CrossMark

ORIGINAL PAPER

Are genetic databases sufficiently populated to detect non-indigenous species?

**Elizabeta Briski · Sara Ghabooli ·
Sarah A. Bailey · Hugh J. MacIsaac**

Abstract Correct species identifications are of tremendous importance for invasion ecology, as mistakes could lead to misdirecting limited resources against harmless species or inaction against problematic ones. DNA barcoding is becoming a promising and reliable tool for species identifications, however the efficacy of such molecular taxonomy depends on gene region(s) that provide a unique sequence to differentiate among species and on availability of reference sequences in existing genetic databases. Here, we assembled a list of aquatic and terrestrial non-indigenous species (NIS) and checked two leading genetic databases for corresponding sequences of six genome regions used for DNA barcoding. The genetic databases were checked in 2010, 2012, and 2016. All four aquatic kingdoms (Animalia, Chromista, Plantae and Protozoa) were initially equally represented in the genetic databases, with 64, 65, 69, and 61 % of NIS included, respectively. Sequences for terrestrial NIS were present at rates of 58 and 78 % for Animalia and Plantae, respectively. Six years later, the number of sequences for aquatic NIS increased to 75, 75, 74, and 63 % respectively, while those for terrestrial NIS increased to 74 and 88 % respectively. Genetic databases are marginally better populated with sequences of terrestrial NIS of plants compared to aquatic NIS and terrestrial NIS of animals. The rate at which sequences are added to databases is not equal among taxa. Though some groups of NIS are not detectable at all based on available data—mostly aquatic ones—encouragingly, current availability of sequences of taxa with environmental and/or economic impact is relatively good and continues to increase with time.

E. Briski (✉)
GEOMAR, Helmholtz-Zentrum für Ozeanforschung Kiel, 24105 Kiel, Germany
e-mail: elzabriski@yahoo.com;
ebriski@geomar.de

S. Ghabooli · H. J. MacIsaac
Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON N9B 3P4, Canada

S. A. Bailey
Great Lakes Laboratory for Fisheries and Aquatic Sciences, Fisheries and Oceans Canada, Burlington, ON L7S 1A1, Canada

Introduction

Biological invasions are a complex process that can be viewed as a series of stages, including transport, introduction, establishment and spread (Kolar and

Lodge 2001; Colautti and MacIsaac 2004). Management efforts focused on interrupting the invasion process, particularly at the transport or introduction stage, are of great significance as they are more effective than eradication or control of established populations of non-indigenous species (NIS) (Lodge et al. 2006; Lockwood et al. 2007; Hulme et al. 2008). Many transport vectors, however, are still not effectively managed, and species continue to arrive in new habitats (Hulme et al. 2008; Kelly et al. 2009; Conn et al. 2010; Sephton et al. 2011; Briski et al. 2012a, b, 2013). Additionally, incomplete taxonomic, biogeographic and historical data frequently result in an inability to determine if newly reported species are native or non-indigenous (Carlton 2009). Incorrect species identifications could artificially inflate or depress the number of NIS in an ecosystem, and lead to misdirecting limited resources against harmless species or inaction against problematic ones (Bax et al. 2001; Simberloff 2009). As a result, accurate identification of species is typically highlighted as an essential component of invasion management strategies (Bax et al. 2001).

DNA barcoding is becoming a promising and reliable tool for species identifications (Cross et al. 2010; Briski et al. 2011). Particularly in invasion ecology, where early detection is tremendously important, molecular identification has several advantages over morphological identification (Cross et al. 2010; Briski et al. 2011). The latter often requires examination of mature specimens of a particular sex, or flowering or fruiting specimens for some plant species (Radford et al. 1968; Cross et al. 2010), which may or may not be present in initial collections of individuals from a new habitat. In contrast, molecular methods allow identification of NIS at any life stage, based on successful DNA extraction from a single individual, egg, or seed—possibly facilitating early detection of NIS before an introduced population becomes fully established in an area (Armstrong and Bell 2005; Chown et al. 2008; Briski et al. 2011; Zhan and MacIsaac 2015). Early identification of NIS, followed by immediate eradication before reproductive or flowering phases, may prevent distribution of eggs, seeds or pollen, circumventing the establishment of the next generation, admixture of genetic material among distinct NIS populations or hybridization with closely related species (Kolbe et al. 2007; Ayres et al. 2008; Cross et al. 2010). Furthermore, new sequencing

technologies, collectively called "Next-Generation Sequencing", have the ability to generate massive amounts of sequence data in one run and allow screening of whole ecosystems (Hall 2007; Rokas and Abbot 2009; Zhan et al. 2013; Zhan and MacIsaac 2015). By assessing multiple barcoding regions using universal primers, it is possible to simultaneously identify not only NIS, but also their associated microbiota, parasites and fellow travelers (Cross et al. 2010).

Use of DNA barcodes for species identification has its own weaknesses. The efficacy of DNA barcoding depends on gene region/s that provide a unique sequence to differentiate among species (Hebert et al. 2003; Cross et al. 2010) and availability of reference sequences in existing genetic databases (Darling and Blum 2007; Briski et al. 2011). Originally, the aim was to have one DNA barcode that would discriminate among all species across all phyla (Janzen 2004; Hebert and Gregory 2005), but this objective has proven unlikely as genomes vary considerably (Shearer and Coffroth 2008; Cross et al. 2010). Consequently, the cytochrome c oxidase subunit I (COI) gene has become the standard DNA barcoding marker for most animal groups (Hebert et al. 2003), the internal transcribed spacer (ITS) has been applied for a wide array of groups including plants, fungi, algae, and animals (Kress et al. 2005), while ribulose-bisphosphate carboxylase (rbcL) and maturase K (matK) genes differentiate most plants (Hollingsworth et al. 2009). The availability of reference sequences in genetic databases for these gene regions varies among taxonomic groups (Briski et al. 2011). We recently reported that only 5, 3.5, and 3.5 % of all described Rotifera, Bryozoa, and Copepoda species, respectively, had reference sequences of COI or small subunit ribosomal 16S rDNA (16S) in the Barcode of Life Database (BOLD) or GenBank (Briski et al. 2011); however, 54 % of known Branchiopoda species are represented. The Consortium for the Barcode of Life fosters development of international alliances to build a global barcode library, continuously increasing the number of available species barcode sequences in the BOLD database to create a global bio-identification system covering all eukaryotic taxa (Ratnasingham and Hebert 2007). In contrast, GenBank was designed to provide access within the scientific community to the most up-to-date and comprehensive DNA sequence information. GenBank

is not restricted to specific regions of the genome, and includes sequences developed for a variety of research purposes (NCBI 2015). Consequently, taxa studied, for example for medicine, pharmacy, or model species in ecological and evolutionary studies, may be better represented in GenBank.

Considering the importance of rapid identification of newly reported species in an area, and noting the different goals and applications of the two aforementioned genetic databases, this study explored availability of DNA sequences for identification of NIS. We assembled a global list of aquatic and terrestrial NIS, and then searched these databases for six genome regions relevant for species-level identification to determine the potential utility of molecular methods in invasion management. To check for an enrichment trend in the genetic databases, the databases were searched three times, in summer 2010 and 2012, and in January 2016.

Methods

From May to September 2010 we utilized Thomson's Institute for Science Information (ISI) Web of Knowledge 4.0 to search the scientific literature to assemble a global list of aquatic and terrestrial NIS. Initially, the following search terms were used: non-native OR alien OR exotic OR non-indigenous OR introduced OR colonizing—resulting in 29,975 publications. Our results were narrowed with an additional search term: list—which also improved the prevalence of studies reporting species newly reported in a region and reduced the importance of well-studied high impact NIS (Pyšek et al. 2008). The resulting 436 publications were screened for NIS reports, and 55 were used to assemble our global list (Appendix 1 of ESM). In addition to NIS recovered by Thomson's ISI search, we included species listed in the Global Invasive Species Database of the Invasive Species Specialist Group (ISSG 2010). To reduce geographical bias, we did not include species from regional data sets such as Delivering Alien Invasive Species Inventories for Europe (DAISIE) or Great Lakes Aquatic Nonindigenous Information System (GLANSIS) (Pyšek et al. 2008). Bacteria, virus-like particles and fungi were excluded from our list because these taxa typically have uncertain status as non-indigenous or native. After the list was assembled, the recorded species were assigned to kingdom, phylum, and class by consulting

several taxonomic websites [e.g. BOLD, the European Nature Information System (EUNIS), World Register of Marine Species (WORMS), ZipcodeZoo].

To determine the potential for molecular identification of NIS, we searched BOLD (http://www.boldsystems.org/) and GenBank (http://www.ncbi.nlm.nih.gov/genbank/) for COI, 16S, small subunit ribosomal 18S rDNA (18S), ITS, rbcL and matK gene sequences. To examine the incidence of sequence deposition to genetic databases, we assessed both genetic databases three times: from May to September 2010, from June to August 2012, and in January 2016. In 2010 and 2012, BOLD was assessed only for COI sequences as in these years it contained very few ITS, rbcL or matK, and no 16S or 18S sequences; in 2016, it was assessed for all six genome regions. GenBank was assessed for all six genome regions each time. To determine the rate of sequence deposition to genetic databases, a series of regression analyses were conducted with total number of species with at least one sequence in at least one genetic database as the dependent variables and time as the independent variable. Additionally, to compare the trend of deposition of sequences of NIS on our list to general deposition of sequences to BOLD irrespective of indigenous/non-indigenous status, regression analysis for BOLD with all species in BOLD with at least one sequence as the dependent variable and time as the independent variable was conducted as well (consulted 17 February 2016).

Finally, to explore if some classes (hereafter class/es is used in the systematic sense) of NIS were more or less represented in genetic databases than was the average for taxa within its particular habitat (i.e. aquatic or terrestrial) in the years we examined (i.e. 2010, 2012, and 2016), we constructed scatter plots with number of NIS per class on the x-axis and number of NIS with at least one sequence in at least one genetic database per class on the y-axis; the line of unity was based on the average percentage of NIS with at least one sequence in at least one genetic database. Six different scatter plots and lines of unity were constructed: for aquatic taxa in 2010, 2012, and 2016, and for terrestrial taxa in 2010, 2012, and 2016. Values were log transformed to standardize the data. Primary dataset containing the list of aquatic and terrestrial NIS, their taxonomic determination, and availability of sequences in 2010, 2012 and 2016 is available at: doi:10.1594/PANGAEA.859211.

Results

Aquatic and terrestrial NIS

Our Thomson's ISI Web of Knowledge search identified 3101 NIS, of which 1383 (45 %) were aquatic and 1718 (55 %) terrestrial (Fig. 1; Appendix 2 of ESM). Aquatic taxa belonged to four kingdoms: Animalia (71 %), Chromista (5 %), Plantae (21 %) and Protozoa (3 %), consisting of 26 phyla (Figs. 1, 2; Appendix 2 of ESM). The most prevalent aquatic phyla were Annelida (10 %), Arthropoda (26 %), Chordata (30 %) and Mollusca (18 %) in kingdom Animalia, Ochrophyta (96 %) in kingdom Chromista, and Chlorophyta (15 %), Rodophyta (40 %) and Tracheophyta (44 %) in kingdom Plantae. Protozoa was represented by the lowest number of species. When the most dominant Animalia phyla were explored deeper, Malacostraca and Maxillopoda were revealed as the richest Arthropoda classes, Actinopterygii as richest Chordata class, and Bivalvia and Gastopoda as richest Mollusca classes (Appendices 2 and 3 of ESM). In the case of aquatic Plantae, Ulvophyceae and Florideophyceae were dominant classes within Chlorophyta and Rodophyta kingdoms, respectively (Appendices 2 and 3 of ESM).

Terrestrial taxa belonged to two kingdoms: Animalia (22 %; having six phyla) and Plantae (78 %; one phylum) (Figs. 1, 3; Appendix 2 of ESM). Arthropoda (68 %) and Chordata (25 %) were the most prevalent Animalia phyla; however, Tracheophyta in Plantae phylum, represented by 1333 species (100 % of terrestrial Plantae), was the most prevalent phylum in both aquatic and terrestrial habitats (Figs. 2, 3;

Appendix 2 of ESM). Deeper analyses of terrestrial phyla revealed Insecta as the richest Arthropoda class, and Aves and Mammalia as richest Chordata classes. Liliopsida and Magnoliopsida were the richest Tracheophyta classes (Appendices 2 and 4 of ESM).

Sequence availability in 2010

Eight hundred ninety-five out of 1383 aquatic NIS (65 %) were characterized by at least one sequence (COI, 16S, 18S, ITS, rbcL or matK) in at least one genetic database. All four aquatic kingdoms were similarly represented in the genetic databases; 64, 65, 69, and 61 % of NIS of Animalia, Chromista, Plantae and Protozoa, respectively (Fig. 1; Appendix 2 of ESM). Of 13 Animalia phyla, coverage for ten phyla ranged from 50 to 79 % of NIS; Ctenophora was 100 % covered, while Porifera and Rotifera were 21 and 0 % covered, respectively (Fig. 2; Appendix 2 of ESM). In Chromista phylum, only Ochrophyta had sequences in the genetic databases (68 %), while coverage for Plantae and Protozoa phyla were mixed, ranging from 0 to 100 % (Fig. 2; Appendix 2 of ESM). The majority of aquatic classes were around the average (i.e. 65 %), though twelve classes were not covered at all (Holothuroidea, Turbellaria, Monogononta, Prymnesiophyceae, Labyrinthulomycetes, Xanthophyceae, Marchantiopsida, Compsopogonophyceae, Gromiidea, Ciliatea, Oligohymenophorea, Kinetoplastea; Fig. 4; Appendix 2 of ESM). Classes of the most species-abundant aquatic Animalia and Plantae phyla (i.e. Arthropoda, Chordata, Mollusca, Chlorophyta, Rodophyta, and Tracheophyta) revealed relatively equal sequence representation; most of the

Fig. 1 Number of non-indigenous species (NIS) per kingdom, and number of NIS with at least one sequence in at least one genetic database in 2010, 2012 and 2016 for aquatic and terrestrial taxa. Percentage cover for 2010, 2012 and 2016 are shown in *brackets*, respectively
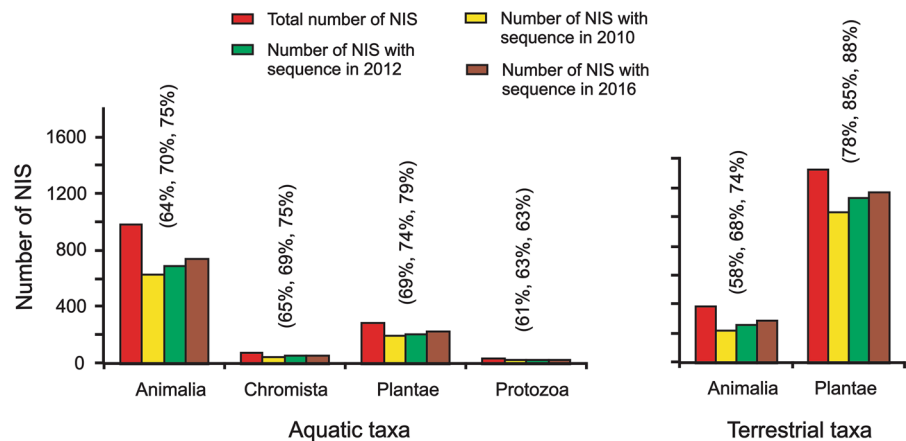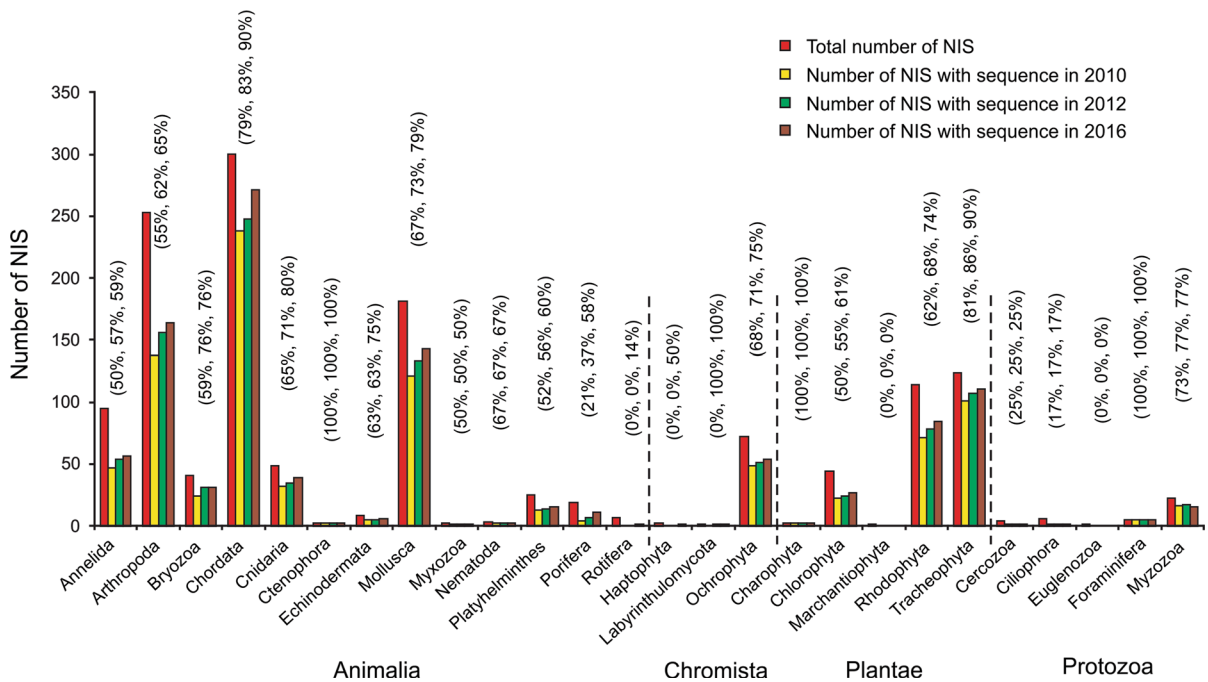
Fig. 2 Number of non-indigenous species (NIS) per phylum, and number of NIS with at least one sequence in at least one genetic database in 2010, 2012 and 2016 for aquatic taxa. Percentage cover for 2010, 2012 and 2016 are shown in *brackets*, respectively

classes' coverage ranged between 50 and 100 % (Appendices 2 and 3 of ESM).

In 2010, out of 1718 terrestrial NIS, 1256 (73 %) were covered by at least one sequence in at least one genetic database (58 % of Animalia and 78 % of Plantae; Fig. 1; Appendix 2 of ESM). Animalia phyla's coverage ranged from 46 to 100 %, though Mollusca had no sequences in the databases (Fig. 3; Appendix 2 of ESM). Tracheophyta, the only Plantae phylum, was covered for 78 % of species (Fig. 3; Appendix 2 of ESM). The majority of terrestrial classes were around the average (i.e. 73 %), though two classes (Chilopoda and Gastropoda) were not covered at all, and Arachnida was very poorly represented (Fig. 4; Appendix 2 of ESM). Coverage for classes of the most species-abundant terrestrial Animalia and Plantae phyla (i.e. Arthropoda, Chordata, and Tracheophyta) were similar to those for aquatic phyla, with most class coverages ranging between 60 and 100 % (Appendices 2 and 4 of ESM).

Sequence availability in 2012

Two years later, 71 % of aquatic NIS were represented in the databases; the number of sequences increased to

70, 69, 74 and 63 % for Animalia, Chromista, Plantae, and Protozoa, respectively (Fig. 1; Appendix 2 of ESM). Out of 13 Animalia phyla, new sequences were available for eight phyla (i.e. Annelida, Arthropoda, Bryozoa, Chordata, Cnidaria, Mollusca, Platyhelminthes, and Porifera; Fig. 2; Appendix 2 of ESM). Sequences for two Chromista, three Plantae and one Protozoa phyla also increased (Fig. 2; Appendix 2 of ESM). Representation of most classes was around the average (i.e. 70 %); eleven classes were still not covered at all (Holothuroidea, Turbellaria, Monogononta, Prymnesiophyceae, Xanthophyceae, Marchantiopsida, Compsopogonophyceae, Gromiidea, Ciliatea, Oligohymenophorea, and Kinetoplastea; Fig. 4; Appendix 2 of ESM). Sequence coverage of terrestrial taxa was 81 % in 2012. The number of sequences increased to 68 and 85 % for Animalia and Plantae, respectively (Fig. 1; Appendix 2 of ESM). Out of five Animalia phyla, new sequences were added for three phyla (i.e. Annelida, Arthropoda, and Chordata; Fig. 3; Appendix 2 of ESM). Coverage of Tracheophyta increased to 85 % (Fig. 3; Appendix 2 of ESM). Coverage for the majority of classes was again around the average (i.e. 81 %). Two classes were still not covered (Chilopoda and Gastropoda), as
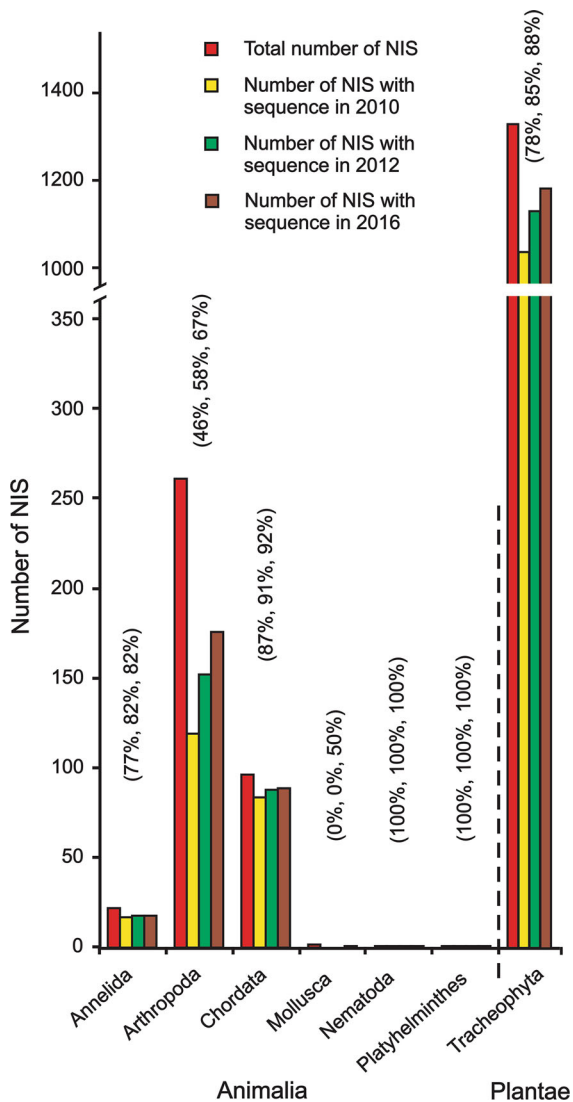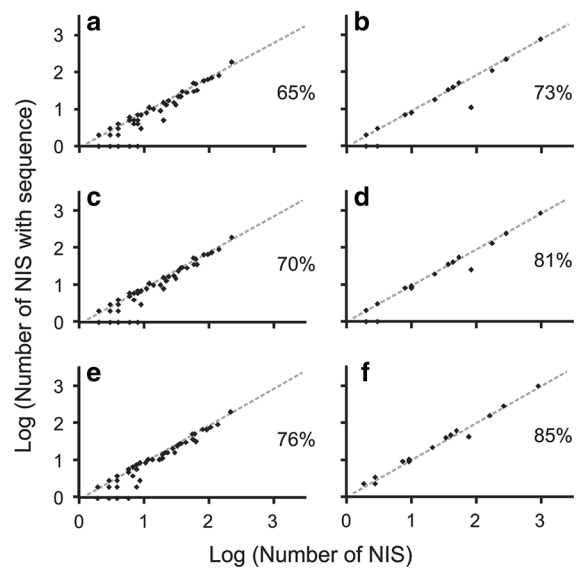
Fig. 3 Number of non-indigenous species (NIS) per phylum, and number of NIS with at least one sequence in at least one genetic database in 2010, 2012 and 2016 for terrestrial taxa. Percentage cover for 2010, 2012 and 2016 are shown in *brackets*, respectively

Fig. 4 Scatter plots with number of NIS per class on *x-axis* and number of NIS with at least one sequence in at least one genetic database per class on *y-axis* for aquatic taxa in 2010 (**a**), terrestrial taxa in 2010 (**b**), aquatic taxa in 2012 (**c**), terrestrial taxa in 2012 (**d**), aquatic taxa in 2016 (**e**), and terrestrial taxa in 2016 (**f**). The lines of unity were based on the average percentage of NIS with at least one sequence in at least one genetic database for aquatic taxa in 2010 (**a**), terrestrial taxa in 2010 (**b**), aquatic taxa in 2012 (**c**), terrestrial taxa in 2012 (**d**), aquatic taxa in 2016 (**e**), and terrestrial taxa in 2016 (**f**). Values are log transformed to standardize the data. The average percentages are given for each panel

well as Arachnida being less covered than the average (Fig. 4; Appendix 2 of ESM).

Sequence availability in 2016

In January 2016, 1047 aquatic NIS (76 %) were represented in the databases; the number of species with at least one sequence increased to 743 (75 %), 56 (75 %) and 224 (74 %) for Animalia, Chromista and Plantae, respectively (Fig. 1; Appendix 2 of ESM). No

new Protozoa species were covered after 2012 (Fig. 1; Appendix 2 of ESM). New sequences were available for nine Animalia phyla (i.e. Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Platy-helminthes, Porifera, and Rotifera; Fig. 2; Appendix 2 of ESM). Sequences for two Chromista and three Plantae phyla also increased (Fig. 2; Appendix 2 of ESM). Representation of most classes was around the average (i.e. 76 %); eight classes were still not covered at all (Turbellaria, Xanthophyceae, Marchan-tiopsida, Compsopogonophyceae, Gromiidea, Cili-atea, Oligohymenophorea, and Kinetoplastea; Fig. 4; Appendix 2 of ESM). Sequence coverage of terrestrial taxa was 85 % (Fig. 1; Appendix 2 of ESM). The number of sequences increased to 74 and 88 % for Animalia and Plantae, respectively (Fig. 1; Appendix 2 of ESM). Out of five Animalia phyla, new sequences were added for three phyla (i.e. Arthropoda, Chordata, and Mollusca; Fig. 3; Appendix 2 of ESM). Coverage of Tracheophyta increased to 88 % (Fig. 3; Appendix

2 of ESM). Coverage for the majority of classes was again around the average (i.e. 85 %; Fig. 4; Appendix 2 of ESM).

Regression analyses revealed no significant increase for either total number of species covered by at least one sequence in at least one database from our NIS list, or for aquatic or terrestrial taxa from our list through time ($P > 0.05$; Fig. 5a). The increase of species with at least one sequence in BOLD independently of indigenous/non-indigenous status was highly significant ($P < 0.05$; Fig. 5b). On average 56 new NIS from our list were covered by at least one sequence per year, while on average sequences for 19,599 new species are entered in BOLD each year (Fig. 5).

Sequence availability for two or more genes per species

When availability of sequences for two or three genes per species were checked, the species coverage for aquatic taxa dropped from 65 % species covered by at least one sequence in at least one database to 49 % species covered by sequences of at least two genes and to 32 % species covered by sequences of at least three genes, in 2010 (Table 1). The coverage of terrestrial taxa dropped from 78 to 56 (two genes) and 33 % (three genes) in 2010 (Table 1). As more sequences were added to the genetic databases through time, the difference between at least one sequence per species and at least two or three sequences per species declined. The species coverage in 2012 dropped from 71 to 56 (two genes) and 41 % (three genes) for aquatic taxa, and from 85 to 75 (two genes) and 61 % (three genes) for terrestrial taxa, respectively (Table 1). The drop in 2016 was from 76 to 66 and 54 % for aquatic taxa, and from 88 to 85 and 79 % for terrestrial taxa for two and three genes per species, respectively (Table 1).

Discussion

Availability of sequences for DNA barcoding

As two-thirds of NIS studied in Web of Science are plants and insects (Pyšek et al. 2008), many ecological hypotheses and theories were tested on plants (Blossey and Nötzold 1995; Davis et al. 2000; Minchinton 2002; Keane and Crawley 2002; Mitchell and Power 2003; Richardson and Pyšek 2006). As it is also easier to manipulate experimental design and to conduct experiments and monitoring programs for terrestrial than for aquatic taxa, one might expect that terrestrial taxa would be more extensively studied and consequently better represented by DNA sequences than
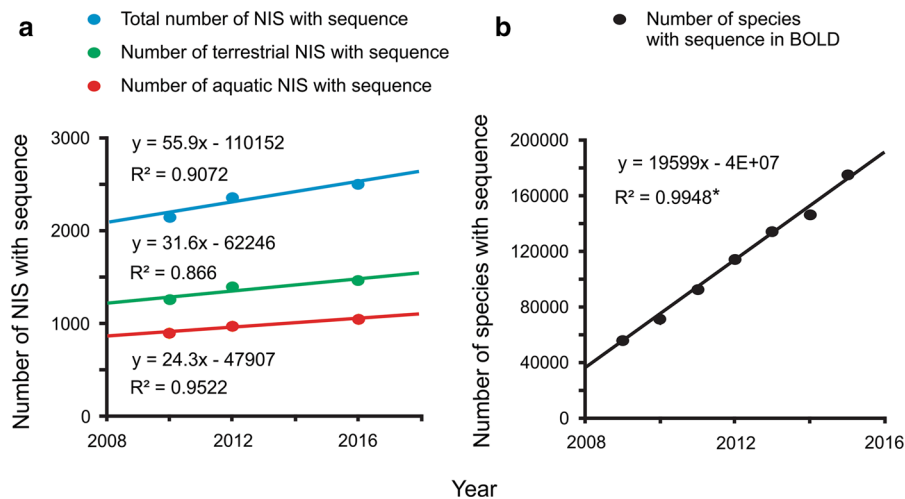
Fig. 5 Scatterplot and fitted regression lines with total number of species with at least one sequence in at least one genetic database as the dependent variables and time as the independent variable for all, terrestrial, and aquatic taxa in our study (**a**), and scatterplot and fitted regression line with all species in Barcode of Life Database (BOLD) with at least one sequence as the dependent variable and time as the independent variable (BOLD 2016) (**b**). An *asterisk* denotes significant difference ($P < 0.05$)

Table 1 Number (#) of species with at least one sequence, at least two sequences, and at least three sequences, in at least one genetic database in 2010, 2012 and 2016 for aquatic and terrestrial taxa

	2010						2012						2016					
	At least one sequence		At least two sequences		At least three sequences		At least one sequence		At least two sequences		At least three sequences		At least one sequence		At least two sequences		At least three sequences	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
Aquatic taxa	895	**65**	672	**49**	430	**32**	975	**71**	780	**56**	567	**41**	1047	**76**	916	**66**	748	**54**
Animalia	627	**64**	483	**49**	316	**32**	688	**70**	550	**56**	400	**41**	743	**75**	650	**66**	524	**53**
Chromista	49	**65**	33	**44**	24	**32**	52	**69**	39	**52**	27	**36**	56	**75**	50	**67**	35	**47**
Plantae	196	**69**	141	**50**	79	**28**	211	**74**	174	**61**	129	**45**	224	**79**	199	**70**	174	**61**
Protozoa	23	**61**	15	**40**	11	**29**	24	**63**	17	**45**	11	**29**	24	**63**	17	**45**	15	**40**
Terrestrial taxa	1256	**73**	914	**52**	530	**31**	1391	**81**	1190	**69**	927	**54**	1460	**85**	1362	**79**	1215	**71**
Animalia	223	**58**	164	**43**	97	**25**	261	**68**	194	**50**	120	**31**	286	**74**	233	**60**	164	**43**
Plantae	1033	**78**	750	**56**	433	**33**	1130	**85**	996	**75**	807	**61**	1174	**88**	1129	**85**	1051	**79**
Total	2151	**69**	1586	**51**	960	**31**	2366	**76**	1970	**64**	1494	**48**	2507	**81**	2278	**74**	1963	**63**

Percentage (%) cover for 2010, 2012 and 2016 are shown in bold

aquatic taxa. Our study demonstrated, however, that there is little difference between the two. Approximately 75 % of species in almost each aquatic kingdom had at least one sequence in at least one genetic database. Only the coverage of aquatic Protozoa was lower (63 %). Similar coverage was available for terrestrial Animalia while terrestrial Plantae were better covered (88 %). Interestingly, our findings were contrary to the findings of Pyšek et al. (2008) who stated that plant NIS are slightly understudied in the general ecological literature compared to other taxa when number of NIS per taxonomic group has been compared to number of studies per taxonomic group. The same authors found that insects, birds, and reptiles are mildly understudied while crustaceans, molluscs, algae, and mammals are more intensively studied (Pyšek et al. 2008). Our examination of sequence availability is mainly in agreement with Pyšek et al. (2008), though there are some discrepancies. We determined that insect sequence availability was slightly lower than average in both aquatic and terrestrial habitats (59 and 78 %, respectively), while birds and reptiles were better covered (78–100 %). The discrepancy between Pyšek et al. (2008) and our sequence availability results demonstrates that intensity of ecological invasion studies is not clearly correlated to intensity of molecular studies of the

same taxa. Encouragingly, some taxonomic groups are mildly understudied in invasion ecology but are well represented in molecular studies with many gene sequences. The opposite pattern has also been observed, however, with more markedly understudied aquatic than terrestrial taxa, particularly those belonging to Chromista and Protozoa kingdoms.

Deposition of sequences to genetic databases

Between 2010 and 2016, species coverage by DNA sequences increased from 65 and 73 % to 76 and 85 % for aquatic and terrestrial taxa, respectively. Assuming that deposition of sequences to the databases follows a linear function, we expect a reasonably brief period (until 2024) before the majority of terrestrial NIS on our list are sequenced, and a slightly more protracted timeframe (until 2030) before the majority of aquatic NIS are likewise surveyed. We cannot confidently demonstrate that the trend is linear since we have only three time points. The regression analyses determined no significant increase in the number of NIS covered, though deposition of sequences to BOLD irrespective of indigenous/non-indigenous status follows a significant linear trend. As more than three-quarters of NIS on our list are already covered, an optimistic explanation for the lack of a significant increase in NIS

coverage may be that the function is saturating and starting to level out. If this is the case, the increase might be significant and much steeper in the period before 2010 than in the last 6 years. However, our list of NIS is not exhaustive, particularly due to uncertainties associated with the status of cryptogenic species, as well as continuous discoveries of new NIS. Bearing in mind that we used the list of NIS assembled in 2010, and did not update it in the consequent years when genetic databases were checked (i.e., 2012 and 2016), it is possible that the rate of increase in NIS coverage is closer to that of total species (irrespective of indigenous/non-indigenous status) in the BOLD than shown by our saturation rates. Furthermore, taking into account the rapid development of molecular techniques and technology, in the near future one may expect the deposition of sequences to follow an exponential rather than linear function. In particular, this might be true for NIS taxa, as studies on invasive species have been rapidly increasing since 1990 (Ricciardi and MacIsaac 2008). In addition, the number of studies of NIS with economic value, such as fishes (e.g. *Cyprinus carpio*, *Salmo trutta*, and *Oncorhynchus mykiss*) and mammals (e.g. *Sus scrofa*), and NIS having severe impact on environment and economy [e.g. *Rattus rattus*, *Dreissena polymorpha*, and *Eichhornia crassipes*; see also Briski et al. (2011) and Trebitz et al. (2015)] is exceptionally high compared to studies of other NIS (MacIsaac et al. 2011). In this study, taxa such as aquatic Malacostraca (many species with environmental or economic impact), Maxillopoda, Bivalvia, and Ulvaceae (many species of economic value and/or causing impact) and terrestrial Insecta (many species causing environmental or economic impact) demonstrate an exceptionally high trend of sequence deposition. Consequently, while there does not appear to be a strong difference in sequence enrichment between aquatic and terrestrial taxa, we may expect that NIS belonging to particular taxonomic groups would be more rapidly described by gene sequences suitable for DNA barcoding than other species.

Perspectives on DNA barcoding for detecting NIS

On average 81 % of NIS were covered by sequences in genetic databases, with terrestrial, and in particular plant taxa, having the best coverage. Most taxonomic classes are covered relatively well, though there are still some taxa not covered at all. Our list of NIS is not exhaustive, and many species which are not reported as NIS today may become NIS in the future. So, as long as most of the world biodiversity is not sequenced, we may expect introductions of species that cannot be identified by DNA barcoding. Furthermore, nuclear pseudogenes, heteroplasmy, hybrid introgression, and mitochondrial and plasmid inheritance modes may also reduce the efficiency of DNA barcoding (Hebert et al. 2004; Buhay 2009; Galtier et al. 2009; Hollingsworth et al. 2011; Comtet et al. 2015). Still, the prospect of DNA barcodes for detection and identification of NIS is more promising than traditional morphological identifications. Beside numerous problems connected to morphological identification, taxonomic experts capable to conduct morphological identification are becoming rare, with some taxonomic groups not covered by experts at all (Segers 2008; Ojaveer et al. 2014).

Metabarcoding, which provides millions of sequences from bulk samples, and its application as an environmental DNA (eDNA) monitoring technique that obtains genetic material directly from environmental samples (e.g. water, sediment, and soil) without any obvious signs of biological source material, provides new approaches to population and biodiversity monitoring (Ficetola et al. 2008; Comtet et al. 2015; Goldberg et al. 2015; Thomsen and Willerslev 2015), and invasion ecologists are already developing and adjusting these techniques for early detection of notorious NIS (Turner et al. 2014; Wilson et al. 2014). Use of metabarcoding and multiple markers are expected to increase identification rates, although at least initially, those techniques would increase work- and cost-loads, particularly since there are still developmental technical problems (Zhan et al. 2014a, b; Comtet et al. 2015). Continued enrichment of genetic databases will be required for the effective use of these techniques, including concerted efforts to sequence genes for under-represented groups, irrespective of their economic value or environmental and/or economic impact. In this process, correct species determination (by traditional taxonomy) and proper management of sequence deposition and voucher storage is vital to preserve connections between morphological and molecular data.

References

Armstrong KF, Bell SL (2005) DNA barcodes for biosecurity: invasive species identification. Philos Trans R Soc B 360:1813–1823. doi:10.1098/rstb.2005.1713

Ayres DR, Zaremba K, Sloop CM, Strong DR (2008) Sexual reproduction of cordgrass hybrids (*Spartina foliosa alterniflora*) invading tidal marshes in San Francisco Bay. Divers Distrib 14:187–195. doi:10.1111/j.1472-4642.2007.00414.x

Bax N, Carlton JT, Mathews-Amos A, Haedrich RL, Howarth FG, Purcell JE, Rieser A, Gray A (2001) The control of biological invasions in the world's oceans. Conserv Biol 15:1234–1246. doi:10.1111/j.1523-1739.2001.99487.x

Blossey B, Nötzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. J Ecol 83:887–889. doi:10.2307/2261425

Briski E, Cristescu ME, Bailey SA, MacIsaac HJ (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. Biol Invasions 13:1325–1340. doi:10.1007/s10530-010-9892-7

Briski E, Ghabooli S, Bailey SA, MacIsaac HJ (2012a) Invasion risk posed by macroinvertebrates transported in ships' ballast tanks. Biol Invasions 14:1843–1850. doi:10.1007/s10530-012-0194-0

Briski E, Wiley CJ, Bailey SA (2012b) Role of domestic shipping in the introduction or secondary spread of nonindigenous species: biological invasions within the Laurentian Great Lakes. J Appl Ecol 49:1124–1130. doi:10.1111/j.1365-2664.2012.02186.x

Briski E, Bailey SA, Casas-Monroy O, DiBacco C, Kaczmarska I, Lawrence JE, Leichsenring J, Levings C, MacGillivary ML, McKindsey CW, Nasmith LE, Parenteau M, Piercey GE, Rivkin RB, Rochon A, Roy S, Simard N, Sun B, Way C, Weise AM, MacIsaac HJ (2013) Taxon- and vector-specific variation in species richness and abundance during the transport stage of biological invasions. Limnol Oceanogr 58:1361–1372. doi:10.4319/lo.2013.58.4.1361

Buhay JE (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. J Crust Biol 29:96–110. doi:10.1651/08-3020.1

Carlton JT (2009) Deep invasion ecology and the assembly of communities in historical time. In: Rilov G, Crooks JA (eds) Biological invasions in marine ecosystems: ecological, management, and geographic perspectives, 1st edn. Springer, Berlin, pp 13–56

Chown S, Sinclair B, Vuuren B (2008) DNA barcoding and the documentation of alien species establishment on sub-Antarctic Marion Island. Polar Biol 31:651–655. doi:10.1007/s00300-007-0402-z

Colautti RI, MacIsaac HJ (2004) A neutral terminology for defining invasive species. Divers Distrib 10:135–141. doi:10.1111/j.1366-9516.2004.00061.x

Comtet T, Sandionigi A, Viard F, Casiraghi M (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. Biol Invsions 17:905–922. doi:10.1007/s10530-015-0854-y

Conn JS, Stockdale CA, Werdin-Pfisterer NR, Morgan JC (2010) Characterizing Pathways of Invasive Plant Spread to Alaska: II. Propagules from imported hay and straw. Invasive Plant Sci Manag 3:276–285. doi:10.1614/IPSM-D-09-00041.1

Cross HB, Lowe AJ, Gurgel FD (2010) DNA barcoding of invasive species. In: Richardson DM (ed) Fifty years of invasion ecology: the legacy of Charles Elton, 1st edn. Blackwell, Oxford, pp 289–300

Darling JA, Blum MJ (2007) DNA-based methods for monitoring invasive species: a review and prospectus. Biol Invasions 9:751–765. doi:10.1007/s10530-006-9079-4

Davis M, Grime JP, Thompson K (2000) Fluctuating resources in plant communities: a general theory of invisibility. J Ecol 88:528–534. doi:10.1046/j.1365-2745.2000.00473.x

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biol Lett 4:423–425. doi:10.1098/rsbl.2008.0118

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol 18:4541–4550. doi:10.1111/j.1365-294X.2009.04380.x

Goldberg CS, Strickler KM, Pilliod DS (2015) Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biol Conserv 183:1–3. doi:10.1016/j.biocon.2014.11.040

Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol 210:1518–1525. doi:10.1242/jeb.001370

Hebert PDN, Gregory T (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54:852–859. doi:10.1080/10635150500354886

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc B Biol Sci 270:313–321. doi:10.1098/rspb.2002.2218

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:1657–1663. doi:10.1371/journal.pbio.0020312

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bankd M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneideri H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacon J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR,

Kim JS, Kim Y-D, Lahaye R, Lee H-L, Long DG, Madrinan S, Maurin O, Meusnier I, Newmaster SG, Park C-W, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi D-K, Little DP (2009) A DNA barcode for land plants. Proc Natl Acad Sci 106:12794–12797. doi:10.1073/pnas.0905845106

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6:e19254. doi:10.1371/journal.pone.0019254

Hulme PE, Bacher S, Kenis M, Klotz S, Kühn I, Minchin D, Nentwig W, Olenin S, Panov V, Pergl J, Pyšek P, Roques A, Sol D, Solarz W, Vilà M (2008) Grasping at the routes of biological invasions: a framework for integrating pathways into policy. J Appl Ecol 45:403–414. doi:10.1111/j.1365-2664.2007.01442.x

Invasive Species Specialist Group (2010). Global invasive species database. http://www.issg.org/database/species/List.asp. Accessed 10 June 2010

Janzen DH (2004) Now is the time. Philos Trans R Soc B 359:731–732. doi:10.1098/rstb.2003.1444

Keane RM, Crawley MJ (2002) Exotic plant invasions and enemy release hypothesis. Trends Ecol Evol 17:164–170. doi:10.1016/S0169-5347(02)02499-0

Kelly DW, Paterson RA, Townsend CR, Poulin R, Tompkins DM (2009) Parasite spillback: a neglected concept in invasion ecology? Ecology 90:2047–2056. doi:10.1890/08-1085.1

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. Trends Ecol Evol 16:199–204. doi:10.1016/S0169-5347(01)02101-2

Kolbe JJ, Glor RE, Rodriguez-Schettino L, Chamizo-Lara A, Larson A, Losos JB (2007) Multiple sources, admixture, and genetic variation in introduced Anolis lizard populations. Conserv Biol 21:1612–1625. doi:10.1111/j.1523-1739.2007.00826.x

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci 105:8369–8374. doi:10.1073/pnas.0503123102

Lockwood JL, Hoopes MF, Marchetti MP (2007) Invasion ecology. Blackwell, Oxford

Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, Carlton JT, McMichael A (2006) Biological invasions: recommendations for US policy and management. Ecol Appl 16:2035–2054. doi:10.1890/1051-0761

MacIsaac HJ, Tedla R, Ricciardi A (2011) Patterns and rate of growth of studies in invasion ecology. In: Richardson RM (ed) Fifty years of invasion ecology—the legacy of Charles Elton, 1st edn. Blackwell, Oxford, pp 51–60

Minchinton TE (2002) Disturbance by wrack facilitates spread of Phragmites australis in a coastal marsh. J Exp Mar Biol Ecol 281:89–107. doi:10.1016/S0022-0981(02)00438-0

Mitchell CE, Power AG (2003) Release of invasive plants from fungal and viral pathogens. Nature 241:625–627. doi:10.1038/nature01317

National Center for Biotechnology Information (NCBI) (2015) http://www.ncbi.nlm.nih.gov/genbank/. Accessed 15 Feb 2015

Ojaveer H, Galil BS, Minchin D, Olenin S, Amorim A, Canning-Clode J, Chainho P, Copp GH, Gollasch S, Jelmert A,

Lehtiniemi M, McKenzie C, Mikuš J, Miossec L, Occhipinti-Ambrogi A, Pećarević M, Pederson J, Quilez-Badia G, Wijsman JWM, Zenetos A (2014) Ten recommendations for advancing the assessment and management of non-indigenous species in marine ecosystems. Mar Policy 44:160–165. doi:10.1016/j.marpol.2013.08.019

Pyšek P, Richardson DM, Pergl J, Jarošík V, Sixtová Z, Weber E (2008) Geographical and taxonomic biases in invasion ecology. Trends Ecol Evol 23:237–244. doi:10.1016/j.tree.2008.02.002

Radford AE, Ahles HE, Bell CR (1968) Manual of the vascular flora of the Carolinas. The University of North Carolina Press, Chapel Hill

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). Mol Ecol Notes 7:355–364. doi:10.1111/j.1471-8286.2007.01678.x

Ricciardi A, MacIsaac HJ (2008) The book that began invasion ecology. Nature 452:34. doi:10.1038/452034a

Richardson DM, Pyšek P (2006) Plant invasions: merging the concepts of species invasiveness and community invisibility. Prog Phys Geogr 30:409–431. doi:10.1191/0309133306pp490pr

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. Trends Ecol Evol 24:192–200. doi:10.1016/j.tree.2008.11.004

Segers H (2008) Global diversity of rotifers (Rotifera) in freshwater. Hydrobiologia 595:49–59. doi:10.1007/978-1-4020-8259-7_6

Sephton D, Vercaemer B, Nicolas JM, Keays J (2011) Monitoring for invasive tunicates in Nova Scotia, Canada (2006–2009). Aquat Invasions 6:391–403. doi:10.3391/ai.2011.6.4.04

Shearer TL, Coffroth MA (2008) Barcoding corals: limited by interspecific divergence, not intraspecific variation. Mol Ecol Resour 8:247–255. doi:10.1111/j.1471-8286.2007.01996.x

Simberloff D (2009) We can eliminate invasions or live with them. Successful management projects. Biol Invasions 11:149–157. doi:10.1007/s10530-008-9317-z

Thomsen PF, Willerslev E (2015) Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18. doi:10.1016/j.biocon.2014.11.019

Trebitz AS, Hoffman JC, Grant GW, Billehus TM, Pilgrim EM (2015) Potential for DNA-based identification of Great Lakes fauna: match and mismatch between taxa inventories and DNA barcode libraries. Sci Rep 5:12162. doi:10.1038/srep12162

Turner CR, Miller DJ, Coyne KJ, Corush J (2014) Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian Bigheaded Carp (Hypophthalmichthys spp.). PLoS ONE 9(12):e114329. doi:10.1371/journal.pone.0114329

Wilson C, Wright E, Bronnenhuber J, MacDonald F, Belore M, Locke B (2014) Tracking ghosts: combined electrofishing and environmental DNA surveillance efforts for Asian carps in Ontario waters of Lake Erie. Manag Biol Invasions 5:225–231. doi:10.3391/mbi.2014.5.3.05

Zhan A, MacIsaac HJ (2015) Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. Conserv Genet 16:513–522. doi:10.1007/s10592-014-0678-9

Zhan A, Hulak M, Sylvester F, Huang X, Adebayo A, Abbott CL, Adamowicz SJ, Heath DD, Cristescu ME, MacIsaac HJ (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol Evol 4:558–565. doi:10.1111/2041-210X.12037

Zhan A, Xiong W, He S, MacIsaac HJ (2014a) Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing. PLoS ONE 9:e96928. doi:10.1371/journal.pone.0096928

Zhan A, Bailey SA, Heath DD, MacIsaac HJ (2014b) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. Mol Ecol Resour 14:1049–1059. doi:10.1111/1755-0998.12254