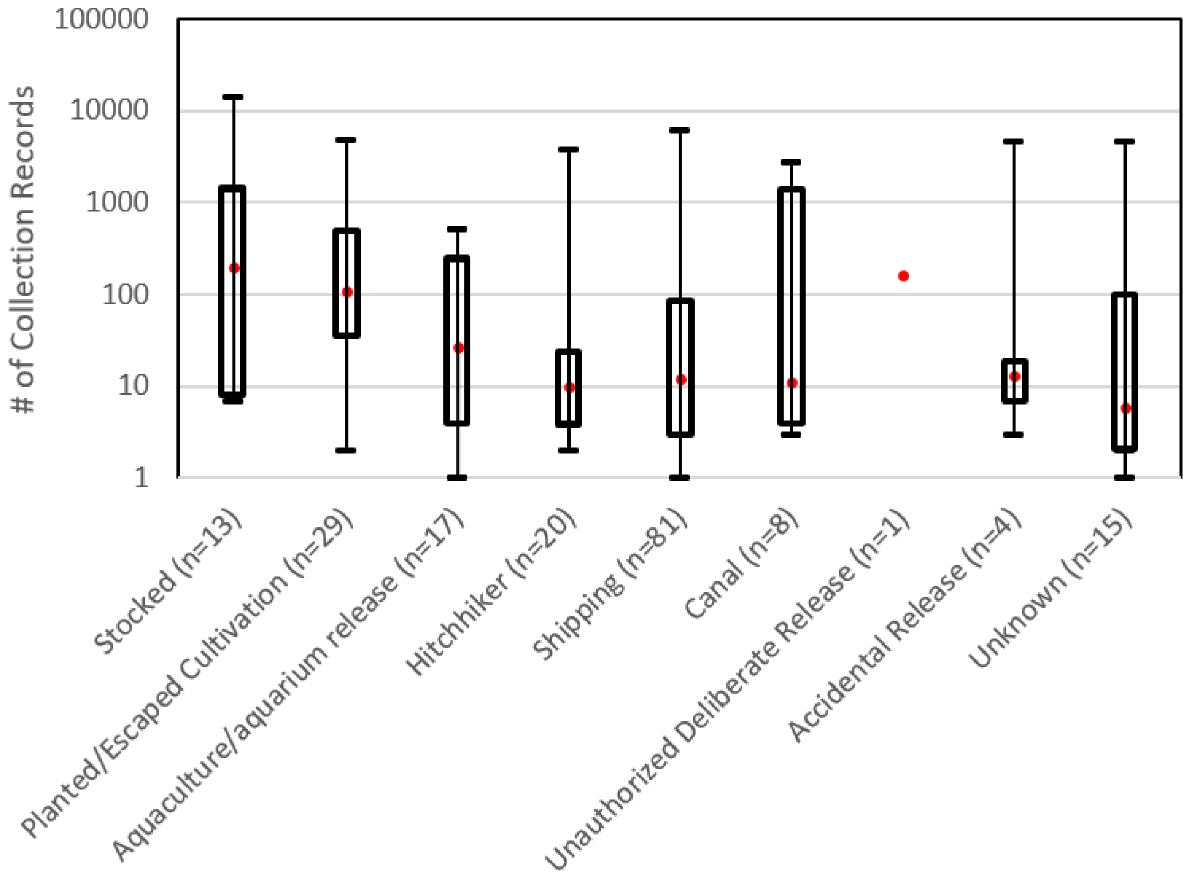


Figure 6. Quartile Analysis of the Collection Record Data for Nonindigenous Species by Vector. Collection records are logarithmically distributed.

Because report data is georeferenced, we also examined this dataset spatially to look for possible geographic gaps in species distributions. The 124 HUC8 watersheds that make up the US Great Lakes basin vary substantially in the number of species (ranging between 5 for Birch-Willow and 130 for Lake Erie) which have been reported (Figure 7). Each watershed within the Great Lakes basin has its own unique assemblage of nonindigenous species; no species has been reported for all watersheds, and no watershed has zero non-native species. Seventeen species are reported as being found in a single watershed, while *Lythrum salicaria* (purple loosestrife), reported in 107 watersheds, has the most wide-spread distribution.

Unfortunately, absence data is difficult to come by because papers tend not to be published on negative data and absences are not consistently reported. As a result, absence of evidence does not necessarily equal evidence of absence. Due to limitations in data availability, GLANSIS does not track absence data or contain consistent data on numbers relative to collection effort, which means that we do not have a reliable way to separate underreporting (GLANSIS effort at collating data) from actual undersampling (by field scientists) versus actual differences in number of species present. As a result, we are unable to determine with confidence if there are watersheds that may actually be resistant to invasion due to habitat or other factors.

In order to handle differences in watershed size, collection record numbers were corrected by watershed area (records per square kilometer) as shown in Figure 8.

Well-studied ‘hotspot’ watersheds in Duluth, Traverse City, the St. Clair-Detroit River System, and the Finger Lakes are immediately apparent in the map of reports per unit area. This is an expected direct reflection of known sampling patterns. Even though there is a high number of nonindigenous species established in Lake Superior, the number of records per square kilometer behind that number is very low. This rate reflects known patterns of research effort on Lake Superior versus the lower lakes. In this way, GLANSIS reports reflect at least some known primary differences in sampling effort.

The high number of records per species per area at some ‘sentinel’ sites suggests that there are areas with high reporting frequency that are accumulating more reports of the same species rather than finding additional species. The average number of records per species varies from 1.9 in the Waiska watershed to 123 in the Wolf watershed. Correcting these for watershed area (Figure 9) reflects high numbers of reports per species in the Beaver-Lester watershed (including Duluth-Superior), the Betsie-Platte watershed south of Grand Traverse, and the Ottawa-Stony and Detroit River watersheds feeding into western Lake Erie, all of which were anticipated as heavily studied regions. Figure 9 also suggests there may be geographic gaps in GLANSIS records, with more reports per species per area in Minnesota and Wisconsin waters.

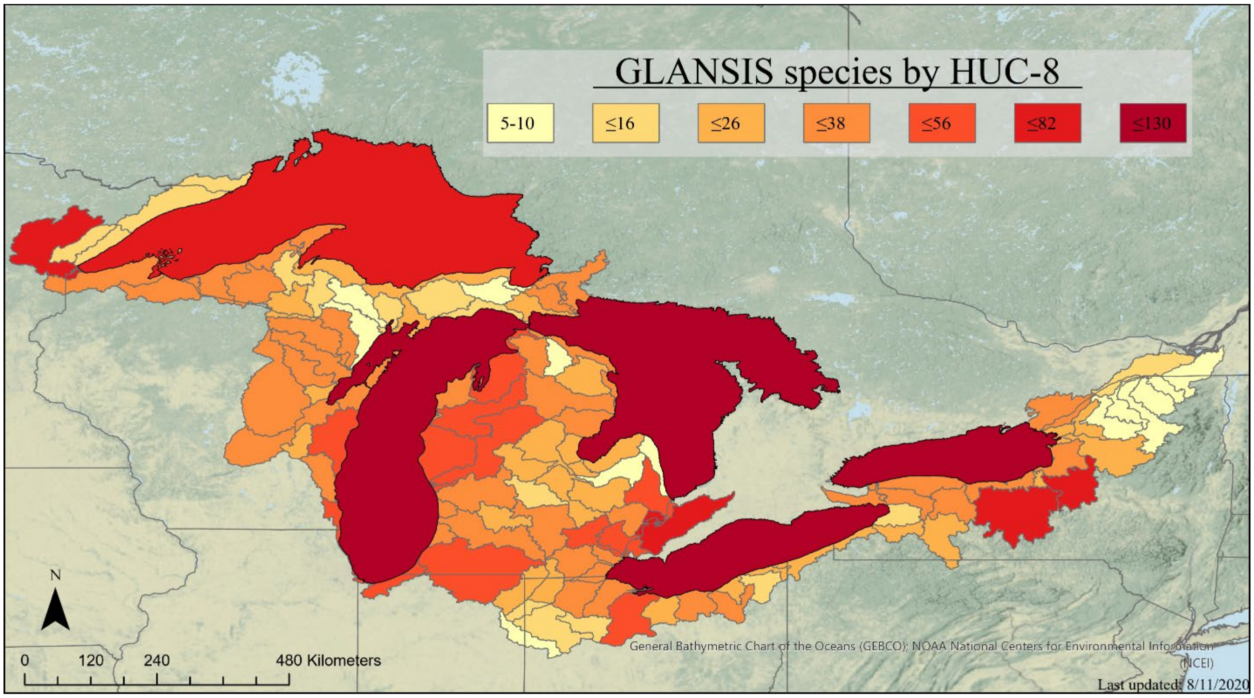


Figure 7. GLANSIS Species per Hydrologic Unit (Watershed) as of August 2020. Each subwatershed in the Great Lakes is color-coded to represent the number of nonindigenous species present in that subwatershed. Categories are optimized with Jenks classification to identify natural breaks in the data.

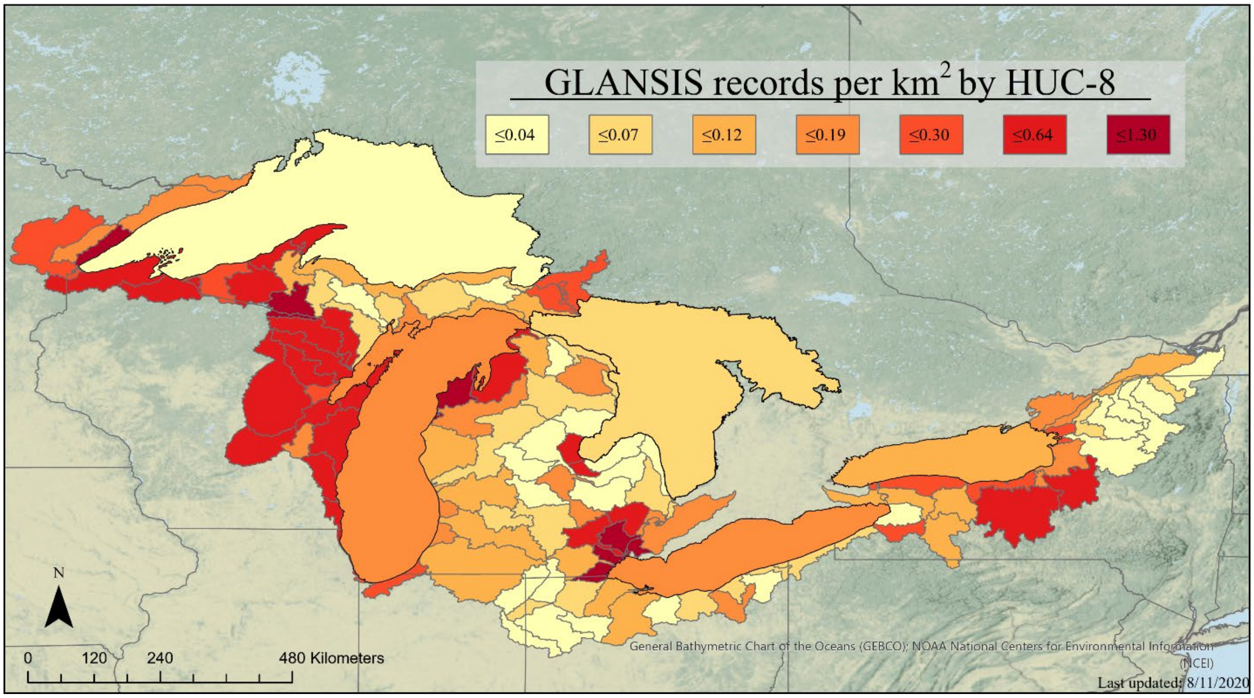


Figure 8. Records per Square Kilometer by Hydrologic Unit (Watershed).

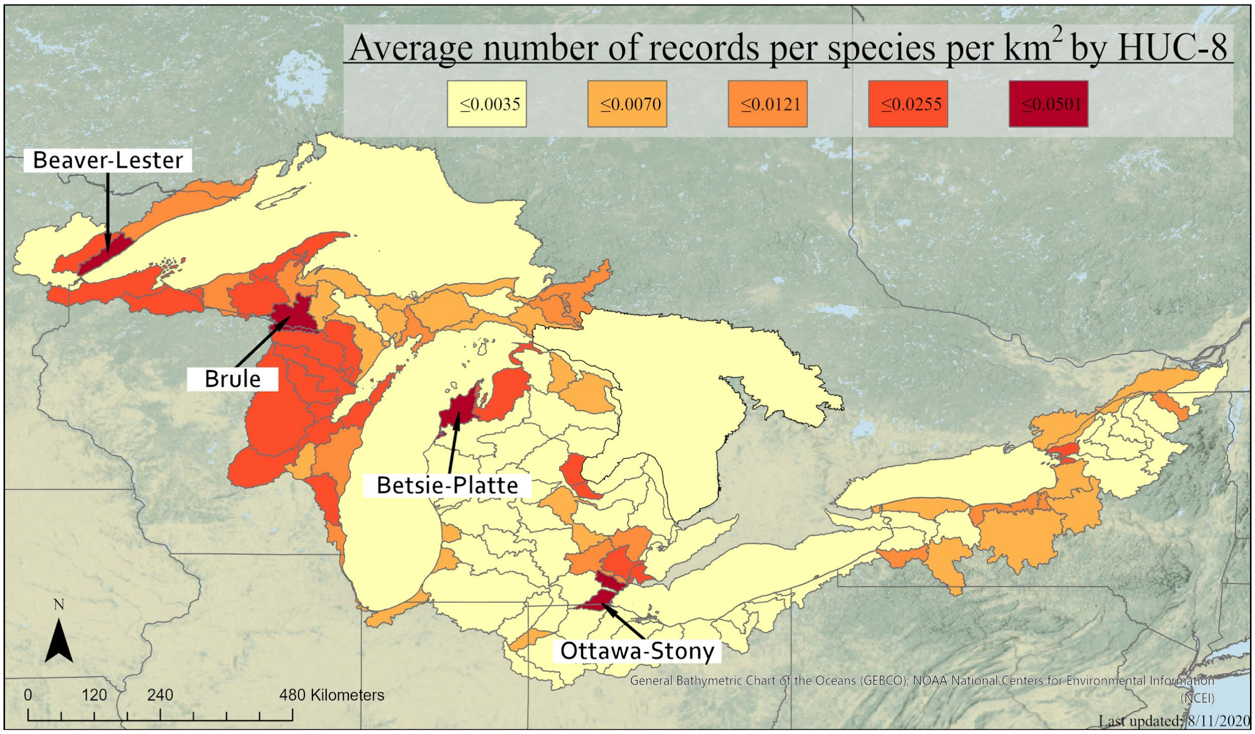
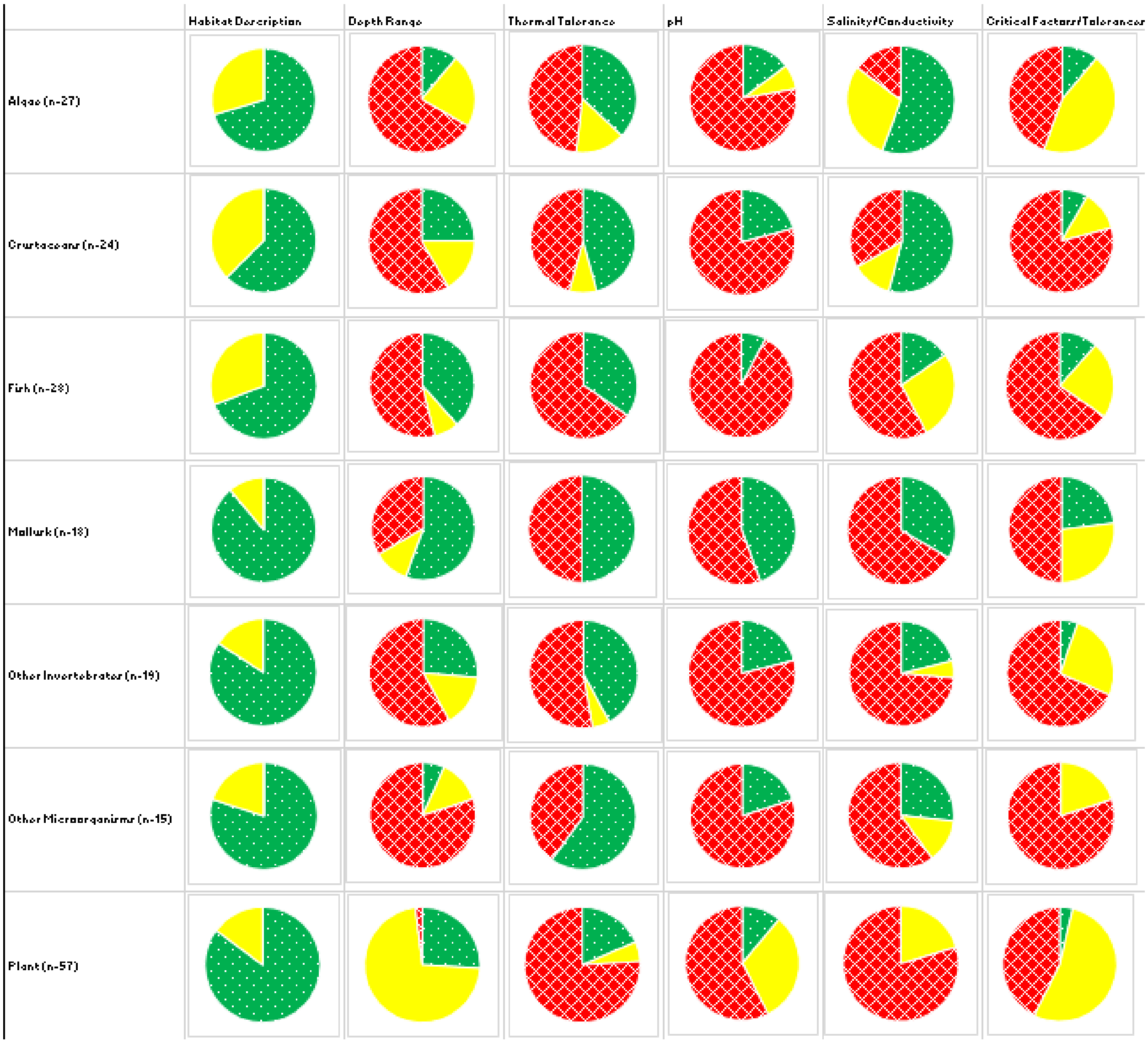


Figure 9. Map of average number of records per species (corrected for watershed area = per square kilometer) by HUC. The watersheds with the highest average number of records per species per area are labeled.

Gaps in Habitat Data

Based on examination articles from the above bibliographies, at least a general description of habitat requirements is available for all Great Lakes nonindigenous species, but our effort to consistently collect more specific information for just 5 categories -- depth range, thermal tolerance, pH, salinity and 'critical factors or tolerances' reveals a high variability in the quality of available information (Table 1) and that for many species critical data is lacking. Just 49 species have references that provide information on numerically specific depth ranges, only 67 have information on thermal tolerances/preferences, and just 31 have information on pH range. Numerous species are noted for having 'salinity tolerance' yet few studies of actual salinity ranges are available to back up those vague statements. A handful of species have quantitative information on other critical factors (e.g., light, flow, nutrients, oxygen) but these vary widely by species rather than being consistent even within taxa (e.g., some might have information on phosphorus, others on iron, yet others on oxygen, etc..)

Table 1. Habitat descriptive data summary. Each species was scored on the quality of information in each category -- clockwise within each chart: green dotted = highly detailed descriptions and/or quantitative statements, yellow solid = general descriptions and/or qualitative statements, red hashed = no data found in literature review.



Understudied and Underrecorded Nonindigenous Species

There are 48 species that each have fewer than 10 references (Figure 10) – for purposes of the gap analysis, we assume these are likely 'understudied' (at least from a database perspective, we would like to have more information on these species). There are 72 species that have fewer than 10 records – these may be truly rare or may be 'underreported'. Looking at these two quasi-independent data types together we see 37 species that have both < 10 references and <10 records. This list of 'understudied and underrecorded' species includes 2 fairly large plants – *Sparganium glomeratum* and *Pluchea odorata* are often considered naturalized. This category also includes 2 fish – *Notropis buchanani* and

Lepisosteus platostomus are both species that experts have fluctuated on classifying as native or not over time. The one mollusk is small peaclam – *Pisidium moitessieranum* -- that is rarely identified to species (GLANSIS excludes records identified only to genus level). The list also includes 3 relatively large (>20cm) nearshore salt-tolerant algae with limited distributions (*Ulva flexuosa, Ulva intestinalis*, and *Ulva prolifera*). The remaining organisms are all microscopic, and many are parasitic (Annelid = *Gianius aquaedulcis*; Cladocera = *Eubosmina maritima*; Copepods = *Cyclops strenuus, Nitokra incerta, Salmincola lotae*; Platyhelminthes = *Dactylogyrus amphibothrium, Dactylogyrus hemiamphibothrium, Ichthyocotylurus pileatus, Neascus brevicaudatus, Scolex pleuronectis, Timoniella sp.;* Algae = *Chaetoceros muelleri, Chroodactylon ornatum, Conticribra guillardi, Discostella woltereckii, Hymenomonas roseola, Pleurosira laevis, Sphacelaria fluviatilis, Sphacelaria lacustris, Thalassiosira bramaputrae*; Protozoans = *Acineta nitocrae, Glugea hertwigi, Heterosporis sutherlandae, Psammonobiotus dziwnowi, Psammonobiotus linearis, Sphaeromyxa sevastopoli, Trypanosoma acerinae*; Bacteria = *Piscirikettsia cf salmonis*; and Virus = *Rhabdovirus carpio.*

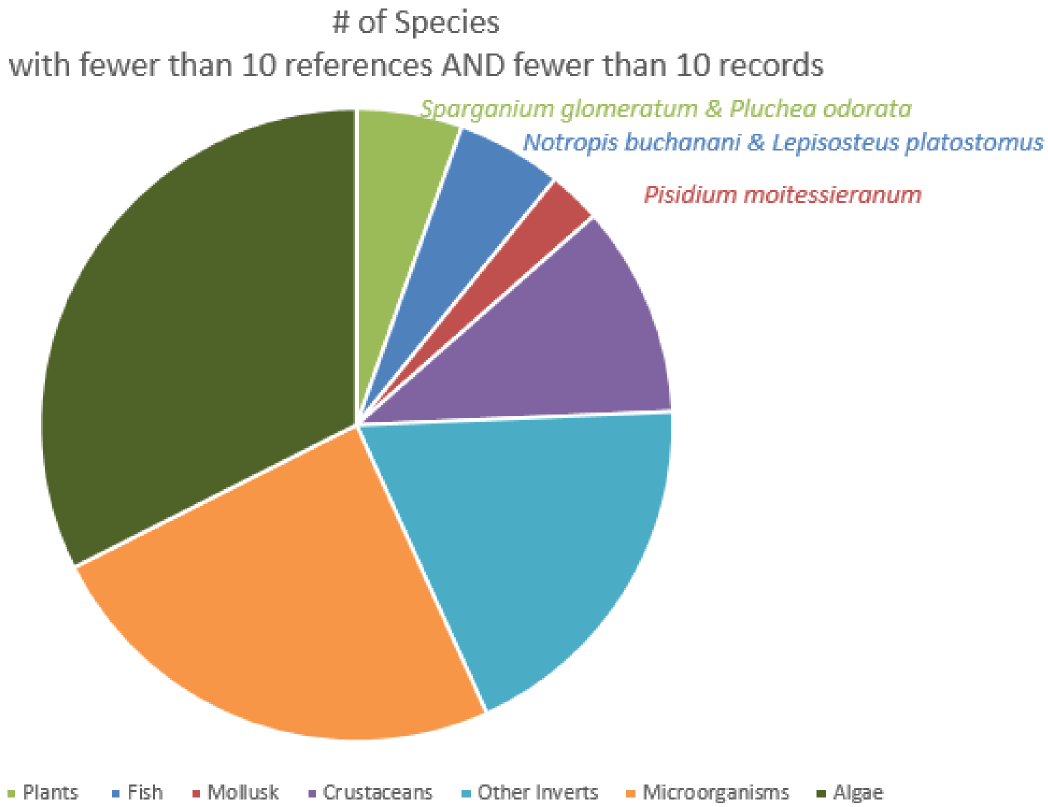


Figure 10. Understudied and Underrecorded Species (<10 references AND <10 reports). 37 of the 188 nonindigenous species (19%) are both understudied and underrecorded.

Gaps in Life History Information

A preliminary effort was made to identify and consistently gather life history information for all nonindigenous species. Only maximum size was actually consistently gathered across all taxa. Age at maturity (first reproductive age for females) was gathered for only 33 species (mostly fish), and maximum lifespan for only 61 species (including many plants just noted as annual versus perennial). Fecundity (variably collected as offspring per brood, offspring per year, or noted as vegetative) was available for just 68 species, despite being considered a critical factor for understanding species spread (See Appendix 5).

Gaps in Impact Data for Nonindigenous Species

Not all nonindigenous species are expected to have equal impacts on the new ecosystem in which they find themselves. GLANSIS conducts comprehensive externally-reviewed impact assessments for each overwintering and reproducing species in the database. References used directly in developing these assessments are accounted separately and documented in the NOAA Tech Memo series TM-161 et seq. These references are primarily from the peer-reviewed literature, but do include published anecdotal data as well as published expert opinion in addition to the direct field and laboratory studies of impact – personal communications (primarily with reviewers) were excluded from this analysis. 60% of species have fewer than 10 impact-relevant studies (Figure 11). No distinct patterns are apparent in taxonomy (Figure 11) or size (Figure 12). Vector continues to follow the trend by which the median number of impact relevant studies is greater for vectors with the greatest human intentional facilitation of introduction (Figure 13), though this is overwhelmed by variation within each vector.

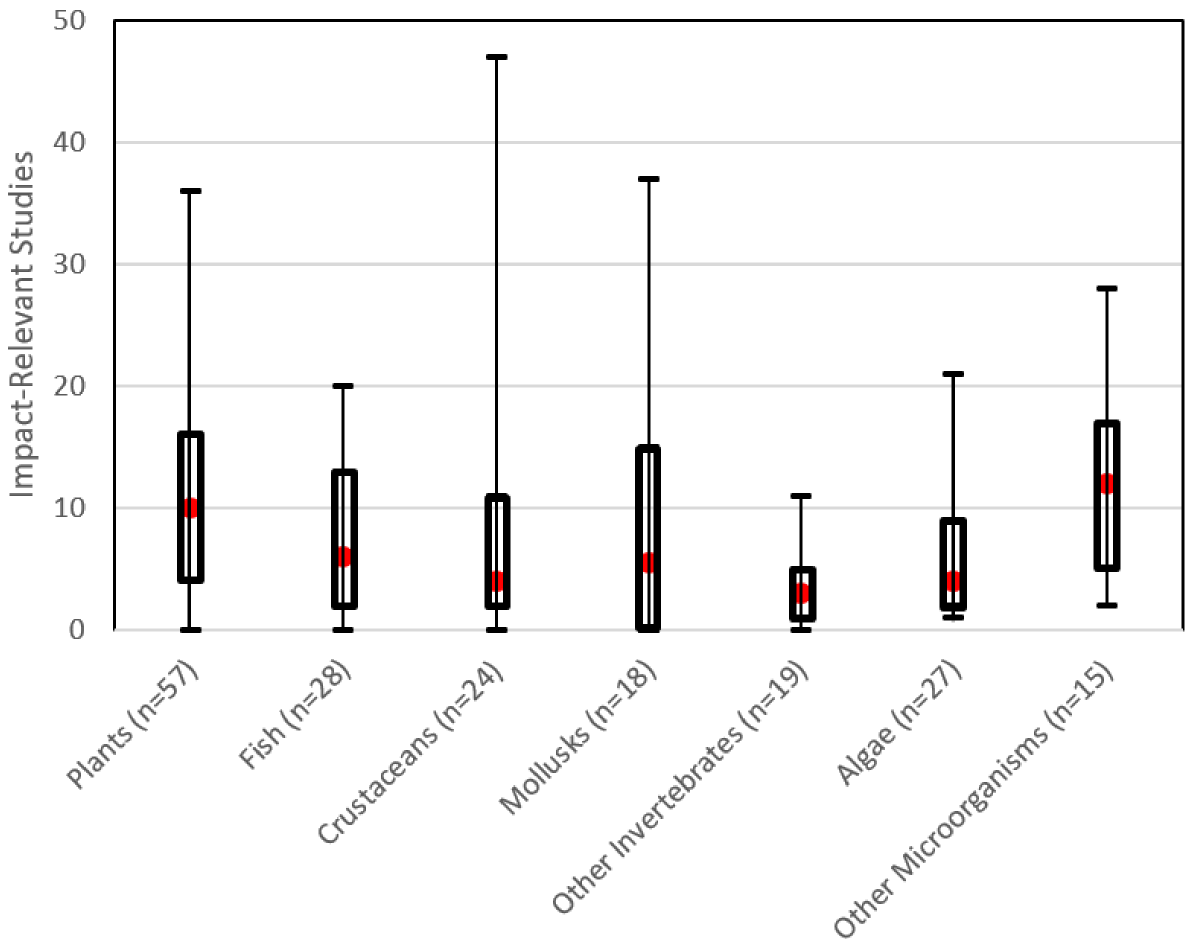


Figure 11. Quartile analysis of Impact-relevant literature by taxa.

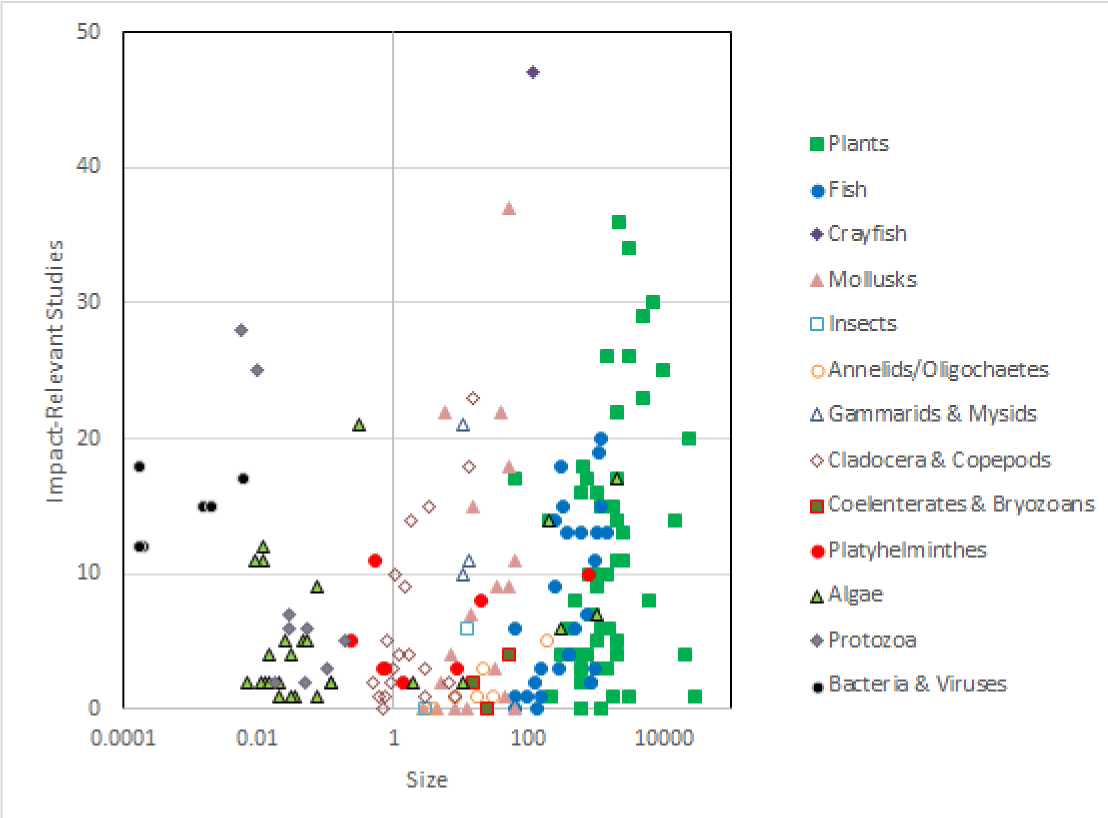


Figure 12. Impact-relevant literature by size.

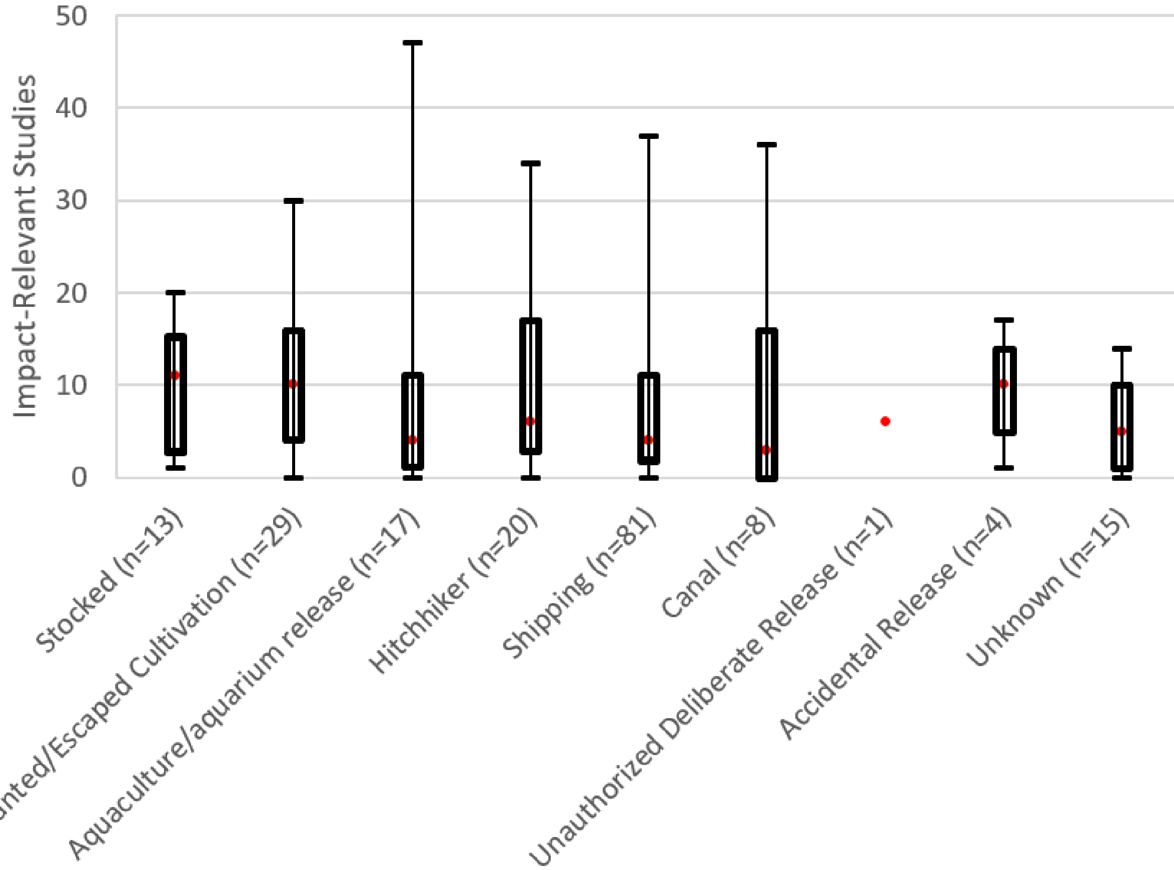


Figure 13. Quartile analysis of Impact-relevant literature by vector

GLANSIS impact assessments include 3 categories of impact (ecological, socioeconomic, and benefits) each based on 6 specific subcategories of impact. Each subcategory is ranked high, moderate, low, or unknown based on the available literature and expert judgment. Unknowns reflect gaps in the available knowledge base. Unknowns are not distributed evenly either taxonomically or by impact type (Table 2). Ecological impacts, particularly impact to predator-prey dynamics remains the greatest source of uncertainty in gauging the overall impact of nonindigenous species to the Great Lakes. Given the paucity of literature for algae and other microorganisms, they surprisingly contribute relatively little to the overall uncertainty with regard to impact.

Table 2. Number of Unknowns for each impact assessment category (n=188). Red indicates more than half of species have an 'unknown' score in this category. Orange indicates more than 25% of species scored unknown.

	Health	Competition	Predator-Prey	Genetics	Water Quality	Physical	Human Health	Infrastructure	Water Quality	Economic	Recreation	Aesthetic	Biocontrol	Commercial	Recreation	Medical	Remediation	Ecological
Unknowns - Overall	32	89	102	42	80	69	4	5	20	9	16	11	6	7	10	8	8	27
Unknowns - Plants (n=57)	13	22	35	25	24	14	1	0	3	1	0	1	2	2	0	0	4	6
Unknowns - Fish (n=28)	4	15	17	4	19	21	0	0	3	2	10	7	0	3	8	3	0	7
Unknowns - Crustaceans (n=24)	3	19	19	2	10	11	1	0	3	3	3	1	3	0	0	2	0	9
Unknowns - Mollusks (n=18)	6	14	15	5	13	13	2	3	1	1	1	1	0	2	0	0	2	0
Unknowns - Other invertebrates (n=19)	6	11	11	6	10	10	0	1	9	2	2	1	1	0	2	1	2	5
Unknowns - Algae (n=27)	0	6	3	0	2	0	0	1	1	0	0	0	0	0	0	2	0	0
Unknowns - Other Microorganisms (n=15)	0	2	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0

GLANSIS also performs a semi-quantitative assessment for each species by summing the scores in each category, which we use to assign each species as 'high Impact' (either a single very impact scoring maximally for the type or a diversity of impact types each scoring moderately), 'moderate impact' (meaning they have a measurable environmental or socioeconomic impact, but not rising to the level of high impact), 'Low Impact', which means they do not have a measurable impact, and 'Unknown', in which it is unclear whether they have an impact or not. 82 of the 188 (44%) nonindigenous species in the Great Lakes score as unknown on their overall assessment of impact. When we look at these groups – High Impact, Moderate Impact, Low Impact, and Unknown – we find that there are more references on impact for the invasive species (Figure 14). This could either be interpreted as impacts being more likely to be discovered for species that are studied extensively, or that obviously impactful species are being studied more frequently.

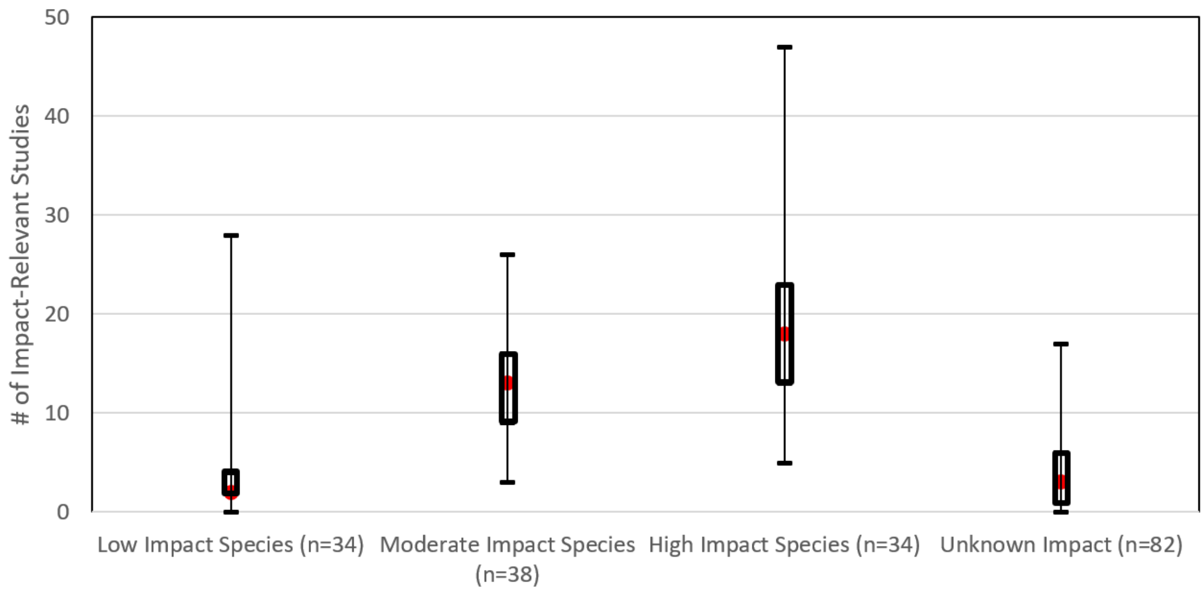


Figure 14. Quartile analysis of the impact-relevant studies by impact category.

Our semi-quantitative overall assessments have numeric 'Impact Factor' scores behind them – such that species that have higher impact and or higher diverse types of impact have higher scores. All the Unknowns and Low Impact species are clustered at IF 0-1, invasives cover a potential range of 2-72, with species scoring above 6 considered high impact (NOAA TM-161). Impact factor does not correlate significantly with the number of references (Figure 15). The extremely high impact species like zebra and quagga mussels have many more studies relevant to impact assessment than do the moderate to low impact species. Often, the additional studies are verifications of a single type of impact. Furthermore, some well-studied species (e.g., *Glugea hertwigi* with 28 impact-relevant studies) are low impact.

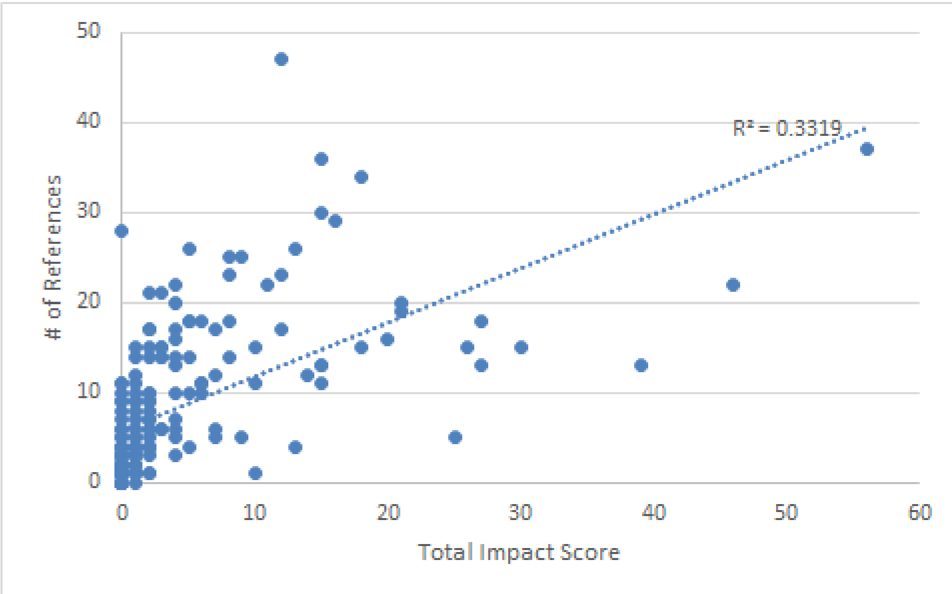


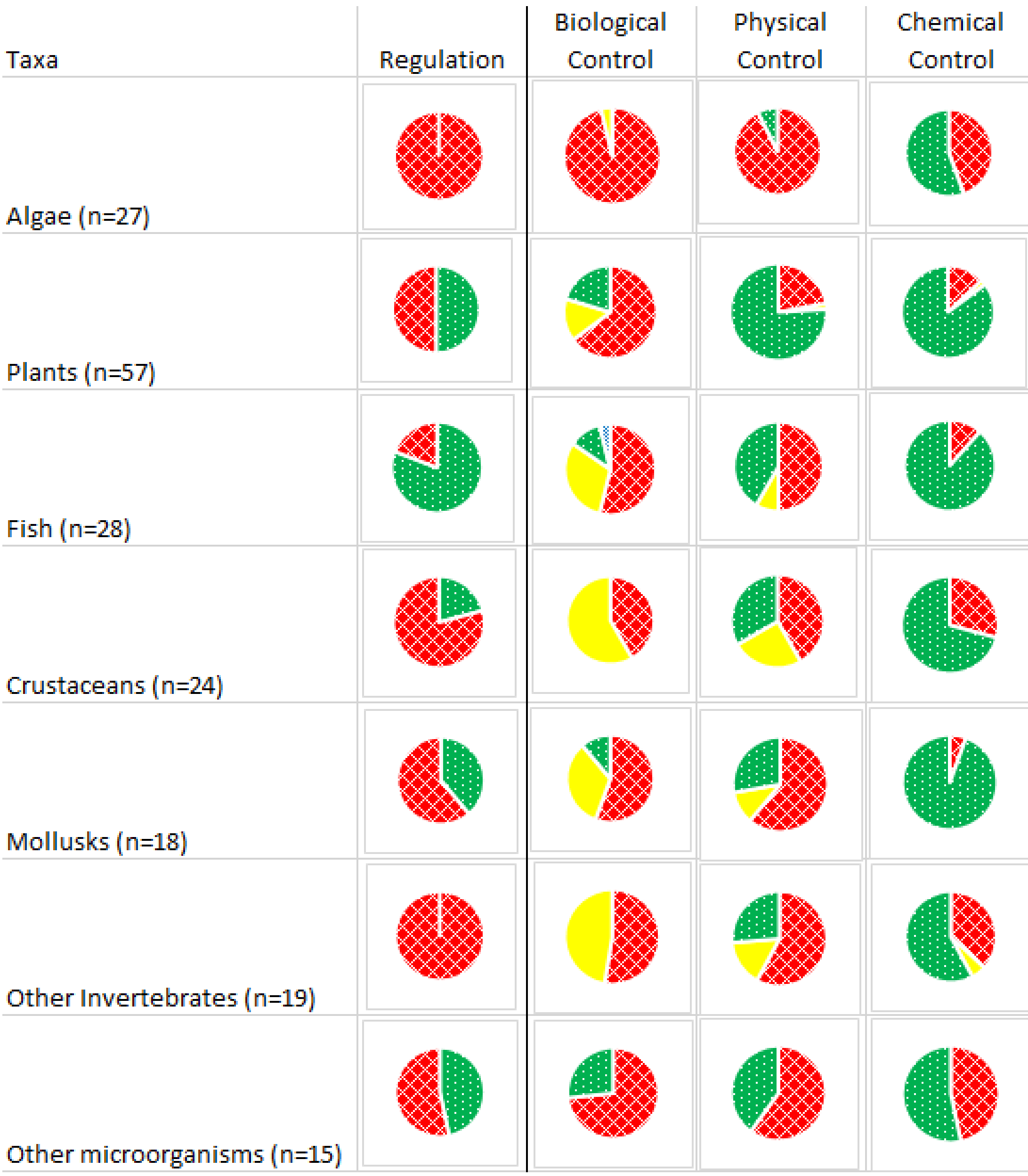
Figure 15. Total impact score does not correlate significantly with the number of impact-relevant studies.

Of the 37 species determined to be understudied and underrepresented based on evaluation of the total bibliography (above) all scored low (not unknown) for socioeconomic impact except for *Ulva flexuosa* (which has a moderate impact score based on its fouling and resistance to anti-fouling paints). Despite the lack of references in the overall bibliography, 3 understudied and underrepresented species scored high for environmental impact (fish parasites - *Ichthyocotylurus pileatus, Heterosporus sutherlandae,* and *Rhabdovirus carpio*) and 3 as moderate environmental impact (*Ulva flexuosa, Ulva intestinalis,* and *Piscirickettsia salmonis* – the first two for mat formation and the third for impacts to fish health, respectively). 18 of the understudied and underrepresented species scored as low environmental impact (due to particular key impact studies or expert review) and 13 as unknown – relative to the overall split (all listed species) of 40 low to 88 unknown, this analysis highlights that the total number of references is not entirely a good surrogate for the state of knowledge.

Gaps in Regulation and Control Information

Only 15 aquatic nonindigenous species are called out by name in International or Federal (US/Canada) regulations. *Aeromonas salmonicida, Piscirickettsia cf. salmonis, Heterosporis sutherlandae, Myxobolus cerebralis*, viral hemorrhagic septicemia (VHSv), largemouth bass virus (LMBv) and spring viremia of carp (SVCv) are restricted/listed/reportable pathogens - internationally and/or under the Fish Health Committee of the Great Lakes Fishery Commission. *Dreissena polymorpha* and *Dreissena rostriformis bugensis* are prohibited in the US by the Lacey Act. Grass carp (*Ctenopharyngodon idella*) is managed in the US by the Asian Carp Regional Coordinating Committee. *Carex acutiformis* is listed by the Canada Botanical Association as a high priority invasive plant. Use of goldfish (*Carassius auratus*) and tubenose goby (*Proterorhinus semilunaris*) as bait is prohibited in Canada. Canada also regulates *Oncorhynchus mykiss* and *Oncorhyncus nerka* as gamefish. Sixty-three species are called out in state or provincial regulations (in some cases those represent additional restrictions relative to the federal level, but not all federally regulated species have corresponding state/provincial regulations). Overall, 36% of nonindigenous aquatic species in the Great Lakes region are regulated at some level, but there are distinct difference among taxa with fish being the most-regulated species, followed by fish-pathogens (a dominant subset of 'other microorganisms'), plants, mollusks and crustaceans (Table 3).

Table 3. Regulation of nonindigenous species. Dotted green = number of species regulated (in some jurisdiction of the Great Lakes basin), hashed red = number of species unregulated (in any jurisdiction). Control of nonindigenous species. Waved blue = number of species for which control is in use for open-lake applications, Dotted green = number of species for which control has been used for limited geographic locations (e.g., small ponds, hatcheries), solid yellow = number of species for which control is available only in the research phase. Hashed red = number of species for which no control is available.



For each species, GLANSIS collates available information on control methods (both available and in research phases) without making direct comparisons or recommendations. These are subdivided as biological, physical and chemical controls. The only control which has been successfully conducted at the lakewide scale is biological control of alewife (*Alosa pseudoharengus*) by Pacific salmonids. All other successful control applications have been at limited geographic scales or particular habitat types - typically small embayments or even just in hatcheries.

True biological control agents have been approved for only a handful of species. Several weevil species have been approved for control of purple loosestrife (*Lythrum salicaria*), one for control of watermint (*Mentha aquatica*), and *Acentria ephemerella* is used as a biocontrol for Eurasian watermilfoil (*Myriophyllum spicatum*). Several Pacific salmonid species were introduced to control Alewife (*Alosa pseudoharengus*). Triploid grass carp (*Ctenopharyngodon idella*) is noted as a biological control agent for *Hydrocharis morsus-ranae, Myriophyllum spicatum*, and *Potamogeton crispus*. Zequanox for control of Dreissenid mussels is a biocontrol technology based on *Pseudomonas* bacteria. A bacteriophage is available for treatment of *Aeromonas salmonicida* in hatchery/aquaculture settings. Beyond this, manipulation of native predator populations is suggested as a potential control for 39 species, bottom-up food web manipulation for 4 species, and parasite manipulation for 10 species. Manipulation of native competitors (e.g., re-seeding with a native competitor following removal) is suggested for several plant species and control of host species (both native and non-native) for several others. Dreissenid removal/control is suggested as a potential control for six nonindigenous species that benefit from the refugia created by mussel beds. Also falling under the header of biological control, sterile male release has been used as a control for sea lamprey (*Petromyzon marinus*) and explored as an option for several other species. Vaccinations are available for hatchery broodstock effective against 3 nonindigenous fish pathogens.

Physical control is most typically removal of individuals -- netting or trapping of fauna, harvesting or mowing plants. A handful of plant species are managed via controlled burn. A handful of barrier methods -- benthic barriers, dams, etc.. - are also included in this category.

Chemical controls are widely available for most taxa - pisicides, molluscides, algicides, herbicides; however, in many cases non-target mortalities are not insignificant and must be given careful consideration in any application. Anti-fouling paints and disinfectants are also included in this category of control. For 36 species, reduction of nutrients/salinity or other manipulations of water quality parameters are suggested as effective potential control options. For most parasites/pathogens, chemical control options are available only for use in confined hatchery/aquaculture settings and for disinfection of gear moved between sites to prevent spread. Potential use of pheromones as lures for trapping or in conjunction with other controls is the subject of research for several species.

Gaps in the Scientific Literature for the Watchlist

Full bibliographies available in GLANSIS for predicted invasive species (watchlist, n=79) average 43 references per species (total is not reported because some references were used for multiple species). The number of references for each species range from 5 (for *Lysimachia punctata*) to 421 (for *Hydrilla verticillata*). Only 3 species qualify as understudied using the definition applied to the nonindigenous list (<10 references available) – *Lysimachia punctata, Leyogonimus polyoon*, and *Sinelobus stanfordii*; however this may be a circular argument, as understudied species are unlikely to have been considered for listing given a criteria of 'identification in the peer reviewed literature as likely to invade the Great Lakes'. There is less variance in the number of references per species (relative to the nonindigenous list), with 89% of species falling between 10 and 100 references. The number of references available does not appear to be biased by taxonomy (Figure 16), size (Figure 17), or vector of introduction (Figure 17). This could be due to species selected for the watchlist having an elevated level of awareness to begin with. Note that more than half of the watchlist is based on risk via the ballast vector. This may reflect a bias in the literature at the time of the list development rather than a true (or on-going) higher risk of invasion via ballast. Similarly, microorganisms (<1 mm maximum length) make up just 8% of the watchlist but 25% of the nonindigenous list, which may reflect a lack of available risk assessment information on microorganisms.

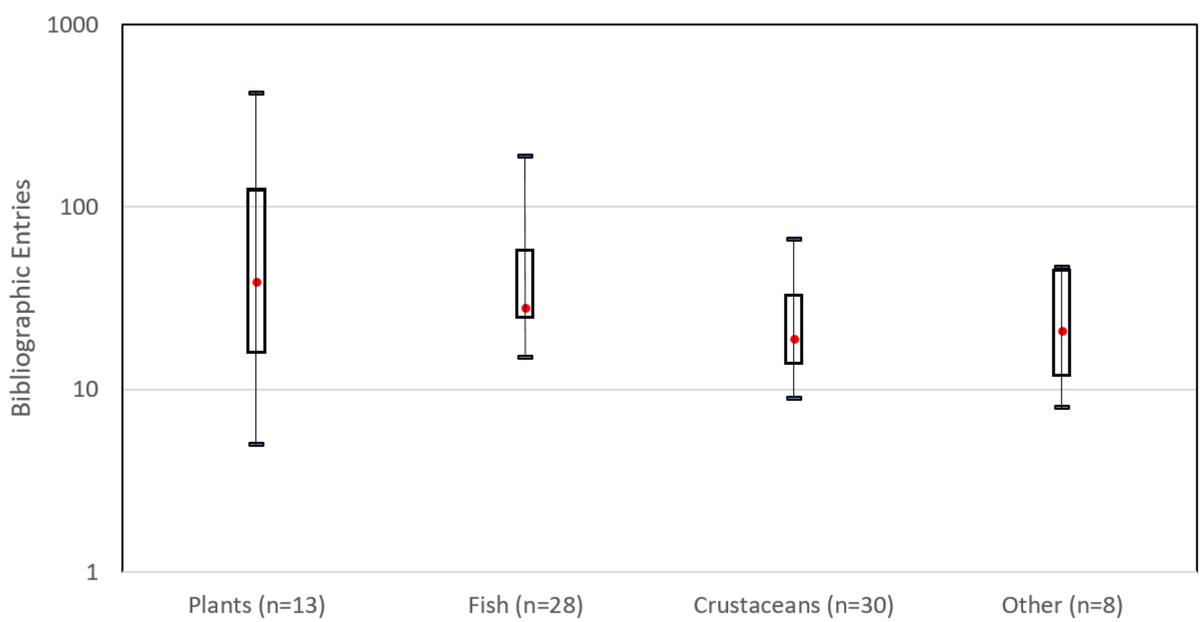


Figure 16. Quartile Analysis of the Total Bibliographic Holdings for Watchlist Species in the Great Lakes by Taxa. 'Other' includes 2 mollusks, 3 rotifers, a platyhelminth, a bryozoan and an annelid.

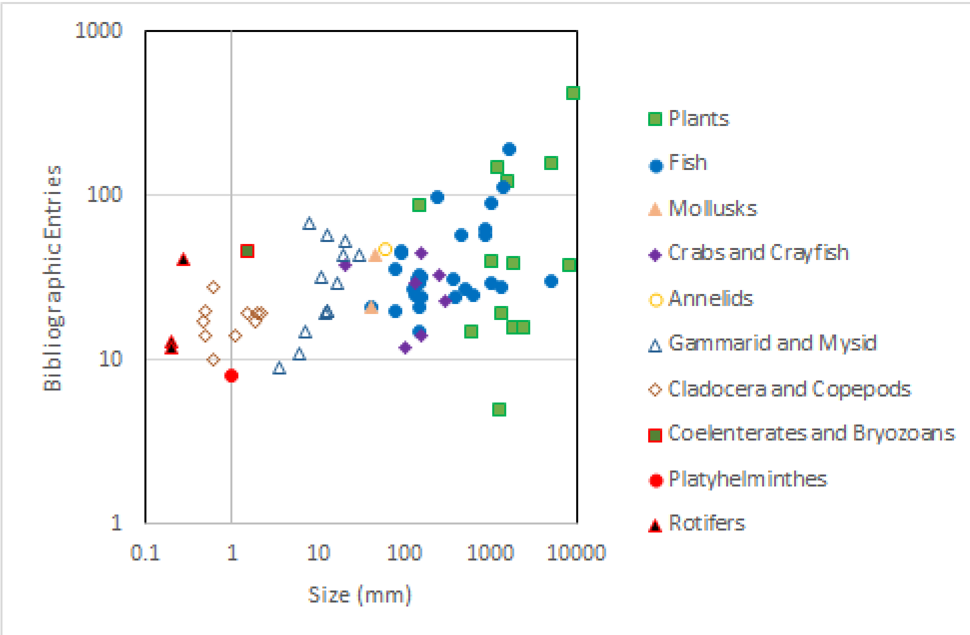


Figure 17. Total Bibliographic Holdings for Watchlist Species in the Great Lakes by Size.

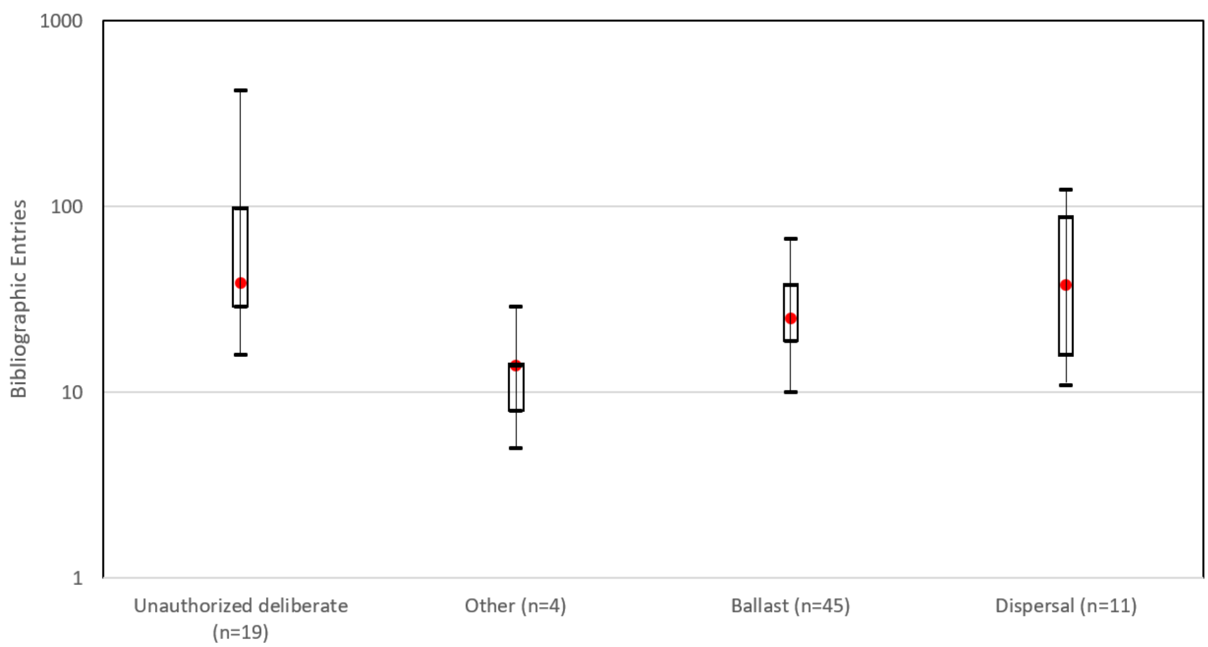


Figure 18. Quartile Analysis of GLANSIS Watchlist Bibliography by Vector. Three species assigned as aquarium/aquaculture escape and 1 hitchhiker are included in the 'Other' category.

Gaps in Risk Assessment Data for the Watchlist

GLANSIS conducts risk assessments for species identified in the peer-reviewed literature as having a significant risk for invading the Great Lakes. Number of references are accounted separately for each segment of the risk assessment – risk of introduction, risk of establishment, and risk of impact.

The watchlist consists of 79 species, with an average of 21 references per species cited in parts A&B (Introduction and Establishment). More than half of the species included in the watchlist are attributed to the ballast vector, which likely reflects a bias within the literature itself at the time the original core watchlist was developed (~2010). Ballast water invaders designated as high risk for establishment also have a significantly higher average number of references per species (34 references per species). One outlier species (*Babka gymnotrachelus*) had an extraordinarily high number of studies. There are no clear trends in the median number of references by taxa, size or vector of introduction (Figures 19, 20 and 21). Only species attributed primarily to the unauthorized intentional release vector (which includes dumping of aquariums and bait in addition to unauthorized stocking) scored as unknown for probability of introduction, and no species scored as unknown for the probability of establishment. Number of references available does not appear to significantly influence the likelihood of a species being designated as high probability of introduction (Figure 22), but there is a trend in the median values for risk of establishment (Figure 23) such that species considered higher risk of establishment have more risk-relevant citations than those deemed lower risk.

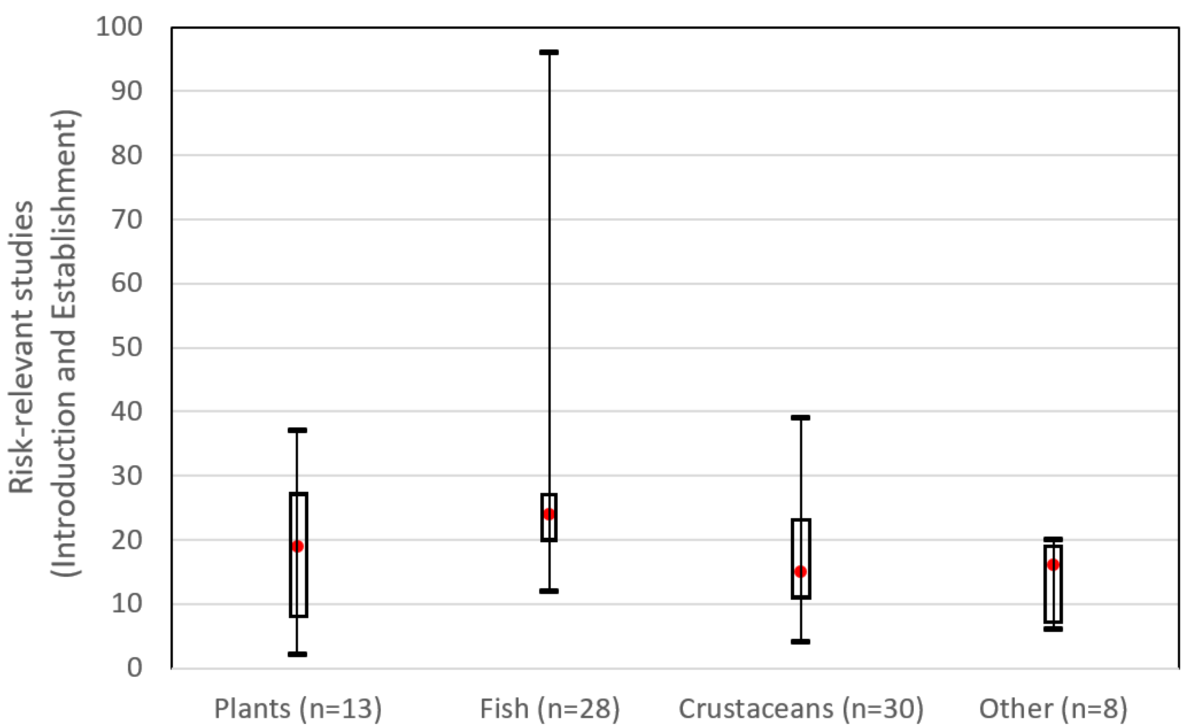


Figure 19. Quartile analysis of the number of references for risk (of introduction and establishment) by taxa. 'Other' includes 2 mollusks, 3 rotifers, a platyhelminth, a bryozoan and an annelid.

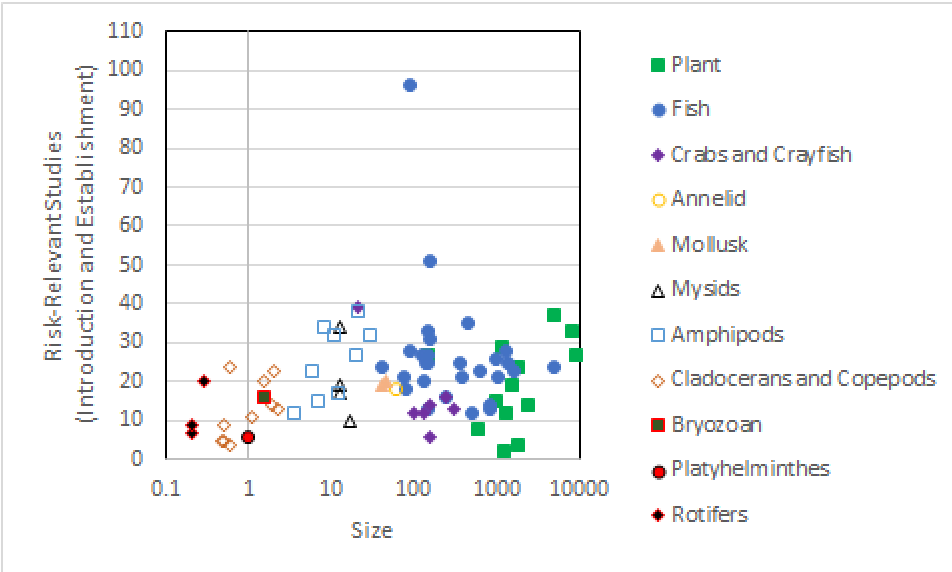


Figure 20. Risk-Relevant Literature (Introduction and Impact) for Watchlist Species by Size.

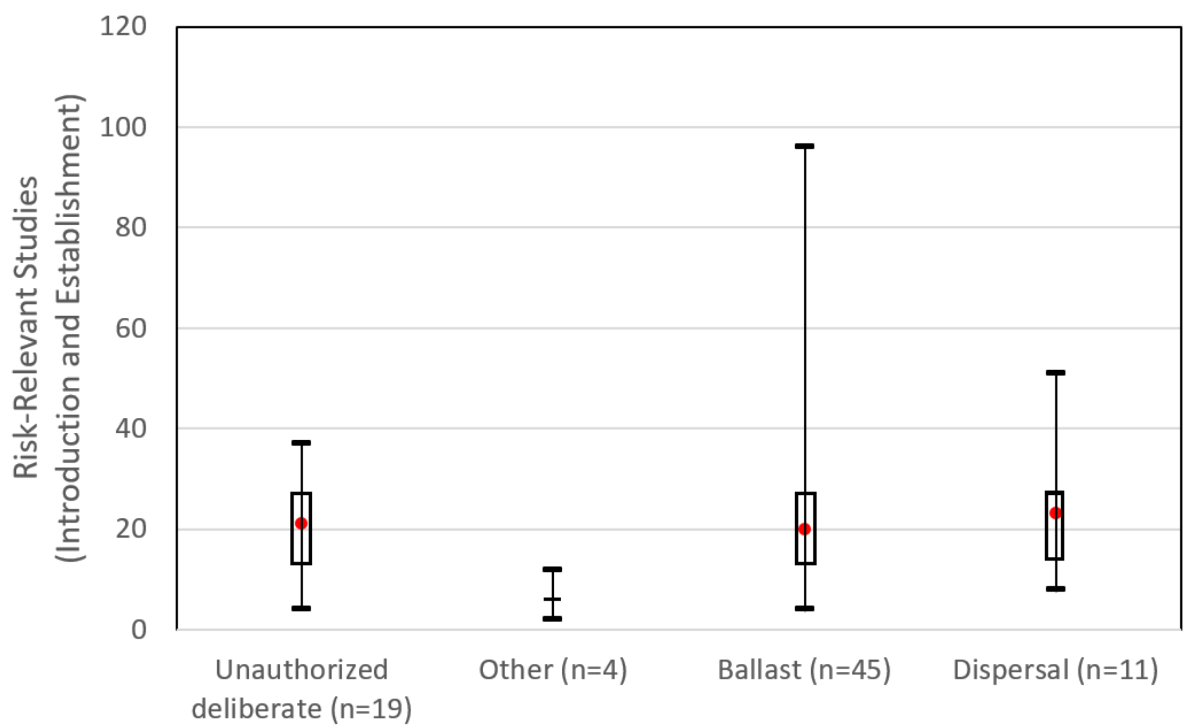


Figure 21. Quartile analysis of the number of references for risk (of introduction and establishment) by vector. Three species assigned as aquarium/aquaculture escape and 1 hitchhiker are included in the 'Other' category.

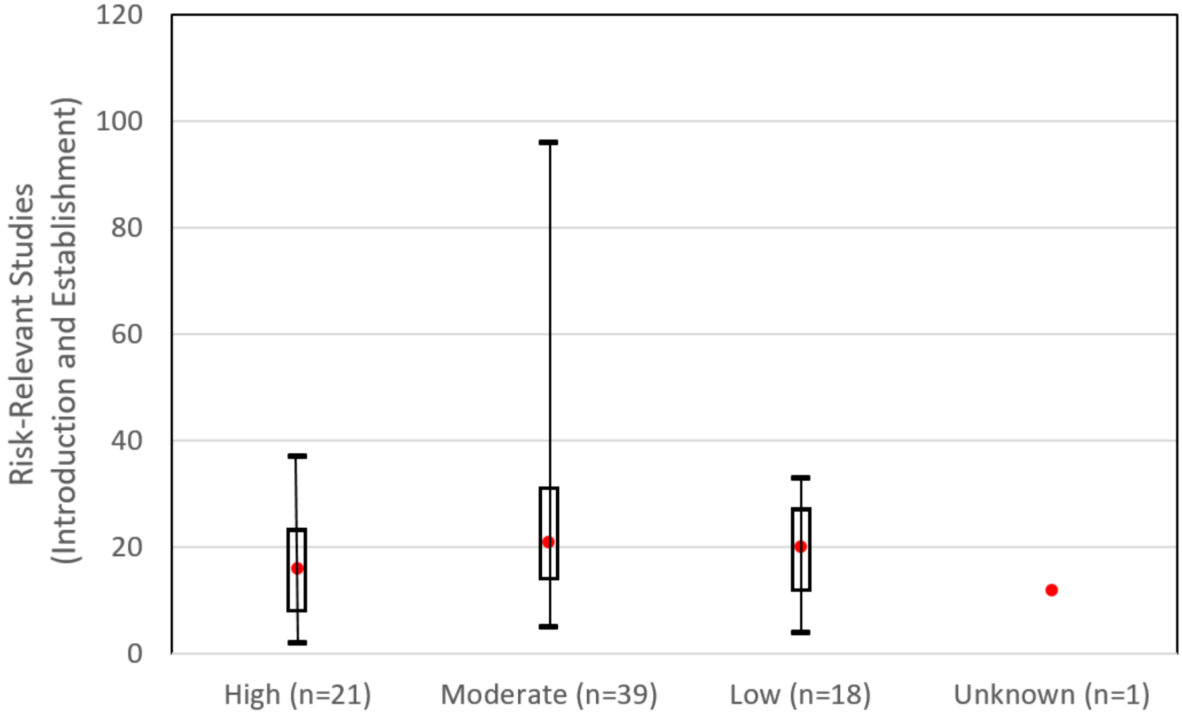


Figure 22. Quartile analysis of the number of references for risk (of introduction and establishment) by Introduction Risk Category (High, Medium, Low and Unknown Risk of Introduction).

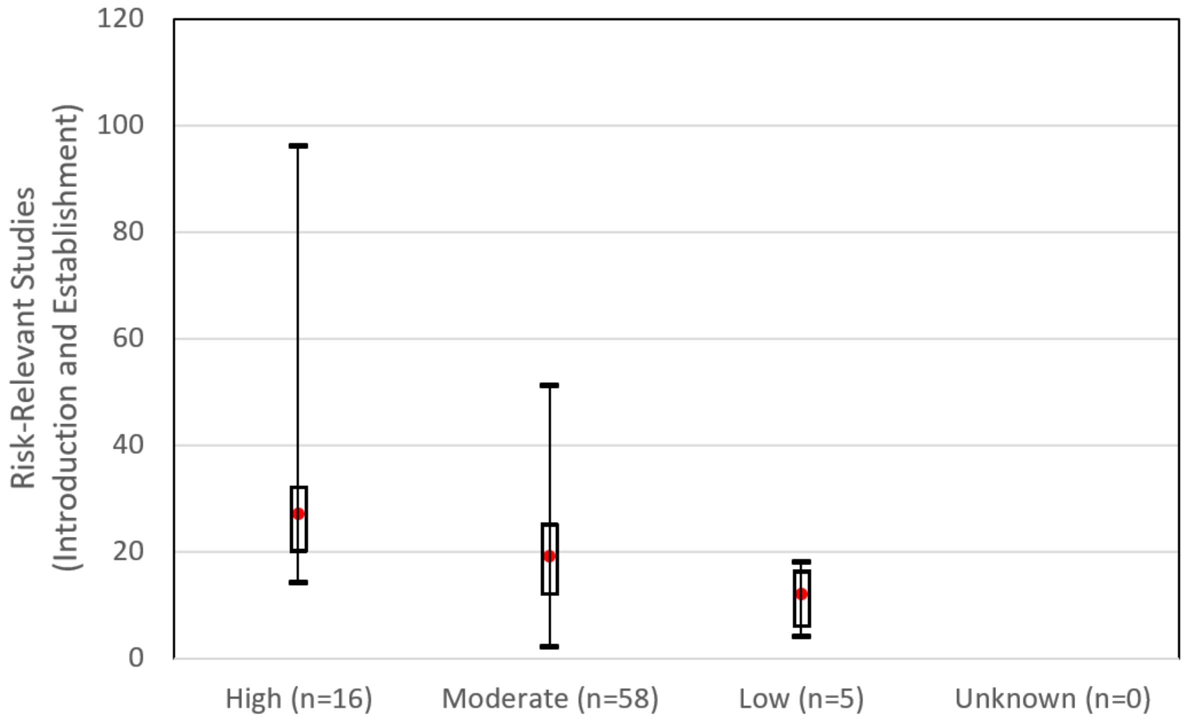


Figure 23. Quartile analysis of the number of references for risk (of introduction and establishment) by Establishment Risk Category (High, Medium, Low and Unknown Risk of Establishment).

The references for part C of the risk assessment (probability of impact) were accounted separately, as they had less overlap and were significantly fewer in number with an average of just 9 impact-relevant references per species, with no apparent trends in the medians across taxa (Figure 24). A distinct trend with size, on the other hand, is apparent in the data, with microorganisms having very few impact-relevant studies (Figure 25) in addition to being a relatively small component of the overall watchlist. Significantly more impact-relevant studies were found for species predicted to invade via dispersal (17 references per species) than for other vectors (Figure 26). All vector types included species for which the potential impact could not be determined (unknown probability of impact). The apparent trend in the medians is for higher impact species to have more impact-relevant studies, but the species scoring unknown did not have significantly lower median number of impact-relevant studies than the low impact species (Figure 27).

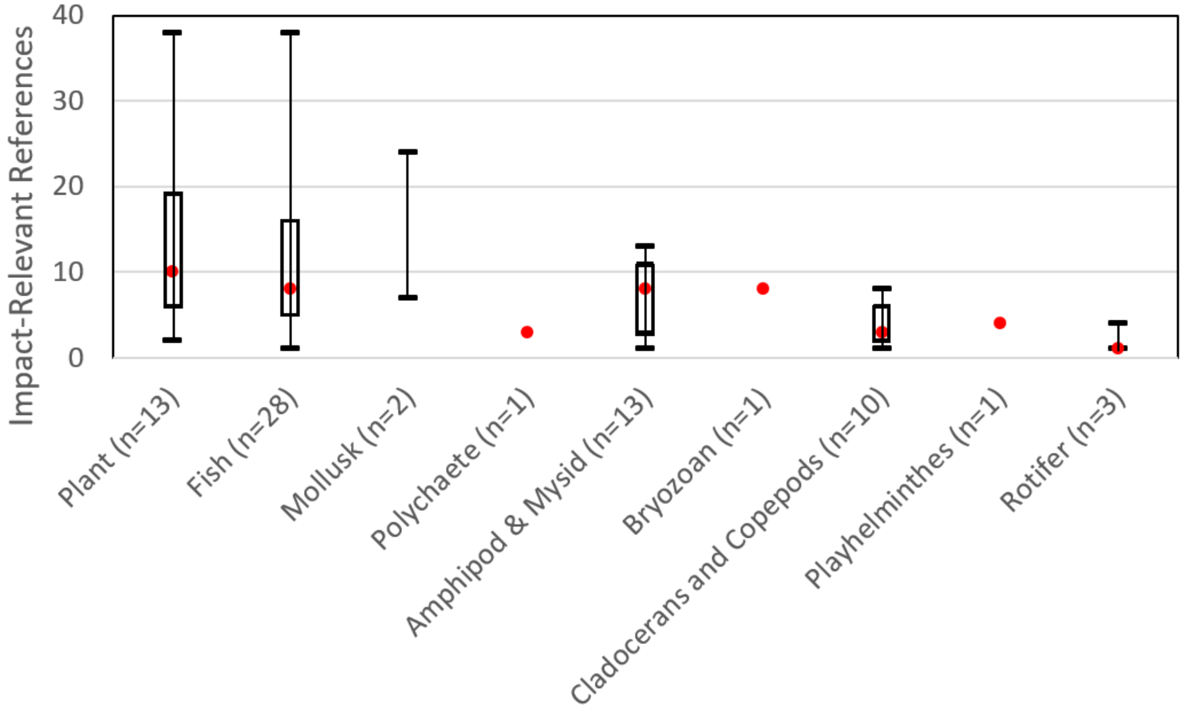


Figure 24. Impact-relevant watchlist references by taxa.

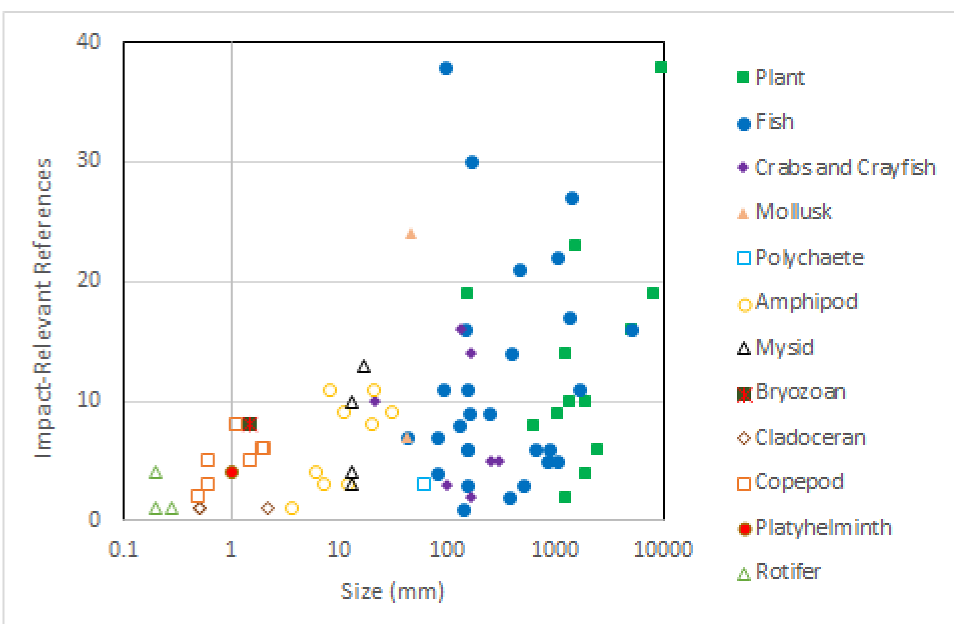


Figure 25. Impact-relevant watchlist references by size.

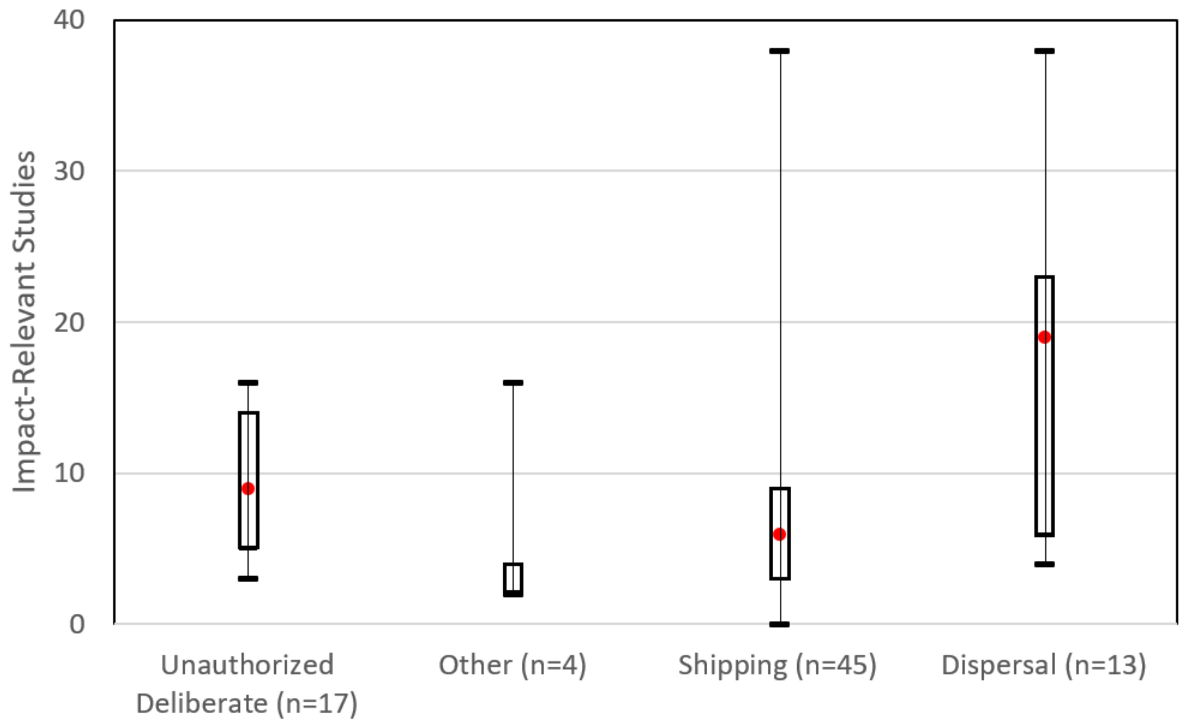


Figure 26. Impact-relevant watchlist references by vector.

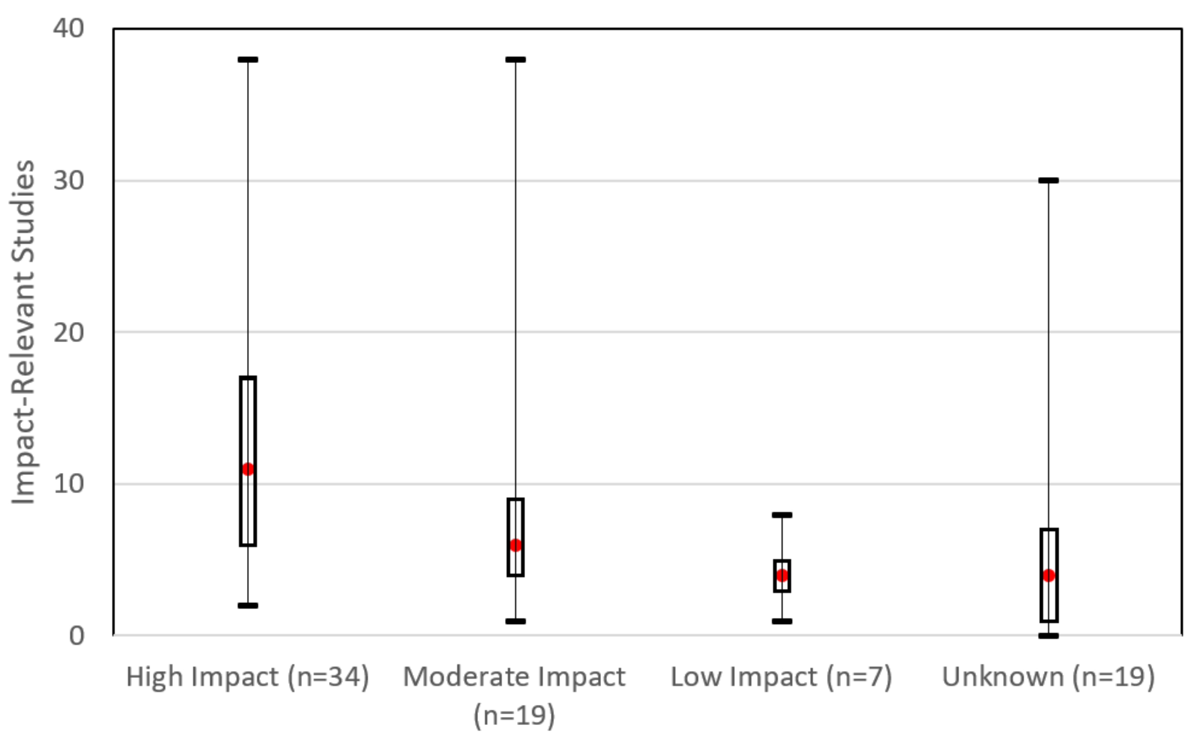


Figure 27. Impact-relevant watchlist references by impact category.

Table 4. Number of Unknowns for each risk assessment category (n=188). Red indicates more than half of species have an 'unknown' score in this category. Orange indicates more than 25% of species scored unknown.

			Overall (n=79)	Fish (n=28)	Plants (n=13)	Decapods (n=8)	Mollusks (n=2)	Amphipods & Mysids (n=12)	Cladocerans & Copepods (n=10)	Other Microorganisms n=6)
Introduction	Dispersal	Occurrence	1	1						
		Proximity	1		1					
	Hitchhiking	Transport Potential	3	1		1			1	
		Proximity	1							1
	Unauthorized intentional Release	Purchase	4	2	1	1				
		Ease	3	1	2					
	Stocking/Escape Culture	Stocking	8	2	3	3				
		Proximity	10	2	5	3				
	Commercial Culture	Commercial Culture	6	1	4	1				
		Proximity	7	1	5	1				
	Shipping	Survival	1							1
		Traffic Patterns	1		1					
Establishment	Attributes	Tolerances	1							1
		Overwinters	1							1
		Diet Flexibility	2			1			1	
		Competition	8	1		2			2	3
		Fecundity	18	1	2			3	7	5
		Reproductive Strategy	7				1	1	1	4
	Environmental Compatibility	Climate	1			1				
		Abiotic Environment	4	1	1	2				
		Habitat	1						1	
		Climate change	5		1				2	2
		Food	4			2			2	
		Critical Species	0							
		Meltdown	3			1		1		1
		Natural Enemy	12	3	1		1	6		1

		Propagule Pressure	44		18	3	3	1	10	6	3
	Invasion History	Establishment	2			1				1	
		Spread History	10		1	2	1		1	2	3
		Control	2		1	1					

Prediction of the number of individuals introduced (propagule pressure) remains the major source of uncertainty in the assessment of the risk of establishment of nonindigenous species. For more than half of the species assessed, the question "*On average, how large and frequent are inoculations (introduction events) from the potential vectors for this species?*" could not be answered (Table 4). Propagule pressure is a critical facet in determining the risk of establishment because dispersed introductions of a few individuals – particularly for sexually reproducing species that need not only survive, but also find a mate – are far more likely to fail than are frequent introductions of large numbers of individuals.

As with the core nonindigenous list, environmental impacts are the largest category of assessment questions that cannot be answered in the impact component of the risk assessment (Table 5).

Table 5. Number of Unknowns for each potential impact assessment category (n=188). Red indicates more than half of species have an 'unknown' score in this category. Orange indicates more than 25% of species scored unknown.

			Overall (n=79)	Fish (n=28)	Plants (n=13)	Decapods (n=8)	Mollusks (n=2)	Amphipods & Mysids (n=12)	Cladocerans & Copepods (n=10)	Other Microorganisms
Impact	Environmental	Health of Native	20	8	1	1		4	5	1
		Competition	14	4		3		3	3	1
		Predator-Prey	26	9	4	5		2	4	2
		Genetic	24	7	6	2		3	4	2
		Water Quality	24	10		2	1	4	4	3
		Physical	21	6	2	2	1	4	4	2
	Socioeconomic	Human Health	9	4	2				3	
		Infrastructure	0							
		Water Quality	2	1		1				
		Markets	8	3	1	1		3		
		Recreation	4	1		1		1		1
		Aesthetic	4	1	2					1
	Benefit	Biocontrol	5	4		1				
		Commercial	0							
		Recreation	2	1		1				
		Medical/Research	1	1						
		Bioremediation	3	1	1					1
		+ Ecological	3	1	2					

Risk Clearinghouse

GLANSIS developed a Risk Assessment Clearinghouse in 2019 which is currently populated with risk assessments for 2,333 species (including species already listed by GLANSIS). GLANSIS protocols captured (and listed, rejected or had immediate plans to assess) approximately 20% of the species in the Clearinghouse (Figure 28). As intended, the Clearinghouse contains a substantial number of assessments for species that are shown to prove no particular risk to the Great Lakes region (68%) – these are species that it is unlikely GLANSIS would have assessed in the standard protocols because the original literature did not suggest the species posed a risk to the region. The remaining 12% of species are those which were not previously captured in GLANSIS protocols but which are demonstrated by the other risk assessments to pose some substantial risk to the Great Lakes. This subset of 261 species (Figure 29) has now been added to the GLANSIS development plan for future investigation. This critical subset is heavily weighted toward crayfish (49%), which either reflects a bias against crayfish in GLANSIS literature searches, a strong crayfish component in the Clearinghouse reflecting the STAIR-Crayfish assessment, or both. Only 1 microorganism (a copepod) is included in the subset of organisms missed by GLANSIS (261 possible risk, 1594 no risk) in contrast to the 56 microorganisms captured by GLANSIS, suggesting that GLANSIS protocols do better than most other assessments included in the Clearinghouse at capturing the risks posed by microscopic invaders. Nonetheless, the contrast between the GLANSIS nonindigenous list (47 microorganisms representing 25% of the list) and the GLANSIS watchlist (9 microorganisms representing 11% of the watchlist) suggests more research is needed in this area.

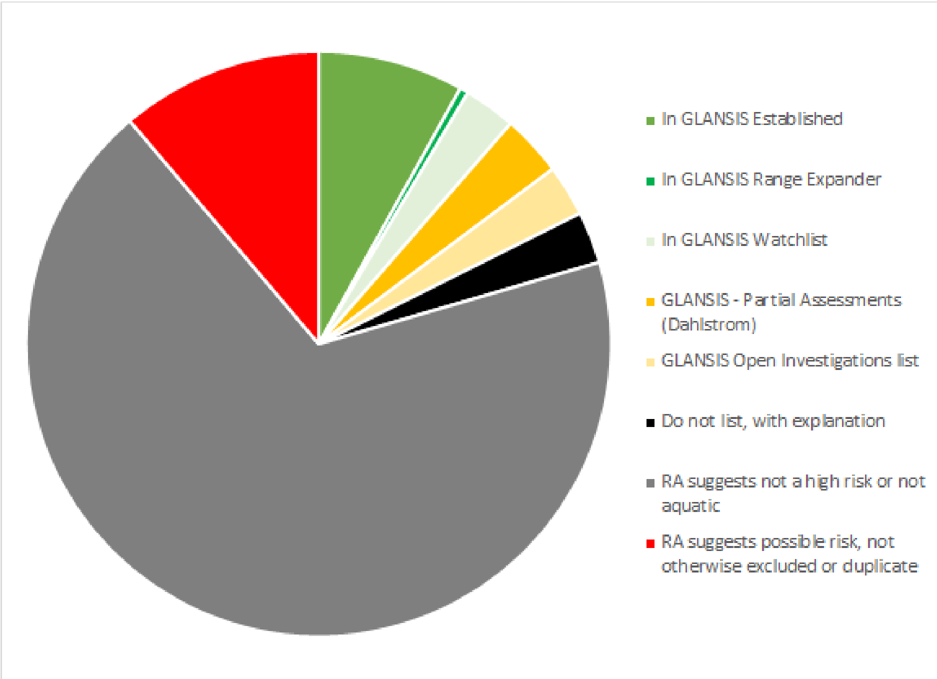


Figure 28. GLANSIS Risk Clearinghouse (n=2333) disposition of species relative to GLANSIS lists and process. In addition to the lists available publicly on GLANSIS, GLANSIS has 2 sets of species 'in progress' – a set of 79 species assessed by A. Dahlstrom pending final review and a set of 66 open investigations which passed initial screening and are pending literature review. "RA not a high risk" and "RA possible risk" are species in the Clearinghouse that had not previously been considered by GLANSIS.

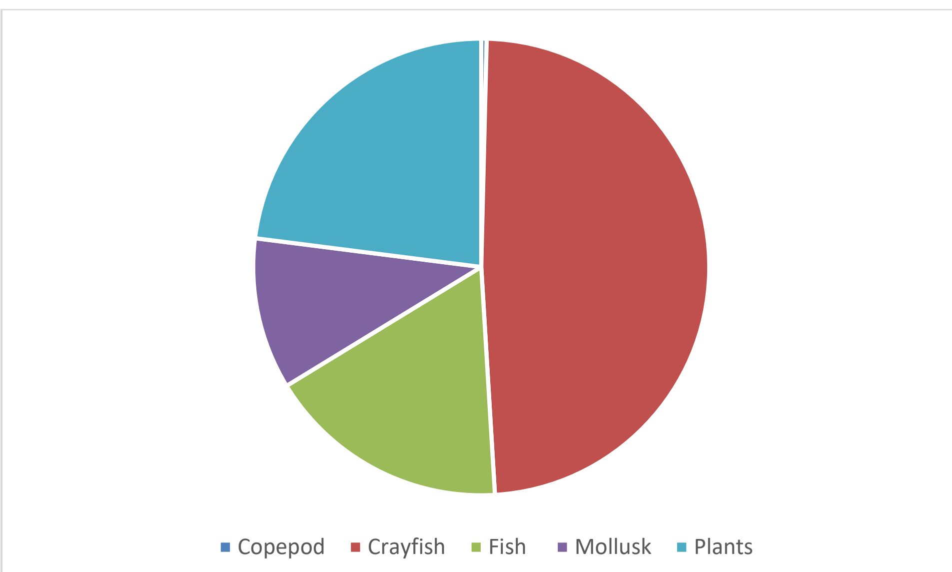


Figure 29. Possible risk species in the Clearinghouse that were not previously captured by GLANSIS (n=261, red in figure 26) by taxa.

Conclusions and Recommendations

The purpose of this gap analysis was to discover *unknown unknowns*, to identify gaps in the body of scientific knowledge of Great Lakes aquatic invasive species represented by the GLANSIS database. Despite equal effort in collating literature, we found 3 major, consistent gaps in the body of literature compiled within and supporting GLANSIS which we believe are reflective of the state of invasion science. We put these forward as needs for future regional research.

- More study is needed on the risks posed by microscopic species both free-living and parasitic. Microscopic species represent 25% of the GLANSIS nonindigenous species list (overwintering and reproducing in the Great Lakes), but 78% of the understudied and underreported species among this same group. Further, microscopic species form 80% of the nonindigenous species established in the Great Lakes in the last 25 years. However, only 11% of the watchlist and just 2% of the Risk Clearinghouse are microscopic species. This discrepancy is a strong indicator that current risk assessment methods are failing to adequately capture potential microscopic invaders. Previous work does not indicate that microorganisms are consistently lower impact as the top 8 highest impact species include XX microorganisms (Lower et al., 2020) and microorganisms are included in high, medium and low impact categories (TM-169et seq).
- More evidence is needed to document the ecological impacts of nonindigenous species, particularly impacts to predator-prey dynamics. These impacts are scored as unknown for more than half of the reproducing and overwintering nonindigenous species listed in GLANSIS, even allowing for inclusion of expert judgement, anecdotal and indirect evidence. Similarly, potential ecological impacts are scored as unknown for more than 25% of the current watchlist species.
- In order to evaluate risk of establishment in the Great Lakes region, more study is needed to predict propagule pressure for each potential invader. For more than half of the species assessed, the question "*On average, how large and frequent are inoculations (introduction events) from the potential vectors for this species?*" could not be answered. Prediction of the number of individuals introduced (propagule pressure) remains the major source of uncertainty in the assessment of the risk of establishment of nonindigenous species. Propagule pressure is a critical facet in determining the risk of establishment because dispersed introductions of a few individuals – particularly for sexually reproducing species that need not only survive, but also find a mate – are far more likely to fail than are frequent introductions of large numbers of individuals.

Potential geographic gaps in the state of knowledge for nonindigenous species in the Great Lakes basin remain unclear. GLANSIS does not track absence data or contain consistent data on numbers relative to collection effort, which means that we do not have a reliable way to separate underreporting (GLANSIS effort at collating data) from actual undersampling (by field scientists) versus actual differences in number of species present (resistance of watersheds/habitats to invasion). The high number of records (for the nonindigenous list species) per species per area at some 'sentinel' sites relative to the average number of species suggests that areas with high reporting frequency are merely accumulating more reports of the same species rather than finding additional species.

Literature Cited

Baker, R., R. Black, G. Copp, K. Haysom, P. Hulme, M. Thomas, A. Brown, M. Brown, R. Cannon, J. Ellis, E. Ellis, R. Ferris, P. Glaves, R. Gozlan, H. Holt, L. Howe, J. Knight, A. MacLeod, N. Moore, J. Mumford, S. Murphy, D. Parrott, C. Sansford, G. Smith, S. St Hilaire, N. Ward. 2007. The UK risk assessment scheme for all non-native species. *NeoBiota* 7: 46–57

Davidson, AD, AJ Fusaro, RA Sturtevant and DR Kashian. 2017. Development of a risk assessment framework to predict invasive species establishment for multiple taxonomic groups and vectors of introduction. *Management of Biological Invasions* 8(1)25-36.

Fusaro, A., E. Baker, W. Conard, A. Davidson, K. Dettloff, J. Li, G. Nunez-Mir, R.A. Sturtevant, and E.S. Rutherford. 2016. A risk assessment of potential Great Lakes aquatic invaders. NOAA GLERL-TM 169.

Lower, E.K., N. Boucher, R.A. Sturtevant, and A.K. Elgin. 2020. 2019 Update to "An Impact Assessment of Great Lakes Aquatic Nonindigenous Species". NOAA GLERL-TM 161c.

Lower, E.K., N. Boucher, P.J. Alsip, A. Davidson, and R.A. Sturtevant. 2019. 2018 Update to "A risk assessment of potential Great Lakes aquatic invaders". NOAA GLERL-TM 169b.

Lower, E.K., N. Boucher, A. Davidson, A.K. ELGIN, R.A. Sturtevant. 2020. 2019 Update to "A risk assessment of potential Great Lakes aquatic invaders". NOAA GLERL-TM 169c.

NOAA. 2020. Great Lakes Aquatic Nonindigenous Species Information System (GLANSIS). https://www.glerl.noaa.gov/glansis/

Sturtevant, R., J. Larson, L. Berent, M. McCarthy, A. Bogdanoff, A. Fusaro, and E. Rutherford. 2014. An Impact Assessment of Great Lakes Aquatic Nonindigenous Species. NOAA GLERL-TM 161.

Sturtevant, R.A., E.K. Lower, N. Boucher, P.J. Alsip, K. Hopper, S. Iott, D.M. Mason, A.K. Elgin, and F.A. Martinez. 2019. 2018 Update to "An Impact Assessment of Great Lakes Aquatic Nonindigenous Species". NOAA GLERL-TM 161b.

Appendix

Data collected for this report can be downloaded in this Microsoft Excel spreadsheet here: https://www.glerl.noaa.gov/pubs/tech_reports/glerl-176/tm-176_appendix.xlsx.

Please see the tab descriptions below.

Tab 1: Bibliographic Holdings for Nonindigenous Species

Bibliography = all papers available in the database with information pertinent to the species.

Profile = all papers cited in the GLANSIS profile (unique studies of significant contribution to understanding the species).

OIA = all papers cited in the organism impact assessment (providing significant information on the impact of the species).

Note that there should be autocorrelation among these categories. All papers included in Profile and OIA categories should also be in Bibliography. Most of the papers in OIA are also in the Profile.

Color-coding:

Green = 'Highly Studied' – Top Quartile with the greatest number of publications.

Red = 'Understudied' – Bottom Quartile with least number of publications.

Tab 2: Bibliographic Holdings for Watchlist Species

Comments:

Bibliography = all papers available in the database with information pertinent to the species.

Profile = all papers cited in the GLANSIS profile (unique studies of significant contribution to understanding the species).

Corrected = uses the profile to supplement the bibliography for species which (a) were not available through NAS or (b) for which numerous references included in the profile were not included or not tagged to the species in NAS.

Note that there should be autocorrelation among these categories. All papers included in Profile and OIA categories should also be in Bibliography. Most of the papers in OIA are also in the Profile.

TM-169 a, b, c indicate the update version used as a source for the risk assessment data.

I/E = Total number of original papers cited in the Introduction and Establishment components of the risk assessment.

Impact = Total number of original papers cited in the Impact component of the risk assessment.

Color-coding:

Green = ‘Highly Studied’ – Top Quartile with the greatest number of publications.

Red = ‘Understudied’ – Bottom Quartile with least number of publications.

Tab 3: Impact of Nonindigenous Species

Comments: Ranked from greatest to least unknowns. See Columns A-U (with N-S hidden).

Tab 4: Risk Assessment of Watchlist Species

Tabs 5a-c: Profile assessment of Nonindigenous Species

- Tab 5a – Habitat
- Tab 5b – Life History
- Tab 5c – Regulation and Control