Diversity and Distributions, (Diversity Distrib.) (2016) 22, 1045–1059

BIODIVERSITY RESEARCH

Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports

Emily A. Brown[1,2]*, Frédéric J. J. Chain[1], Aibin Zhan[3], Hugh J. MacIsaac[2] and Melania E. Cristescu[1]

[1]Department of Biology, McGill University, 1205 Docteur Penfield, Montreal, QC H3A 1B1, Canada, [2]Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON N9B 3P4, Canada, [3]Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Haidian District, Beijing 100085, China

*Correspondence: Emily A. Brown, Department of Biology, McGill University, 1205 Docteur Penfield, Montreal, QC H3A 1B1, Canada.
E-mail: emily.angharad@gmail.com

ABSTRACT

Aim Invasive species represent one of the greatest threats to biodiversity. The ability to detect non-indigenous species (NIS), particularly those present at low abundance, is limited by difficulties in performing exhaustive sampling and in identifying species. Here we sample zooplankton from 16 major Canadian ports and apply a metabarcoding approach to detect NIS.

Location Marine and freshwater ports along Canadian coastlines (Pacific, Arctic, Atlantic) and the Great Lakes.

Methods We amplified the V4 region of the small subunit ribosomal DNA (18S) and used two distinct analytic protocols to identify species present at low abundance. Taxonomic assignment was conducted using BLAST searches against a local 18S sequence database of either (i) individual reads (totalling 7,733,541 reads) or (ii) operational taxonomic units (OTUs) generated by sequence clustering. Phylogenetic analyses were performed to confirm the identity of reads with ambiguous taxonomic assignment.

Results Taxonomic assignment of individual reads identified 379 zooplankton species at a minimum sequence identity of 97%. Of these, 24 species were identified as NIS, 11 of which were detected in previously unreported locations. When reads were clustered into OTUs prior to taxonomic assignment, six NIS were no longer detected and an additional NIS was falsely identified. Phylogenetic analyses revealed that sequences belonging to closely related species clustered together into shared OTUs as a result of low interspecific variation. NIS can thus be misidentified when their sequences join the OTUs of more abundant native species.

Main conclusions Our results reveal the power of the metabarcoding approach, whilst also highlighting the need to account for potentially low levels of genetic diversity when processing data, to use barcode markers that allow differentiation of closely related species and to continue building comprehensive sequence databases that allow reliable and fine-scale taxonomic designation.

Keywords
18S, biodiversity, biomonitoring, high-throughput sequencing, invasive species, metabarcoding, operational taxonomic unit.

INTRODUCTION

Invasive species are recognized as a significant global threat. The introduction and spread of non-indigenous species (NIS) into novel environments can result in declines in local biodiversity and ecosystem function (Molnar *et al.*, 2008; Pejchar & Mooney, 2009), which can in turn lead to huge economic losses (e.g. Pimentel *et al.*, 2000, 2005; Molnar

DOI: 10.1111/ddi.12465
http://wileyonlinelibrary.com/journal/ddi

A Journal of Conservation Biogeography
Diversity and Distributions

et al., 2008). Continuing climate change, habitat alteration and species exploitation, along with the accelerated movement of species across the globe, have left ecosystems particularly vulnerable to invasion (Brook *et al.*, 2008). Rigorous monitoring programmes that allow early detection of NIS are recommended as a priority strategy for conservation and management efforts (Lodge *et al.*, 2006; Vander Zanden *et al.*, 2010). Traditional methods that rely on visual identification of specimens have been criticized for their poor ability to identify juvenile life stages that may be critically important in the establishment and spread of invasive populations, and for their limited taxonomic resolution in many taxa (Caesar *et al.*, 2006). Moreover, typical sampling protocols have been shown to have a low probability (< 0.2) of detecting species unless population density is high (Harvey *et al.*, 2009), making sampling reliable only for species that are moderately to highly abundant (Cao *et al.*, 1998; Jerde *et al.*, 2011). NIS may therefore not be detected until they have established large populations and/or spread (Crooks & Soulé, 1999).

Given the difficulties associated with traditional methods of identifying NIS, interest has arisen in developing tools with greater detection probabilities (McDonald, 2004). Molecular identification of species, through the analysis of a small fragment of the genome (a 'barcode' region; Hebert *et al.*, 2003), has been used to detect NIS with greater sensitivity than traditional survey approaches (e.g. Jerde *et al.*, 2011; Dejean *et al.*, 2012; Takahara *et al.*, 2013). The majority of these studies have involved the detection of one or a few species at a given location, using primers designed to amplify target species such as Asian carp, the American bullfrog, the bluegill sunfish, the New Zealand mudsnail and the red swamp crayfish (Ficetola *et al.*, 2008; Dejean *et al.*, 2011, 2012; Jerde *et al.*, 2011, 2013; Goldberg *et al.*, 2013; Mahon *et al.*, 2013; Takahara *et al.*, 2013; Tréguier *et al.*, 2014). Targeted assays such as these are appropriate for active surveillance of priority species, but are disadvantageous in that they miss non-target NIS that may be present in samples (Handley, 2015). Passive surveillance of NIS through description of whole communities from environmental samples would provide a substantial benefit to NIS management (Handley, 2015).

Through PCR amplification of genes conserved across phyla, the barcoding method can potentially be used to identify multiple species present within environmental samples. Whilst traditional barcoding has involved the identification of single specimens, recent advances in high-throughput sequencing (HTS) technology have allowed the barcoding approach to develop dramatically. The combination of HTS with barcoding has been termed 'metabarcoding' and typically involves bulk DNA extraction, PCR amplification and HTS of complex species assemblages to identify multiple taxa. Samples can be processed either as a homogenized 'soup' of whole organisms (e.g. Fonseca *et al.*, 2010; Yu *et al.*, 2012; Leray *et al.*, 2013) or as environmental DNA typically extracted from soil or water (e.g. Thomsen *et al.*, 2012; Porco *et al.*, 2013; Simmons *et al.*, 2015). The metabarcoding method has been demonstrated to have high sensitivity; that is, species can be detected when present at very low biomass (Hajibabaei *et al.*, 2011; Pochon *et al.*, 2013; Zhan *et al.*, 2013), making it possible to sample rare taxa that are present but not detectable by traditional means. However, very few studies aimed at detecting NIS have applied the metabarcoding method. Pochon *et al.* (2013) used artificial communities containing 10 marine pests at varying concentrations to confirm the high sensitivity of metabarcoding to detect NIS. More recently, Zaiko *et al.* (2015) successfully applied the metabarcoding approach to identify NIS present in ballast water collected from a cruise ship. As yet, the metabarcoding approach has not been applied to survey a broad variety of NIS across a wide geographic range and in complex natural communities.

With rapidly changing environments and species distributions, predicting which species and areas to prioritize for NIS surveillance may be difficult. Perhaps one of the most valuable applications of the metabarcoding method is in the routine surveillance of invasion vectors or key entry points. The threat of invasive species is particularly high in aquatic environments, where shipping traffic and the discharge of foreign ballast water present ideal vectors for their spread (Lodge *et al.*, 2006; Keller *et al.*, 2009). Indeed, shipping has been confirmed as an important vector for dispersal of NIS such as the European green crab (*Carcinus maenas*), mud crab (*Rhithropanopeus harrisii*) and soft shell clam (*Mya arenaria*) (Briski *et al.*, 2012). Our study represents one of the most comprehensive surveys of zooplankton and is, as far as we know, the first to apply the metabarcoding method to detect NIS in natural communities. Here we survey 16 Canadian ports, covering four geographic regions (Atlantic and Pacific Coasts, Arctic and the Great Lakes). We apply two approaches to identify species including NIS that may occur in low abundance, which involve taxonomic assignment of (i) individual reads without prior sequence clustering and (ii) representative sequences generated by the more common practice of operational taxonomic unit (OTU) clustering.

METHODS

Biological sampling

A total of 16 major Canadian ports in four geographic regions (four ports per region) were chosen for sampling based on vessel traffic and ballast water discharge (Fig. 1). Zooplankton samples were collected in the Atlantic coast (Bayside, Baie de Sept-Îles, Halifax, Hawkesbury); Pacific coast (Nanaimo, Robert's Bank, Victoria, Vancouver); Arctic (Churchill, Deception Bay, Iqaluit, Steensby Inlet); and Great Lakes (Hamilton Harbour, Montreal, Nanticoke, Thunder Bay) (see Table S1 in Supporting Information). Samples were collected during ice-off periods over two seasons between May 2011 and December 2012. Within each port, samples were collected in six different sites with geo-referenced 80 and 250 μm oblique plankton net hauls (50 cm diameter

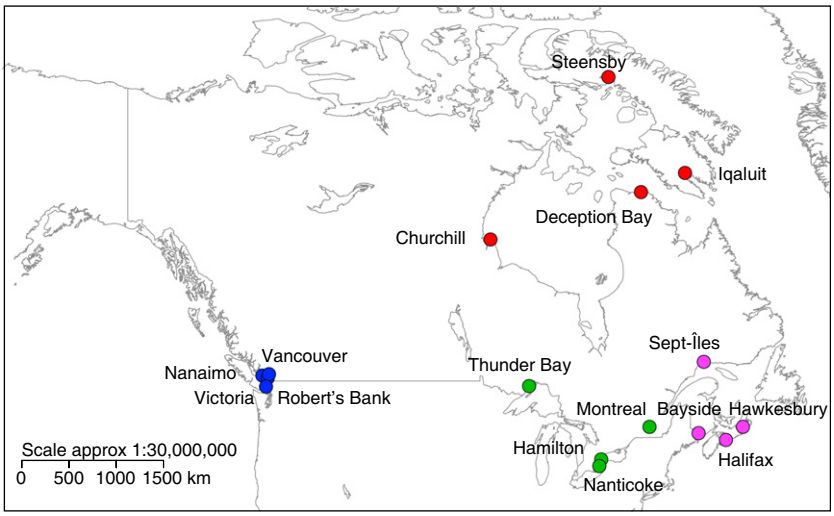

Figure 1 Location of the 16 sampled ports. Samples are from four geographic regions: Pacific Ocean (blue), Arctic Ocean (red), Atlantic Ocean (magenta) and Great Lakes (green). More detailed information on sampling dates and locations is given in Table S1.

opening and 250 cm long, see Supporting Information for more details). Each region was sampled by different teams using separate equipment, and within regions, nets were washed between sites to prevent contamination. Samples were immediately preserved in 95% ethanol.

Sample preparation, DNA extraction, PCR amplification and pyrosequencing

Depending on the available amount of plankton for each sample, 50–150 mg of sample was used for DNA extraction (see Table S2 for more detailed information). Total genomic DNA was isolated using DNeasy Blood and Tissue Kits (Qiagen, Venlo, Limburg, Netherlands) following the manufacturer's protocol. The quality and quantity of each DNA extraction were assessed using gel electrophoresis and Quant-iT PicroGreen dsDNA Assay kit (Invitrogen). Two replicate DNA extractions were carried out per sample, and four replicate PCRs were carried out per DNA extraction, totalling eight PCRs per sample and between 72 and 96 PCRs per port. Negative PCR controls were included to detect contamination and verified via gel electrophoresis; these controls were not sequenced. Each port was also processed separately to avoid cross-contamination. Approximately 400–600 bp of the hypervariable V4 region of the 18S rDNA gene was amplified using the primer pair developed by Zhan *et al.* (2013) (Uni18S: AGGGCAAKYCTGGTGCCAGC; Uni18SR: GRCGGTATCTRATCGYCTT) to amplify crustaceans, molluscs and tunicates. This primer pair was selected due to its ability to recover a wide-range of zooplankton groups (Zhan *et al.*, 2014). PCR mixtures (25 μL) contained approximately 100 ng of genomic DNA, 1 × PCR buffer, 2 mM of Mg^{2+}, 0.2 mM of dNTPs, 0.4 μM of each primer and 2 U of *Taq* DNA polymerase (Genscript, Piscataway, NJ, USA). PCR cycling parameters consisted of an initial denaturation step at 95 °C for 5 min, followed by 25 amplification cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 90 s and a final elongation step at 72 °C for 10 min. With the exception of Hamilton and Vancouver, each sample was amplified with tagged primers that included unique 10-bp tags (MID sequences) approved by Roche (Technical bulletin 005-2009; Roche Diagnostics Corp., Basel, Switzerland) to ensure sample recognition in downstream analyses (see Table 1 for total numbers of samples per port). All PCR products were cleaned using the solid-phase reversible immobilization paramagnetic bead-based method (ChargeSwitch; Invitrogen, Carlsbad, CA, USA). Cleaned PCR products were quantified using PicoGreen and pooled together such that each sample was at equimolar concentration for each port. Each port was pyrosequenced at one-half PicoTiter plate scale using 454 FLX Adaptor A on a GS-FLX Titanium platform (454 Life Sciences, Branford, CT, USA) by Engencore at the University of South Carolina and Genome Quebec at McGill University.

Analytical protocols for non-indigenous species detection

Quality filtering of reads

Reads were assigned to their particular sample based on their tagged primers using a python script provided with UPARSE (Edgar, 2013). This script was also used to trim the tags and forward primer and to remove sequences with errors in these regions, allowing no tag mismatches and two primer mismatches. Reverse primers were removed using the FASTX-Toolkit with default settings, and reads were trimmed for quality (minimum Phred score of 20), whilst only retaining reads longer than 200 bp (http://hannonlab.cshl.edu/fastx_toolkit/). To improve computational efficiency, reads were first dereplicated (collapsed into unique sequences) with UPARSE in each sample separately. Chimera detection and removal was performed either with UCHIME (Edgar *et al.*, 2011) or within the UPARSE-OTU algorithm during sequence clustering (see below).

Taxonomic assignment

Metabarcoding data are often clustered into OTUs based on sequence similarities. If low levels of interspecific genetic variation are present within marker regions, OTU clustering

Table 1 Summary of the reads sequenced for the 16 sampled ports. Quality-filtered reads are those filtered using a minimum Phred quality score of 20 and a minimum length of 200 bp, followed by chimera detection and removal.

		Samples	Raw reads			Quality-filtered reads		
Geo	Port	Total samples	Total reads	Average quality score	Average length	Total reads	Average quality score	Average length
	Total	147	10,277,272			7,733,541		
AR	Churchill	11	605,049	37	479	563,537	37	441
AR	Deception Bay	10	787,293	37	474	735,666	37	433
AR	Iqaluit	11	767,297	37	472	726,701	37	431
AR	Steensby Inlet	6	799,089	37	481	756,776	37	441
PA	Nanaimo	11	789,405	32	345	426,311	33	407
PA	Roberts Bank	10	715,442	36	473	671,760	36	434
PA	Vancouver	1	1,008,358	32	223	290,852	31	413
PA	Victoria	12	456,391	32	346	224,201	33	417
AT	Bayside	12	656,488	33	400	456,232	34	421
AT	Halifax	12	770,511	37	475	726,883	37	434
AT	Hawkesbury	12	444,315	33	374	272,206	34	421
AT	Sept Iles	6	502,688	32	377	342,461	33	392
GL	Hamilton	1	1,099,458	32	181	304,202	32	412
GL	Montreal	8	634,126	37	524	582,848	37	483
GL	Nanticoke	12	480,962	32	359	284,612	33	429
GL	Thunder Bay	12	556,984	34	401	368,293	35	428

AT, Atlantic; PA, Pacific; AR, Arctic; GL, Great Lakes.

will not allow full description of species diversity when closely related species are clustered together (Brown *et al.*, 2015). In such cases, OTU clustering allows assessment of higher taxonomic levels but remains a crude method for assessing species richness. To circumvent this problem, we conducted taxonomic assignment in two ways, using (i) all quality-filtered reads without prior OTU clustering and (ii) representative OTU sequences generated by sequence clustering. We OTU clustered quality-filtered reads from all ports using default settings in UPARSE (Edgar, 2013), implemented in USEARCH v. 7.0.1090. BLAST searches were performed to assign taxonomy to both quality-filtered reads and to the representative sequences of OTUs using a local reference sequence database. The local database, which enabled faster computational processing, consisted of 957,467 18S sequences acquired from the NCBI nucleotide database (in August 2014) and SILVA/SINA version NR99_119 database (Pruesse *et al.*, 2007). These sequences were not verified or quality checked, but previous analyses revealed that most of the common zooplankton groups reported within Canada are detectable using the local database (Chain *et al.*, 2016). Taxonomy was assigned based on the best BLAST hit to a sequence with taxonomic information available. Best hits to Metazoans with a minimum of 370 bp and 97% sequence identity were retrieved using custom Perl scripts and retained for further analysis. Reads/OTUs were classified to the species level if they generated a single best BLAST hit. Reads/OTUs with multiple best BLAST hits (as defined by having the same sequence % identity) were kept but flagged as these cannot be resolved at the species level using our 18S sequences. To further examine reads/OTUs with multiple best BLAST hits, multiple sequence alignments of these reads or the representative sequences of the OTUs, and the representative sequences of their best BLAST hits, were performed with default settings in MUSCLE version 3.8.31 (Edgar, 2004). Neighbour-joining (NJ) trees were then created in MEGA version 5.2.2 (Tamura *et al.*, 2011), enabling the assessment of taxonomic assignment of reads and OTUs relative to phylogenetic placement. Taxonomic incidence-based richness within a port was evaluated using perl and R scripts. Taxonomic diversity was measured as the total number of taxa detected, with these taxa defined as in Fig. 2a, for example Copepoda, Mollusca, Malacostraca.

Detection of non-indigenous species

A list of 124 aquatic NIS of particular relevance to Canada was compiled using online resources and relevant literature (see Table S3). We verified that all species included in the list of NIS had representative 18S sequences in our BLAST database that covered the V4 region. A custom perl script was used to search for the 124 target NIS among the best BLAST hits of quality-filtered reads or OTUs (see *Taxonomic identification of reads/OTUs*). We considered single best BLAST hits with the reference sequences of NIS to signify unambiguous NIS detection.

RESULTS

Taxonomic assignment

A total of 10,277,272 raw reads were pyrosequenced from 147 samples generated from 16 ports (Table 1). Following quality filtering, a total of 7,733,541 reads were retained for

Table 2 NIS detected in Canadian ports located in four geographic regions (AT-Atlantic; PA-Pacific; AR-Arctic; GL-Great Lakes) via 18S metabarcoding when using BLAST searches of quality-filtered reads or OTUs for taxonomic assignment. Species starred with an asterisk are those that were detected only when using one of the two assignment methods. Species in bold were detected by reads that matched reference sequences with > 99% sequence identity.

		AT				PA				AR				GL			
Species	#	BS	HF	HB	SI	NN	RB	VC	VT	CH	DB	IQ	SB	HM	MT	NC	TB
Acartia omorii																	
Reads	1		1														
OTUs	1		1														
Acartia tonsa																	
Reads	21		16	2	3												
OTUs	3		15	2	4												
*Amphibalanus amphitrite**																	
Reads	1					1											
Brachionus plicatilis																	
Reads	2															2	
OTUs	1													1		67	4
*Calocalanus styliremis**																	
Reads	1							1									
Carcinus maenas																	
Reads	1134	291	340	503													
OTUs	3	244	260	433		887	752	326	261								3
Cercopagis pengoi																	
Reads	108															107	1
OTUs	3															92	1
Ciona intestinalis																	
Reads	27		24	3													
OTUs	1		17	3													
Clausocalanus furcatus																	
Reads	3410	1	1601	383	274	15	70	50	882	3	35	92				1	3
OTUs	2	40	757	541	435	21	65	382	16	159	63	173	14				
*Cyclops kolensis**																	
Reads	10											9	1				
Daphnia galeata																	
Reads	10,660	2		1										3043	38	698	6878
OTUs	10	2		1										2645	62	728	5166
Dreissena rostriformis																	
Reads	3974	1												68	502	3402	1
OTUs	12	1												76	4274	2520	1
Eurytemora affinis																	
Reads	755,230	7458	152,393	4682	82	339	38,515	3	114	44,027	2141	174		795	308,843	104,789	90,875

Table 2 Continued.

		AT				PA				AR				GL			
Species	#	BS	HF	HB	SI	NN	RB	VC	VT	CH	DB	IQ	SB	HM	MT	NC	TB
OTUs	45	27	414	263	1377	54	3033		1	5	19			371	55,901	30,018	12,432
Euterpina acutifrons																	
Reads	5870			27						5	818	5011	9				
OTUs	5			24						6	534	2293	4				
Littorina littorea																	
Reads	7	1			4	1		1									
OTUs	2	1			4	1			4								
Mecynocera clausi																	
Reads	1		1														
OTUs	1		1														
Mya arenaria*																	
Reads	37		32							5							
*Mytilopsis leucophaeata**																	
Reads	6														6		
*Mytilus galloprovincialis**																	
OTUs	5		31		68		1			91							
Nitokra hibernica																	
Reads	408													6	116	267	19
OTUs	2													7	79	202	16
Oithona brevicornis[*]																	
Reads	1			1													
Oithona similis																	
Reads	581,964	200	218,836	47,418	33,147	18,641	11,399	492	4629	132,930	104,598	9341	1			315	17
OTUs	87	157	26,983	23,792	22,756	12,150	3527	409	2864	39,071	26,122	2487	1			240	10
Paracalanus parvus																	
Reads	352,590	211	1430	153	1	168,344	108,640	14,866	58,727							66	152
OTUs	75	160	542	151		102,278	24,969	6467	30,464							50	103
Pseudocalanus																	
Reads	1,175,249	197	105,610	44,270	22,571	13,653	193,506	21,171	14,914	104,182	302,997	215,897	136,212			13	56
Elongates																	
OTUs	45	89	17,002	14,452	11,293	4633	36,809	6554	2203	29,231	64,338	34,728	12,287			8	33
Temora turbinata																	
Reads	72,136	478	29,126	12,982	20,224		1			193	105	2			26	3522	5477
OTUs	5	1	5318	1282	1424		597			12		2			28	103	122

taxonomic identification, of which 75% matched a metazoan reference sequence in our database with at least 97% identity. Of these reads, 76% were identified to the species level; those that could not be described to the species level either had multiple best BLAST hits (22%, 7% of all reads) or the best BLAST hit had no species-level information available (2%). Overall, 379 zooplankton species from 320 genera were identified. Crustaceans were the most diverse taxa, representing a third of all identified families (Fig. 2a). Within the arthropods, the taxa Copepoda, Malacostraca and Branchiopoda were particularly species-rich, with 67, 28 and 13 species, respectively. Other species-rich phyla included the molluscs (33 gastropods, 18 bivalves), annelids (44 polychaetes, 7 Clitellata), rotifers (34 Monogononta, 3 Bdelloidea) and cnidarians (31 Hydrozoa, 5 others).

Detection of aquatic non-indigenous species: a taxonomy-dependent approach with greater sensitivity

A list of 124 aquatic NIS with published 18S sequences was compiled to examine whether these species were detected among our samples. A total of 24 NIS (19% of the 124 queried) were detected (Tables 2 & S4), 10 of which are considered to be invasive, that is are NIS known to have caused damage to the environment or economy (*Amphibalanus*

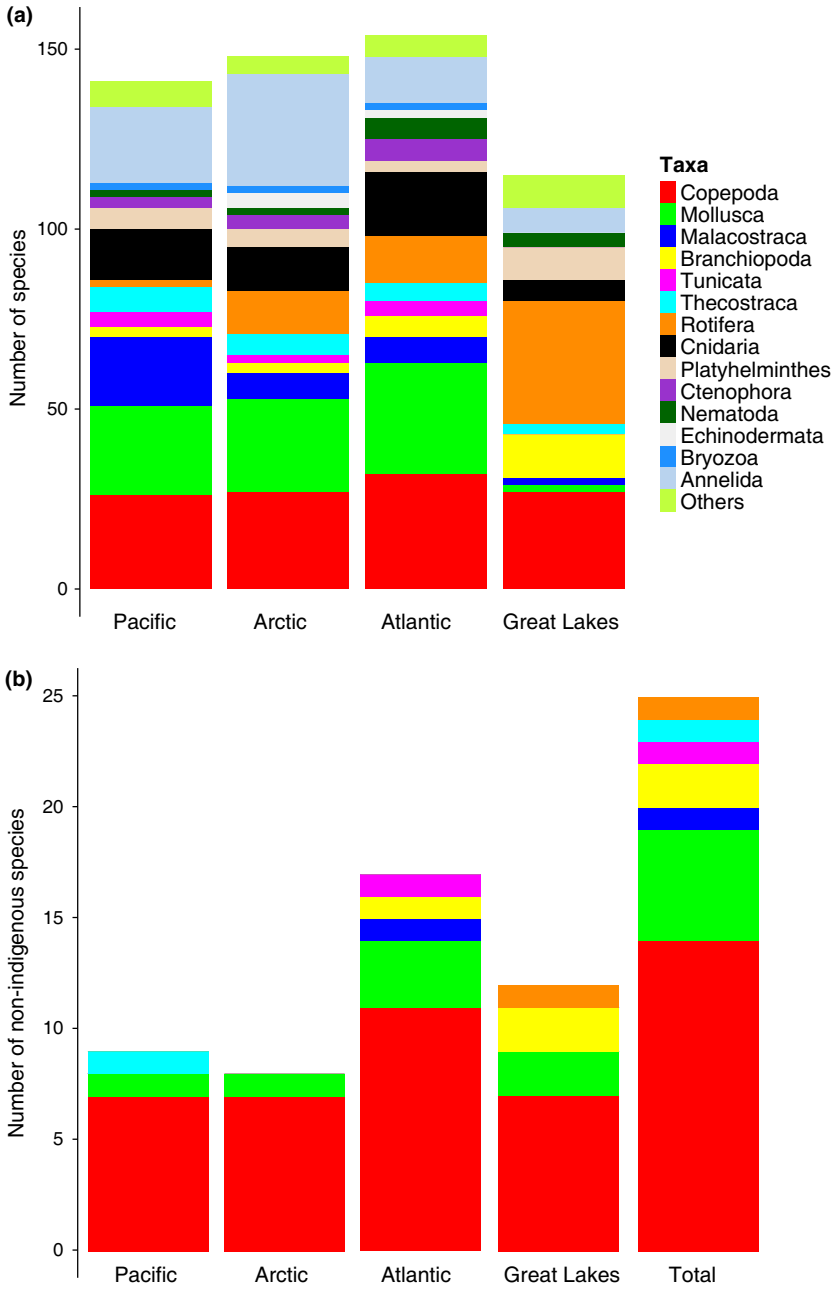

Figure 2 Number of species detected across ports from the four sampled regions when conducting BLAST searches of quality-filtered reads. (a) Total number of species and (b) number of non-indigenous species categorized into higher taxonomic groups.

amphitrite, Carcinus maenas, Cercopagis pengoi, Ciona intestinalis, Daphnia galeata, Dreissena rostriformis, Eurytemora affinis, Littorina littorea, Mya arenaria, Mytilopsis leucophaeata). Of the 24 NIS, five were detected by only a single read (*Acartia omorii, Amphibalanus amphitrite, Calocalanus styliremis, Mecynocera clausi, Oithona brevicornis*). When a more stringent sequence identity threshold (> 99%) was applied for species-level identification, most of the 24 NIS (18) were still detected (Table 2).

The majority of NIS detected across all regions were copepods, followed by molluscs and branchiopods (Fig. 2b). The greatest number of NIS was detected in the Atlantic, and these NIS were the most diverse taxonomically, belonging to the taxa Copepoda, Mollusca, Malacostraca, Branchiopoda and Tunicata. Overall, the Atlantic was also the most species-rich region, although we did not find a correlation between the total number of species and the number of NIS detected in a region (Pearson's product moment correlation coefficient: $r = 0.163$, $P = 0.837$). We also found no correlation between the taxonomic diversity of a region and the number of NIS detected (Pearson's $r = 0.022$, $P = 0.978$). We did, however, find evidence for a correlation between the number of NIS detected in a region and the taxonomic diversity of those NIS (Pearson's $r = 0.958$, $P = 0.042$). Indeed, the fewest NIS were detected in the Arctic and these were also the least rich taxonomically, belonging to Copepoda and Mollusca. Whilst certain NIS were detected in all regions, some taxa were detected only in a particular port or region, including Rotifera (*Brachionus plicatilis*) in the Great Lakes, Tunicata (*Ciona intestinalis*) and Malacostraca (*Carcinus maenas*) in the Atlantic, and Thecostraca (*Amphibalanus amphitrite*) in the Pacific (Figs 2b & 3).

We identified 11 NIS (*Acartia omorii, Clausocalanus furcatus, Cyclops kolensis, Eurytemora affinis, Oithona brevicornis, Pseudocalanus elongatus, Temora turbinata, Amphibalanus amphitrite, Dreissena rostriformis, Mya arenaria, Mytilopsis leucophaeata*) in regions where they have not previously been reported or predicted to occur (Table S4). Three of these NIS were detected by a single read (*Acartia omorrii,*

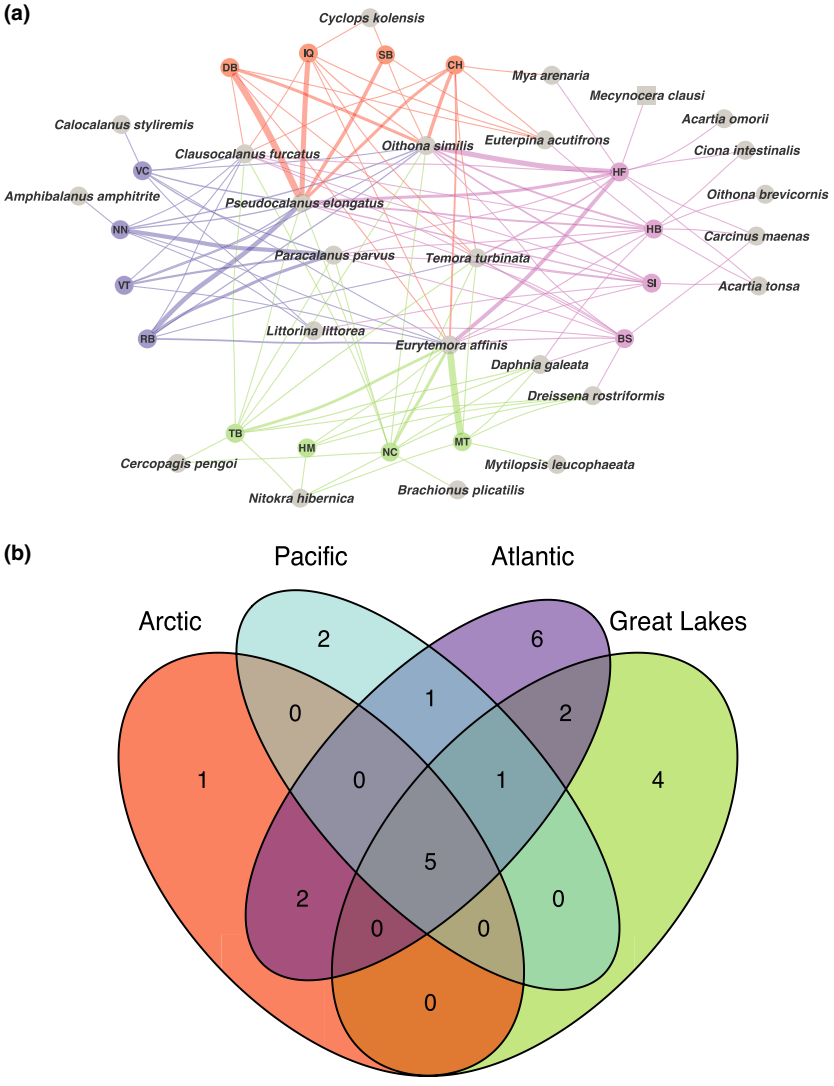

Figure 3 NIS occurrence across our samples. (a) The detection of NIS among 16 ports from the Pacific (blue), Arctic (red), Atlantic (purple) and Great Lakes (green). Connecting lines show the locations in which a species was detected, with greater number of reads represented by thicker line width. Port name abbreviations correspond to those used in Table 2. Created with CYTOSCAPE v. 3.1.1 (Shannon *et al.*, 2003). (b) Venn diagram showing the number of NIS detected across the four geographic regions.

Amphibalanus Amphitrite, Oithona brevicornis), whereas most others are supported by hundreds or thousands of reads. There were also 13 NIS (*Calocalanus styliremis*, *Eurytemora affinis, Mecynocera clausi, Oithona similis, Paracalanus parvus, Pseudocalanus elongatus, Daphnia galeata, Carcinus maenas*, *Dreissena rostifornis, L. littorea, Mya arenaria, B. plicatilis, Ciona intestinalis*) that were detected in geographic region(s) or port(s) in which they have previously been reported. For seven NIS (*Acartia tonsa, Clausocalanus furcatus, Euterpina acutifrons, Nitokra hivernica, T. turbinata, Cercopagis pengoi, L. littorea*), although the species were detected in regions where they have been previously identified, we detected these species in ports that we believe to be beyond their reported limits.

A number of reads (22%) could not be identified at the species level when the 97% sequence identity threshold was applied because they generated multiple best BLAST hits. A total of 748,146 reads equally matched NIS on our list and a number of closely related species (Table S5). This was due to high sequence similarity among 18S sequences, suggesting that the V4 region of 18S offers insufficient resolution to identify these taxa. For example, a total of nine reads from the Churchill and Steensby Inlet samples generated multiple best BLAST hits with *Mya truncata*, *Mya arenaria*, and *Corbula coxi*. A NJ tree of these reads, together with reference sequences for the three species, shows that some reads cannot be resolved taxonomically due to insufficient informative genetic differences (Fig. 4). The aligned and trimmed reference sequences for *Mya truncata* and *Corbula coxi* included in the phylogeny were identical, and the reference sequence for *Mya arenaria* differed from these species at only one base pair position. As another example, the 18S sequence of the invasive European green crab *Carcinus maenas* is 99% similar to other decapods not included in our list of NIS, such as *Cancer pagurus, Acantholobulus bermudensis, Hepatus epheliticus* and *Praebebalia longidactyla*. Three reads from the Pacific dataset matched all five of these species with equivalent BLAST scores.

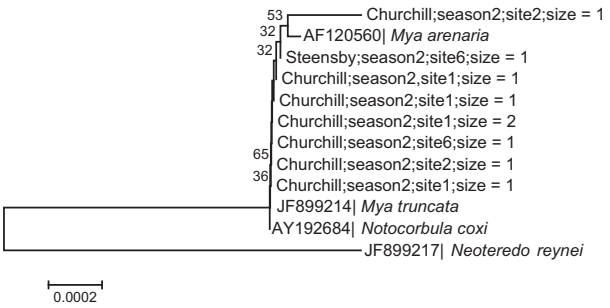

Figure 4 Phylogenetic tree of quality-filtered reads that matched *Mya arenaria*, *Mya truncata* and *Corbula coxi* with equivalent BLAST score, including representative sequences for these three species and the outgroup *Neoteredo reynei* (98% similar). The tree was generated using the neighbour-joining method with 1000 bootstrap replicates. Representative sequences are labelled with their accession numbers and the species name. Each read is labelled according to the sample it originated from, for example Steensby Inlet, season 2, site 6. Prior to being used in BLAST searches, these reads were dereplicated, and the number of reads (*n*) that were found to be identical to the reads shown on the phylogeny is indicated by 'size = *n*'.

Detection of aquatic non-indigenous species: the OTU clustering approach

In general, OTU clustering resulted in many OTUs being generated per species, with reads originating from multiple geographic regions and ports joining to form shared OTUs. A total of 19 NIS were detected when OTU clustering was performed prior to BLAST searches (Table 2), including eight invasive species (*Carcinus maenas, Cercopagis pengoi, Ciona intestinalis, Daphnia galeata, Dreissena rostiformis, Eurytemora affinis, L. littorea, Mytilus galloprovincialis*) (see Table S4 for more details). Two of the 19 NIS were detected by a single read (*Acartia omorii, Mecycnocera clausi*).

Of the 18 NIS that were identified both using OTU clustering and individual reads, 16 (89%) were detected in the same geographic region using the two methods. One exception was *Clausocalanus furcatus,* which was detected in the Pacific, Atlantic and Arctic by both methods but was detected in the Great Lakes by only four individual reads. We constructed a NJ tree of the four reads and the consensus sequences of the two OTUs that identified *Clausocalanus furcatus* in the Pacific, Arctic and Atlantic, along with a reference sequence for *Clausocalanus furcatus* (Fig. 5a). Based on this tree, three reads are more closely related to the reference sequence of *Clausocalanus furcatus* than the two OTU consensus sequences. The other exception, *Carcinus maenas,* was identified by an OTU composed of three reads from the Great Lakes, but when BLASTing individual reads, none of the Great Lakes reads matched *Carcinus maenas.* The three reads belonging to the OTU matching *Carcinus maenas* were found to BLAST against multiple species not included in our list of NIS at 98–99% identity over 438–462 bp (*Carpilius maculatus*, *Panopeus herbstii*, *Cancer pagurus*, *Acantholobulus bermudensis, H. epheliticus* and *Praebebalia longidactyla*). A NJ tree of the three reads and the reference sequences for the species generating BLAST hits suggests that the three reads were likely not generated by *Carcinus maenas* (Fig. 5b). Thus in this case, reads belonging to closely related species are potentially misidentified as belonging to an NIS.

A total of six NIS identified when BLASTing reads were not detected with OTU clustering (*Amphibalanus amphitrite*, *Calocalanus styliremis, Cyclops kolensis, Mya arenaria, Mytilopsis leucophaeata, Oithona brevicornis*) (Table 2). For these six species, we traced the reads that BLASTed against these NIS to see what OTUs they joined when OTU clustering was performed. We found that these reads joined OTUs together with the reads of closely related species, and the species that generated the best BLAST hit was not the NIS. For example, a total of 37 reads BLASTed uniquely against *Mya arenaria,* which were found to map back to one OTU. This OTU BLASTed against *Mya arenaria,* at 99.77%, but

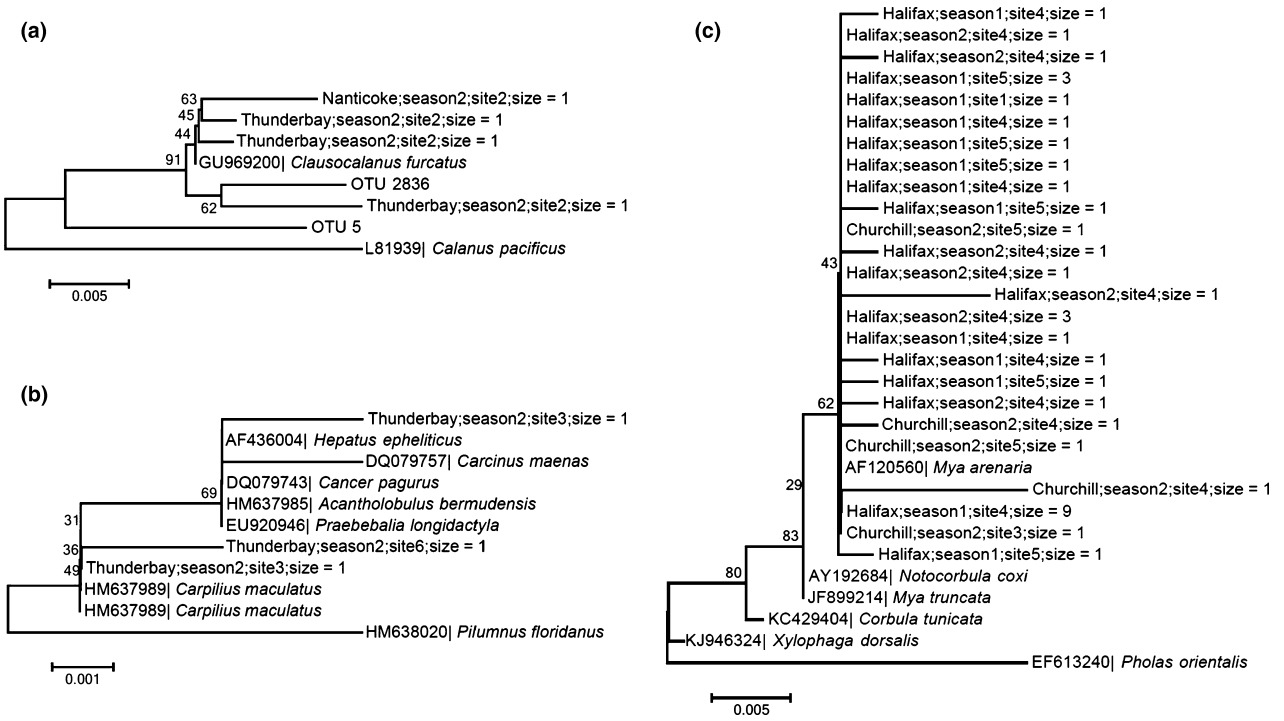

Figure 5 Phylogenetic trees generated using the neighbour-joining (NJ) method with 1000 bootstrap replicates demonstrating both false-negative (a, c) and false-positive (b) identifications of non-indigenous species. Representative sequences for the species are labelled with their accession numbers and the species name. Each read is labelled according to the sample it originated from, for example Nanticoke, season 2, site 2. Prior to being used in BLAST searches, these reads were dereplicated, and the number of reads (*n*) that were found to be identical to the reads shown on the phylogeny is indicated by 'size = *n*'. (a) NJ phylogeny of four quality-filtered reads originating from Great Lakes ports that generated BLAST hits to *Clausocalanus furcatus* and the consensus sequences of the two OTUs that generated hits to the same species. Representative sequences for *C. furcatus* and the outgroup *Calanus pacificus* (96% similar) were also included. (b) NJ phylogeny of three reads belonging to an OTU that generated a BLAST hit to *Carcinus maenas.* When used individually in BLAST searches, these reads generated BLAST hits with a number of closely related species, also included in the phylogeny. *Pilummnus floridanus* (98% similar) is included as an outgroup. (c) NJ phylogeny of quality-filtered reads BLASTing against *Mya arenaria* and representative sequences of *Corbula tunicata, Corbulidae* gen. sp., *M. arenaria, Notocorbula coxi, M. truncata, Xylophaga dorsalis* and the outgroup *Pholas orientalis* (96% similar).

also with two closely related species (98–99% sequence similarity), *Mya truncata* and *Corbula coxi,* at 100% over the same length. A NJ tree of the reads that generated single best BLAST hits to *Mya arenaria*, together with reference sequences of species matching the OTU that these reads joined, suggests that the taxonomic assignment of individual reads before clustering was accurate (Fig. 5c). Over the 430 bp aligned and trimmed V4 region included in the phylogeny, *Mya arenaria, Notocorbula coxi, Mya truncata, Xylophaga dorsalis* and *Corbula tunicata* differed at fewer than five base positions. Despite this low sequence divergence, the reads are monophyletic with *Mya arenaria.* The reads share a unique nucleotide substitution present in the reference sequence of *Mya arenaria* but absent in the reference sequences of the closely related species *X. dorsalis*, *Corbula tunicata, Notocorbula coxi* and *Mya truncata* (Fig. S1).

One NIS, *Mytilus galloprovincialis,* was only identified when OTU clustering was performed prior to BLAST searches. Although this species was identified by five different OTUs and over 100 reads, none of the individual reads that formed these OTUs could be unambiguously identified to the species level as they generated multiple top BLAST hits with *Mytilus trossulus, Mytilus galloprovincialis* and *Mytilus edulis.*

DISCUSSION

Genetic tools have become increasingly prevalent in conservation and management efforts (Schwartz *et al.*, 2007). In this study, we demonstrate the applicability and value of the metabarcoding approach for the detection of NIS, specifically in complex zooplankton communities sampled across 16 major Canadian ports. Whilst previous NIS surveillance efforts have focused almost exclusively on one or a few species and have been geographically narrow (e.g. Ficetola *et al.*, 2008; Jerde *et al.*, 2011; Takahara *et al.*, 2013), we use a broadly amplifying primer pair that allows for passive surveillance of many species. By conducting BLAST searches of individual reads, we detected 24 NIS from 379 identified species, with some NIS found with very low read number.

One of the major advantages of metabarcoding over alternatives that rely on traditional taxonomy is the increased sensitivity gained for species occurring at low abundance. Of the 24 NIS detected, five were identified by a single read, and whilst there has been debate over the validity of singletons, these reads could represent low abundance eukaryotes such as recently introduced NIS. Although we use 454 sequencing rather than Illumina, our findings are not sequencing platform specific and are of relevance to any researcher posed with the challenge of processing large volumes of metabarcoding data.

Of the 24 NIS detected, to the best of our knowledge, 11 were detected in regions where they have not previously been reported. For example, *Cyclops kolensis, Eurytemora affinis, T. turbinata* and *Mya arenaria* were detected in the Arctic, above their reported northern limits. Decreasing Arctic sea ice generates new opportunities for trans-Arctic shipping routes, and coupled with the changing climate, leaves the region increasingly vulnerable to invasion by NIS (Vermeij & Roopnarine, 2008; Smith & Stephenson, 2013; Ware *et al.*, 2014). Whilst the detection of these species in the Arctic could indicate the existence of invasive populations, we cannot confirm that these species exist as viable, reproducing populations. Ships predominantly transfer aquatic organisms in ballast water tanks or as biofouling. Mortality of organisms present within ballast water is known to increase with time, and the thermal reproductive requirements of surviving animals may pose a barrier to species invasion (Ware *et al.*, 2014). We also detected the barnacle *Amphibalanus amphitrite* in Nanaimo, British Columbia, above its reported northern limit of San Francisco. This barnacle may occur sporadically farther north, but it is not thought to be capable of surviving the winter (Fofonoff *et al.*, 2003). We detected *Amphibalanus amphitrite* with a single read, perhaps suggesting that the species has not established a reproducing population. Similarly, *Mytilopsis leucophaeata,* a species that has not previously been recorded in Canadian waters but has been reported as an industrial pest in some parts of Europe (Kennedy, 2011), was detected by six reads in the Great Lakes port of Montreal. To confirm the presence of NIS detected by a small number of reads, passive surveillance could be followed by a targeted approach, such as active surveillance with digital drop PCR (e.g. Simmons *et al.*, 2015).

Perhaps surprisingly, several NIS were detected among all sampled regions, including the Arctic and Great Lakes (Table 2; Fig. 3). This finding could result from previous underestimation of species' occurrences due to less intensive sampling or could represent an alarming scenario in which many of these NIS have recently attained widespread distributions. Field or laboratory-based contamination is also a possibility, although the four geographic regions were sampled by independent teams using different equipment and care was taken to reduce contamination between sites. As each port was separately processed in the laboratory at different times, it is also highly unlikely that cross-contamination occurred at this step. The detection of freshwater species in marine ports, and vice-versa, may raise concern over the validity of these findings. It is not unusual, however, to occasionally detect freshwater organisms in marine ports, particularly in ports that receive heavy freshwater input. Moreover, marine ballast water release in freshwater ports is expected to result in the release of marine organisms. The detection of a number of NIS in previously reported locations also lends support to the reliability of the metabarcoding method to accurately detect NIS. For example, the European green crab, *Carcinus maenas*, is known to have successfully invaded the Atlantic coast of North America (Carlton & Cohen, 2003; Klassen & Locke, 2007) and was detected by 1134 reads in three Atlantic ports. We detected the copepod *Calocalanus styliremis* in Vancouver, British Columbia, where it has previously been identified in the ballast water of ships entering the port (DiBacco *et al.*, 2012). The cladoceran *Daphnia galeata* has previously been detected in the ballast water of vessels entering the Great Lakes and Atlantic, and we identified this species in both of these regions.

It should be noted that our classification of NIS in relation to their previously reported distributions (see Table S4) is based on information available through online resources and primary literature. It is possible that more recent species distribution data are available elsewhere. In addition, many of the species examined here, such as *Paracalanus parvus, Amphibalanus amphitrite* and *B. plicatilis*, belong to species complexes that are difficult to classify morphologically, making it difficult to track their invasion history. The development of databases that report up-to-date information on species presence over a fine geographic scale would greatly aid attempts to document and track species identifications. Our study presents an ideal means by which such information can be generated and effectively compiled.

DNA-based species identification also relies on the existence of a comprehensive sequence database. The majority (75%) of our quality-filtered reads could be matched to a reference sequence in our local database with at least 97% identity. Given that our reads were quality-filtered prior to BLAST, we suspect that most reads that did not match a reference sequence at 97% identity belonged to organisms missing from current 18S sequence databases. It is also possible that reads matching reference sequences at 97% and above could belong to species missing from databases, leading to false species detections. Using a stricter threshold of 99% sequence identity, 18 of the 24 NIS were still detected (Table 2). Global initiatives addressing the need to develop well-populated and regulated sequence libraries have been established, and databases are rapidly growing (e.g. the BOLD database, http://www.boldsystems.org/). However, the remaining knowledge gap will for some time impair the usefulness of DNA-based monitoring in groups where all species have not yet been DNA barcoded. This will be particularly problematic for the most complex and species-rich systems, which are the most challenging to monitor with conventional methods and thus also the most likely to benefit from DNA-based monitoring. For example, a number of our reads

matched sequences labelled as 'uncultured eukaryote' that were most likely generated by previous metabarcoding studies. Unless these organisms are investigated taxonomically, it will not be possible to provide Linnaean species descriptions for them.

In addition, we found that species-level resolution of our marker was poor for some groups. The rate of evolution of nuclear ribosomal RNA genes is slow in comparison with mitochondrial DNA, which can result in insufficient taxonomic resolution for the identification of closely related species (Porazinska *et al.*, 2009). For example, the 18S marker can easily differentiate between two *Acartia* species with sequence similarity of 86% (*Acartia tonsa* and *Acartia omorii*), but cannot discriminate between *Calanus* species with sequence similarity of 100% (*Calanus helgolandicus* and *Calanus pacificus*). When metabarcoding is used to broadly estimate biodiversity, identification at a higher taxonomic level, such as family or order, is often sufficient (Valentini *et al.*, 2009; Riaz *et al.*, 2011). However, in the case of NIS detection, species-level identification is often essential, unless an entire group of higher-level taxa (i.e. genus, family) is known to be non-native to regions targeted by a study. Overall, 22% of our reads matched multiple species equally well beyond 97% identity. This problem could be alleviated by further development of sequence databases that allow interspecific variability to be quantified within groups of interest and facilitate the detection of diagnostic substitutions that differentiate closely related species. For example, we detected the presence of a nucleotide substitution within the V4 region of *Mya arenaria* that allowed this species to be distinguished from related species with very low sequence divergence (Fig. S1). Employing a multiple-marker approach that involves a cocktail of wide-range and group-specific primers could also allow multiple taxa to be amplified whilst retaining species-level resolution (Aylagas *et al.*, 2014).

Metabarcoding data are often clustered into OTUs to account for intraspecific variation and artefactual sequences that can be generated during PCR and/or sequencing. Low levels of interspecific variation, however, mean that both false-positive and false-negative NIS detections can be generated when OTU clustering is employed. Our analyses suggest that a taxonomy-dependent approach that does not involve prior clustering of sequences has greater sensitivity and species resolution; we found that six NIS identified when reads were BLASTed against our local database were no longer detected among OTUs. The reads matching these NIS formed OTUs with closely related species, such that the consensus sequences for the OTUs did not generate a BLAST hit with the NIS (Fig. 4). In addition, OTU clustering appears to have falsely identified the presence of an NIS (*Carcinus maenas*) in some ports due to low sequence divergence between closely related species. *Carcinus maenas* was detected by an OTU that included reads from Atlantic and Pacific ports as well as three reads originating from the Great Lakes port Thunder Bay, which is a highly unlikely habitat for this species. When the three reads from freshwater ports were individually used in BLAST searches, they matched a number of other species with equivalent BLAST scores. A NJ tree of the reads suggested that they do not belong to *Carcinus maenas* (Fig. 5b). Our findings encourage the development of sequence databases and a move away from OTU clustering for taxonomic identification and detection of target species, especially those at low abundance.

In conclusion, our results suggest that metabarcoding is a powerful method not only for zooplankton biodiversity assessment but also for the detection of NIS. In total, we detected 25 NIS, a number of which were present at apparently low abundance. In certain groups, low interspecific divergence within our marker made distinguishing closely related species problematic. As a result, NIS may go undetected or be falsely identified if reads belonging to closely related species are clustered together into shared OTUs or cannot be distinguished based on BLAST scores. We suggest that future work should focus on the development of well-populated and regulated sequence databases that allow individual reads to be directly used for taxonomic assignment without the need for OTU clustering, and on the use of multiple markers that allow the full taxonomic breadth of communities to be described to an appropriate taxonomic level. Once an invasive species becomes established in an aquatic habitat, it can be very challenging to eliminate (Thresher & Kuris, 2004), making accurate and reliable identification of NIS crucial. With limited funds, establishing priorities is key, which requires knowledge of the species most likely to harm native ecosystems, current distributions of these species, and how they are likely to be transported to new regions (Byers *et al.*, 2002). This study provides a methodological framework for the analysis of zooplankton diversity and NIS presence, informing on species distributions and the potential mechanisms shaping them.

ACKNOWLEDGEMENTS

We thank our NSERC Canadian Aquatic Invasive Species Network (CAISN) colleagues, in particular the sample collection teams including K. Howland, R. Young, J. Goldsmit and S. Qureshi. We also thank C. Abbott, S. Adamowicz and T. Therriault for helpful discussions. We would also like to thank the anonymous referees, whose comments improved this manuscript. This research was supported by the NSERC Canadian Aquatic Invasive Species Network (CAISN) and Discovery grants to H.J.M. and M.E.C., by the NSERC CREATE training programme in Aquatic Ecosystem Health to M.E.C., by the 100-talent programme of the Chinese Academy of Sciences to A.Z., and Canada Research Chairs to H.J.M. and M.E.C.

DATA ACCESSIBILITY

The 454 data generated in this study are accessible in the European Nucleotide Archive (Accession Number PRJEB11768).

REFERENCES

Aylagas, E., Borja, A. & Rodríguez-Ezpeleta, N. (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). *PLoS One*, **9**, e90529.

Briski, E., Ghabooli, S., Bailey, S.A. & MacIsaac, H.J. (2012a) Invasion risk posed by macroinvertebrates transported in ships' ballast tanks. *Biological Invasions*, **14**, 1843–1850.

Brook, B.W., Sodhi, N.S. & Bradshaw, C.J.A. (2008) Synergies among extinction drivers under global change. *Trends in Ecology and Evolution*, **23**, 453–460.

Brown, E.A., Chain, F.J.J., MacIsaac, H.J. & Cristescu, M.E. (2015) Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecology and Evolution*, **5**, 2234–2251.

Byers, J.E., Reichard, S., Randall, J.M., Parker, I.M., Smith, C.S., Lonsdale, W.M., Atkinson, I.A.E., Seastedt, T.R., Williamson, M., Chornesky, E. & Hayes, D. (2002) Direction research to reduce the impacts of nonindigenous species. *Conservation Biology*, **16**, 630–640.

Caesar, R.M., Sorensson, M. & Cognato, A.I. (2006) Integrating DNA data and traditional taxonomy to streamline biodiversity assessment: an example from edaphic beetles in the Klamath ecoregion, California USA. *Diversity and Distributions*, **12**, 483–489.

Cao, Y., Williams, D.D. & Williams, N.E. (1998) How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, **43**, 1403–1409.

Carlton, J.T. & Cohen, A.N. (2003) Episodic global dispersal in shallow water marine organisms: the case history of the European shore crabs *Carcinus maenas* and *C. aestuarii*. *Journal of Biogeography*, **30**, 1809–1820.

Chain, FJJ, Brown, EA, MacIsaac, HJ & Cristescu, ME (2016) Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions*, **22**, 493–504.

Crooks, J.A. & Soulé, M.E. (1999) Lag times in population explosions of invasive species: causes and implications. *Invasive species and biodiversity management* (ed. by O.T. Sandlund, P.J. Schei and A. Viken), pp. 103–125. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Dejean, T., Valentini, A., Duparc, S., Pellier-Cuit, F., Pompanon, F., Taberlet, P. & Miaud, C. (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS One*, **6**, e23398.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E. & Miaud, C. (2012) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, **49**, 953–959.

DiBacco, C., Humphrey, D.B., Nasmith, L.E. & Levings, C.D. (2012) Ballast water transport of non-indigenous zooplankton to Canadian ports. *ICES Journal of Marine Science*, **69**, 483–491.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quine, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Fofonoff, P.W., Ruiz, G.M., Steves, B. & Carlton, J.T. (2003) California Non-native Estuarine and Marine Organisms (Cal-NEMO) System. Available at: http://invasions.si.edu/nemesis/ (accessed 3 July 2015).

Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., Packer, M., Blaxter, M.L., Lambshead, P.J., Thomas, W.K. & Creer, S. (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, **1**, 98.

Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L.P. (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, **32**, 792–800.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, **6**, e17497.

Handley, L.L. (2015) How will the 'molecular revolution' contribute to biological recording? *Biological Journal of the Linnean Society*, **115**, 750–766.

Harvey, C.T., Qureshi, S.A. & MacIsaac, H.J. (2009) Detection of a colonising, aquatic, non-indigenous species. *Diversity and Distributions*, **15**, 429–437.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identification through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Jerde, C.L., Mahon, A.R., Chadderton, L. & Lodge, D.M. (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.

Jerde, C.L., Chadderton, W.L., Mahon, A.R., Renshaw, M.A., Corush, J., Budny, M.L., Mysorekar, S. & Lodge, D.M. (2013) Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Science*, **70**, 522–526.

Keller, R.P., Philine, S.E., Ermgassen, Z.U. & Aldridge, D.C. (2009) Vectors and timing of freshwater invasions in Great Britain. *Conservation Biology*, **23**, 1526–1534.

Kennedy, V.S. (2011) The invasive dark falsemussel *Mytilopsis leucophaeata* (Bivalvia: Dreissenidae): a literature review. *Aquatic Ecology*, **45**, 163–183.

Klassen, G. & Locke, A. (2007) A biological synopsis of the European green crab, *Carcinus maenas*. Canadian

Manuscript Report of Fisheries and Aquatic Sciences, 2818: vii + 75 pp.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & Machida, R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterising coral reef fish gut contents. *Frontiers in Zoology*, **10**, 34.

Lodge, D.M., Williams, S., MacIsaac, H.J., Hayes, K.R., Leung, B., Reichard, S., Mack, R.N., Moyle, P.B., Smith, M., Andow, D.A., Carlton, J.T. & McMichael, A. (2006) Biological invasions: recommendations for US policy and management. *Ecological Applications*, **16**, 2035–2054.

Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, W.L., Lodge, D.M., Hunter, M.E. & Nico, L.G. (2013) Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS One*, **8**, e58316.

McDonald, L.L. (2004) Sampling rare populations. *Sampling rare or elusive species* (ed. by W.L. Thompson), pp. 11–42. Island Press, New York.

Molnar, J.L., Gamboa, R.L., Revenga, C. & Spalding, M.D. (2008) Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, **6**, 485–492.

Pejchar, L. & Mooney, H. (2009) Invasive species, ecosystem services and human well-being. *Trends in Ecology and Evolution*, **24**, 497–504.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000) Environmental and economic costs associated with non-indigenous species in the United States. *BioScience*, **50**, 53–65.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Pochon, X., Bott, N.J., Smith, K.F. & Wood, S.A. (2013) Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS One*, **8**, e73935.

Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N. & Morris, K. (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, **9**, 1439–1450.

Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarzyński, D., Erséus, C., Butt, K.R., Richard, B. & Hebert, P.D.N. (2013) Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biological Invasions*, **15**, 899–910.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. & Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARM. *Nucleic Acids Research*, **35**, 7188–7196.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. & Coissac, E. (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.

Schwartz, M.K., Luikart, G. & Waples, R.S. (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, **22**, 25–33.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.

Simmons, M., Tucker, A., Chadderton, L., Jerde, C.L. & Mahon, A.R. (2015) Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences*, **73**, 76–83.

Smith, L.C. & Stephenson, S.R. (2013) New Trans-Arctic shipping routes navigable by mid-century. *Proceedings of the National Academy of Sciences USA*, **110**, e1191–e1195.

Takahara, T., Minamoto, T. & Doi, H. (2013) Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS One*, **8**, e56584.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. & Willerslev, E. (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One*, **8**, e41732.

Thresher, R.E. & Kuris, A.M. (2004) Options for managing invasive marine species. *Biological Invasions*, **6**, 295–300.

Tréguier, A., Paillisson, J.M., Dejean, T., Valentini, A., Schlaepfer, M.A. & Roussel, J.M. (2014) Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkia* in freshwater ponds. *Journal of Applied Ecology*, **51**, 871–879.

Valentini, A., Miquel, C., Nawaz, M.A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruad, C., Nascetti, G., Wincker, P., Swenson, J.E. & Taberlet, P. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trn*L approach. *Molecular Ecology Resources*, **9**, 51–60.

Vander Zanden, M.J., Hansen, G.J.A., Higgins, S.N. & Kornis, M.S. (2010) A pound of prevention, plus a pound of cure: early detection and eradication of invasive species in the Laurentian Great Lakes. *Journal of Great Lakes Research*, **36**, 199–205.

Vermeij, G.J. & Roopnarine, P.D. (2008) The coming Arctic invasion. *Science*, **321**, 780–781.

Ware, C., Berge, J., Sundet, J.H., Kirkpatrick, J.B., Coutts, A.D.M., Jelmert, A., Olsen, S.M., Floerl, O., Wisz, M.S. & Alsos, I.G. (2014) Climate change, non-indigenous species and shipping: assessing the risk of species introduction to a high-Arctic archipelago. *Diversity and Distributions*, **20**, 10–19.

Yu, D.W., Yinqiu, J., Emerson, B.C., Wang, X., Ye, C., Yang, C. & Dung, Z. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

Zaiko, A., Martinez, J.L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A. & Garcia-Vazquez, E. (2015) Metabarcoding approach for the ballast water surveillance – an advantageous solution or an awkward challenge? *Marine Pollution Bulletin*, **92**, 25–34.

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Abebayo, A.A., Abbott, C.L., Adamowicz, S.J., Heath, D.D., Cristescu, M.E. & MacIsaac, H.J. (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, **4**, 558–565.

Zhan, A., Bailey, S.A., Heath, D.D. & MacIsaac, H.J. (2014) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources*, **14**, 1049–1059.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Zooplankton collection dates and sample locations.

Table S2 Sample preparation method used when preparing for DNA extractions.

Table S3 The 124 NIS targeted by this study.

Table S4 NIS detected in Canadian ports using 18S metabarcoding.

Table S5 Reads and OTUs with multiple best BLAST hits of > 97% identity and 370 bp length.

Figure S1 Divergent nucleotide (at position 224) potentially allowing identification of reads belonging to *Mya arenaria*.

BIOSKETCH

Emily A. Brown is an evolutionary biologist interested in studying the interplay between ecology and evolution. All authors are part of the Canadian Aquatic Invasive Species Network (CAISN) www.caisn.ca.

Author contributions: H.J.M. and M.E.C. conceived and designed the research. E.A.B. wrote this manuscript, and all authors contributed to its editing. E.A.B. and A.Z conducted the laboratory work. Taxonomic assignment of the reads/OTUs was performed by F.J.J.C. E.A.B. worked with F.J.J.C. to carry out the additional data analyses.

Editor: Andrew Lowe