




Environment and shipping drive environmental DNA beta-diversity among commercial ports

Jose Andrés^{1,2} | Paul Czechowski^{1,3,4}  | Erin Grey^{5,6}  | Mandana Saebi⁷ | Kara Andres^{1,2} | Christopher Brown⁸ | Nitesh Chawla⁷ | James J. Corbett⁹ | Rein Brys¹⁰  | Phillip Cassey¹¹ | Nancy Correa^{12,13} | Marty R. Deveney¹⁴ | Scott P. Egan¹⁵ | Joshua P. Fisher¹⁶ | Rian vanden Hooff¹⁷ | Charles R. Knapp¹⁸ | Sandric Chee Yew Leong¹⁹ | Brian J. Neilson²⁰ | Esteban M. Paolucci²¹ | Michael E. Pfrender²² | Meredith R. Pochardt²³ | Thomas A. A. Prowse¹¹ | Steven S. Rumrill²⁴ | Chris Scianni^{25,26} | Francisco Sylvester²⁷ | Mario N. Tamburri²⁸ | Thomas W. Therriault²⁹ | Darren C. J. Yeo^{30,31} | David M. Lodge^{1,2}

¹Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA

²Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, New York, USA

³Department of Anatomy, University of Otago, Dunedin, New Zealand

⁴Helmholtz Institute for Metabolic, Obesity and Vascular Research, Leipzig, Germany

⁵School of Biology and Ecology and Maine Center for Genetics in the Environment, University of Maine, Orono, Maine, USA

⁶Division of Science, Mathematics and Technology, Governors State University, University Park, Illinois, USA

⁷Center for Network and Data Science (CNDS), University of Notre Dame, Notre Dame, Indiana, USA

⁸Golden Bear Research Center, California State University Maritime Academy, Vallejo, California, USA

⁹College of Earth, Ocean, and Environment, University of Delaware, Newark, Delaware, USA

¹⁰Research Institute for Nature and Forest, Geraardsbergen, Belgium

¹¹School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

¹²Servicio de Hidrografía Naval (Ministerio de Defensa), Buenos Aires, Argentina

¹³Escuela de Ciencias del Mar, Sede Educativa Universitaria, Facultad de la Armada, UNDEF, Buenos Aires, Argentina

¹⁴SARDI Aquatic Science and Marine Innovation SA, South Australian Research and Development Institute, West Beach, South Australia, Australia

¹⁵Department of BioSciences, Rice University, Houston, Texas, USA

¹⁶United States Fish and Wildlife Service, Pacific Islands Fish and Wildlife Office, Honolulu, Hawaii, USA

¹⁷Oregon Department of Environmental Quality, Portland, Oregon, USA

¹⁸Daniel P. Haerther Center for Conservation and Research, Chicago, Illinois, USA

¹⁹St. John's Island National Marine Laboratory, Tropical Marine Science Institute, National University of Singapore, Singapore, Singapore

²⁰State of Hawaii Division of Aquatic Resources, Honolulu, Hawaii, USA

²¹Museo Argentino de Ciencias Naturales "Bernardino Rivadavia"-CONICET, Buenos Aires, Argentina

²²Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame, Notre Dame, Indiana, USA

²³M. Rose Consulting, Haines, Alaska, USA

²⁴Marine Resources Program, Oregon Department of Fish and Wildlife, Newport, Oregon, USA

²⁵California State Lands Commission, Marine Invasive Species Program, Long Beach, California, USA

²⁶Instituto para el Estudio de la Biodiversidad de Invertebrados, Facultad de Ciencias Naturales, Universidad Nacional de Salta, Salta, Argentina

²⁷Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Salta, Argentina

Jose Andrés, Paul Czechowski and Erin Grey contributed equally and names are listed alphabetically.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

²⁸Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, Maryland, USA

²⁹Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, British Columbia, Canada

³⁰Department of Biological Sciences, National University of Singapore, Singapore, Singapore

³¹Lee Kong Chian Natural History Museum, National University of Singapore, Singapore, Singapore

Correspondence

Erin Grey, School of Biology and Ecology and Maine Center for Genetics in the Environment, University of Maine, Orono, Maine, USA.

Email: erin.grey@maine.edu

Funding information

National Science Foundation, Grant/Award Number: #OIA-1849227 and 1748389

Handling Editor: Rosemary Gillespie

Abstract

The spread of nonindigenous species by shipping is a large and growing global problem that harms coastal ecosystems and economies and may blur coastal biogeographical patterns. This study coupled eukaryotic environmental DNA (eDNA) metabarcoding with dissimilarity regression to test the hypothesis that ship-borne species spread homogenizes port communities. We first collected and metabarcoded water samples from ports in Europe, Asia, Australia and the Americas. We then calculated community dissimilarities between port pairs and tested for effects of environmental dissimilarity, biogeographical region and four alternative measures of ship-borne species transport risk. We predicted that higher shipping between ports would decrease community dissimilarity, that the effect of shipping would be small compared to that of environment dissimilarity and shared biogeography, and that more complex shipping risk metrics (which account for ballast water and stepping-stone spread) would perform better. Consistent with our hypotheses, community dissimilarities increased significantly with environmental dissimilarity and, to a lesser extent, decreased with ship-borne species transport risks, particularly if the ports had similar environments and stepping-stone risks were considered. Unexpectedly, we found no clear effect of shared biogeography, and that risk metrics incorporating estimates of ballast discharge did not offer more explanatory power than simpler traffic-based risks. Overall, we found that shipping homogenizes eukaryotic communities between ports in predictable ways, which could inform improvements in invasive species policy and management. We demonstrated the usefulness of eDNA metabarcoding and dissimilarity regression for disentangling the drivers of large-scale biodiversity patterns. We conclude by outlining logistical considerations and recommendations for future studies using this approach.

KEYWORDS

18S, dissimilarity analysis, eDNA, metabarcoding, ports, shipping

1 | INTRODUCTION

Commercial shipping transports roughly 80% of international trade by volume and 70% by value via 100,000 vessels making millions of voyages across thousands of ports each year (UNCTAD, 2021). These voyages comprise a complex network that connects all ports on Earth and redistributes aquatic species among them (Carlton & Geller, 1993). This ship-borne spread of nonindigenous species (NIS) potentially blurs biogeographical boundaries, harms coastal ecosystems (Bax et al., 2003; Carlton, 1996) and causes billions of dollars in damage each year (Cuthbert et al., 2021). Despite decades of research to understand ship-borne species spread, especially via ballast water discharge, several uncertainties remain in our understanding of coastal biodiversity and the relative importance of biotic and abiotic drivers of species transport and establishment success (e.g., Bailey, 2015; Lodge, 1993; Sax et al., 2007). Filling these gaps would enable more efficient use of NIS prevention and management resources and ultimately mitigate the harms of this growing global phenomenon.

Ship-borne species spread is a complex process wherein each voyage potentially transports a diversity of organisms, from bacteria to fish, to new regions where they may or may not establish, depending on the number and state of individuals introduced, environmental conditions, biotic interactions, mitigation efforts or simply chance (Ruiz et al., 2000; Wonham et al., 2013). Further, because individual vessels visit multiple ports, predicting ship-borne species spread may require considering emergent network properties such as higher-order pathways and indirect stepping-stone connections (Apte et al., 2000; Saebi, Xu, Grey, et al., 2020).

The growing availability of data describing global shipping and environmental characteristics enables the modelling of ship-borne species spread at global scales (Drake & Lodge, 2004; Keller et al., 2011; Saebi, Xu, Grey, et al., 2020; Seebens et al., 2013). These models offer the potential to predict global patterns of species spread and invasion hotspots, and to evaluate alternative mitigation strategies at relevant regional scales (Wan et al., 2021; Wang et al., 2020, 2022; Wang, Saebi, et al., 2021; Wang, Silberman, & Corbett, 2021). However, the lack of standardized biodiversity data

at accompanying large scales (Bailey, 2015) has limited testing and refinement of models of global ship-borne species spread, leaving their accuracy and usefulness unknown.

Environmental DNA (eDNA) metabarcoding can help close this biodiversity data gap because it provides robust estimates of community dissimilarity between sites (Coutant et al., 2021; Dully et al., 2021; Grey et al., 2018), and has been successfully used to survey for known NIS in commercial ships and ports, characterize community compositions, and infer biodiversity patterns within and between ballast water and ports (Borrell et al., 2017; Darling et al., 2020; Deiner et al., 2017; Grey et al., 2018; Koziol et al., 2019; Pearman et al., 2021; Rey et al., 2019).

Here we capitalize on eDNA metabarcoding's robust β -diversity estimates to test the hypothesis that shipping homogenizes port communities and to evaluate alternative models of global ship-borne species spread. Our main hypothesis was that all else being equal, ship-borne spread of species lowers community dissimilarity between ports because successful spread results in more species shared between ports. However, we expected that two other factors, environmental dissimilarity and biogeography, would have relatively larger effects than shipping on community dissimilarity between ports at the global scale because they have been influencing coastal biodiversity patterns far longer than shipping. Specifically, we hypothesized that environmental dissimilarity between ports increases community dissimilarity, as ports with different temperature and salinity environments (Keller et al., 2011) will harbour species

with different physiological requirements, and that ports in the same biogeographical region (Costello et al., 2017) will have lower community dissimilarities due to their shared evolutionary history. As its dependent variable, our approach relies on Jaccard and Unifrac community dissimilarity metrics that can be calculated from eDNA sequence data alone. By using these metrics, we avoid the need to identify taxa and their indigenous/nonindigenous status at each port, which is a formidable barrier limiting traditional methods. Similar dissimilarity regression approaches have been used to understand and predict β -diversity across sparsely sampled landscapes (Ferrier et al., 2002; Mokany et al., 2022; Tuomisto & Ruokolainen, 2006), including NIS applications (Capinha et al., 2015), but to our knowledge have not yet been widely applied to aquatic metabarcoding studies (but see Clarke & Deagle, 2020).

We tested our predictions about community dissimilarities of port pairs using a regression model with shipping spread risk, abiotic environmental dissimilarity and biogeographical region as independent variables. Since previous research leaves unclear which shipping risk metrics are most informative for management and policy, we compared models that used four different ways of estimating species transport risk by ships between ports. To calculate these different shipping risk metrics, we created network models where ports are *nodes* which are connected to each other by *edges* that represent the relative risk of species transport along a particular shipping route (Figure 1). Specifically, we considered two methods for estimating the species transport risk for each voyage (either as a

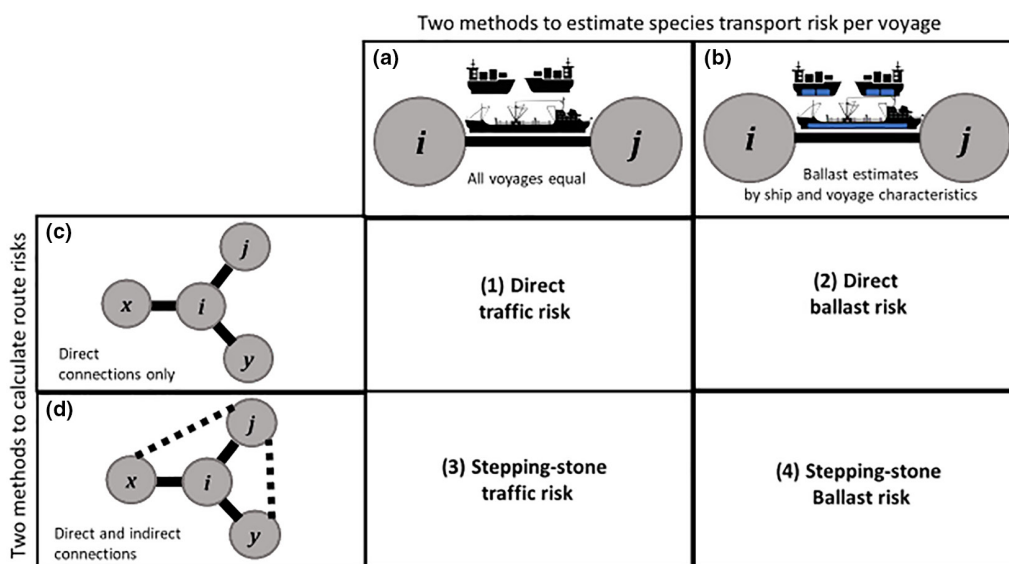


FIGURE 1 Diagram outlining the four ship-borne species transport scenarios tested in this study. In the diagram, hypothetical ports i , j , x and y are nodes (circles) which are connected to each other by edges (lines) that we define as undirected species transport risks between ports (route risks). Initially, we estimated edges as the sum of all voyages along that route (e.g., each voyage has a uniform risk) (a) or as the sum of voyage risks, each of which is calculated as a function of ballast water discharge volume (which is a function of ship type and size) and the likelihood of species survival over the voyage length (b). With these voyage risks, we then estimate route risks (transport risks between two ports) as either a sum of direct voyage risks only (c) or as stepping-stone risks where species could be transported through intermediary ports (d, dotted black lines represented indirect stepping-stone risks between ports i and j). This resulted in four alternative models of ship-borne species spread between routes that we named (1) Direct traffic risk, (2) Direct ballast risk, (3) Stepping-stone traffic risk and (4) Stepping-stone ballast risk. See Methods section for more details on transport risk estimation. Ship icons downloaded from <https://publicdomainvectors.org/en/free-clipart/Ship-silhouette/43320.html> under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication.

uniform value among all voyages or based on ballast discharge and species survival estimates), and then we integrate each set of voyage risk estimates into route risks, which considered only direct connections between two ports or considered direct plus indirect connections between two ports. This process yielded four alternative models for ship-borne species spread which we named direct traffic risk, direct ballast risk, stepping-stone traffic risk and stepping-stone ballast risk (Figure 1).

By evaluating these four alternative scenarios of route transport risks with observed biodiversity data, we aim to determine how much complexity is necessary to capture the dynamics of ship-borne species spread at the global scale. At the scale of individual voyages between ports, studies have documented the importance of ballast discharge and transport duration to successful NIS transport (Seebens et al., 2013). However, when aggregating these risks across all voyages along a route, it remains unclear how much predictive value these complex metrics add over voyage number alone. For route transport risks, most scenarios consider only direct risks of a ship travelling from port to port (Seebens et al., 2013; Saebi, Xu, Grey, et al., 2020). However, here we also consider indirect stepping-stone NIS transport, which has been documented at regional scales (Apte et al., 2000). Stepping-stone spread occurs when a species is introduced from Port A to Port B by one ship, establishes in Port B and subsequently is transported by another ship from Port B to Port C. To our knowledge, this is the first study to directly evaluate the importance of global species transport between ports even when ship traffic does not directly connect the ports.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We analysed five eDNA samples from or adjacent to commercial docks in each of 22 ports across seven biogeographical realms as defined in Costello et al. (2017) including Australia, Southeast Asia, Europe and the Americas. Although the number of unique amplicon sequence variants (ASVs) detected continued to increase beyond five samples (Figure S1), Jaccard (Figure S2) and Unifrac (Figure S3) dissimilarities between ports were insensitive to sample numbers greater than five. To reduce diurnal and tidal variation, all samples were collected during daylight hours at slack high tide after at least a 12-h period without precipitation (Table S1). Following the sampling protocol for commercial ports established by Grey et al. (2018), we collected 250 mL surface water for each sample in a sterile wide-mouth bottle that was stored at 4°C. Within 12 h of sample collection, water was filtered through a 0.45- μ m cellulose-nitrate membrane, immediately immersed in 700 μ L Longmire's buffer, and stored at room temperature for up to 2 weeks and then at -20°C until extraction. If 250 mL could not be filtered through a single membrane, as was the case at some ports with high turbidity and algae, then additional membranes were used and stored in the same microtube. For each site, a collection "blank" sample was taken by

filtering 250 mL of bottled or tap water, which was subsequently treated identically to other samples.

2.2 | eDNA extraction

DNA extractions for the samples and collection blanks taken at the ports of Adelaide, Chicago and Singapore were done using phenol-chloroform in a dedicated polymerase chain reaction (PCR)-free laboratory at the University of Notre Dame as described by Grey et al. (2018). Total DNA from all other samples and collection blanks were extracted in an eDNA facility at Cornell University (UV light for 8 h day⁻¹, HEPA filtered air under positive pressure, personnel wearing full body suits, face-shields and breathing masks) using a Blood and Tissue DNA extraction kit (Qiagen) protocol optimized for extracting eDNA from cellulose nitrate filters preserved in Longmire's buffer (detailed in Spens et al., 2017, Figure 1). We tested for the impact of these two different methods on our results in two ways (see Dissimilarity regression modelling section below).

2.3 | Library preparation and sequencing

18S rRNA amplicon libraries were prepared using two primers targeting the eukaryotic V4 region (18S_574F and 18S_952R) that were previously evaluated by Hadziavdic et al. (2014). For 19 ports (all but Adelaide, Chicago and Singapore), we followed a single-step amplification approach using a single set of primers including Illumina adapters and a 12-bp barcode unique to each sample (Figure S4). Three-fold PCR amplifications were carried out for all samples to account for amplification bias. Each reaction of 25 μ L total volume contained 2 μ L template DNA, 1.5 mM MgCl₂, 1 \times Colourless GoTaq Flexi Buffer (Promega), 0.25 mM of each dNTP, 0.2 μ M of forward and reverse primer and 1.25 units GoTaq Hot Start DNA Polymerase (Promega). After initial denaturation at 95°C for 2 min, we performed 40 amplification cycles (95°C for 45 s, 49°C for 1 min, 72°C for 1 min) and a final elongation at 72°C for 5 min. Amplicon triplicates were pooled, visualized on 2% agarose gels and then purified using Mag-Bind Total Pure NGS beads (Omega Bio-Tek). DNA concentration for all samples was quantified using PicoGreen reagent (Thermo Fisher Scientific) and a Spectramax M2 multi-detection microplate reader (Molecular Devices). Libraries were prepared by combining samples in equimolar ratios, and pair-sequenced (2 \times 250 bp) on an Illumina MiSeq platform.

18S rRNA amplicon libraries for the ports of Adelaide, Chicago and Singapore were generated using a two-step protocol described in Grey et al. (2018). The first PCR amplified target fragments from the eDNA samples using the Hadziavdic 18S_574F and 18S_952R primers with Illumina Nextera adaptor overhangs in 50 μ L PCR consisting of 27 μ L H₂O, 10 μ L 5 \times HiFi buffer, 1.5 μ L 50 mM MgCl₂, 1 μ L 10 mM dNTPs, 2.5 μ L of 10 μ M primer-F, 2.5 μ L of 10 μ M primer-R, 0.5 μ L Taq DNA Polymerase (iProof, Bio-Rad) and 5 μ L DNA. The thermocycling protocol for this first PCR started with an initial

denaturation at 98°C for 2 min, followed by 25 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 20 s, extension at 72°C for 30 s and a final extension at 72°C for 10 min. Amplicons were run on a 2% agarose gel, band cut, and cleaned using a QIAquick Gel Extraction Kit (Qiagen) and eluted with 25 µL EB. The second PCR attached the remaining adaptor sequence and library-specific i5/i7 indices to the amplicons. This 50-µl PCR consisted of 22 µL sterile water, 10 µL 5× HF buffer, 1.5 µL 50 mM MgCl₂, 1 µL 10 mM dNTPs, 5 µL of 10 µM Nextera ID (forward primer), 5 µL of 10 µM Nextera ID (reverse primer), 0.5 µL iProof Taq (Bio-Rad) and 5 µL DNA template (from band-cut and cleaned first PCR). Thermocycling consisted of initial denaturation at 98°C for 2 min, and eight cycles of denaturation at 98°C for 10 s, followed by annealing at 60°C for 20 s. Amplicons were sized-selected with AMPure bead clean-up at a bead to volume ratio of 0.8:1. DNA concentrations were estimated with a Qubit HS Assay Kit (Life Technologies). Libraries were prepared by combining samples in equimolar ratios and paired-end sequenced (2×300 bp) on an Illumina MiSeq platform.

2.4 | Negative and positive controls

Alongside port sample DNA libraries, we generated libraries for three different types of negative controls (collection, extraction and amplification controls) and three positive controls: one containing genomic DNA of the freshwater fish *Danio rerio*, and two equimolar mixtures of either four freshwater fishes (*Notropis topeka*, *Noturus taylori*, *Umbra limi*, *Thoburnia atripinnis*) or six marine fishes (*Pseudanthias dispar*, *Ecsenius bicolor*, *Macropharyngodon negrosensis*, *Centropyge bispinosa*, *Salarius fasciatus*, *Amphiprion ocellary*; Figure S5). Reference sequences for all species listed above were generated using genomic DNA and Sanger sequencing on an ABI 3730xl. Except for *M. negrosensis* (the sequencing of which failed), we used reference sequences from these species to inform taxonomic assignment.

2.5 | Sequence data processing

Amplicon sequence variants were generated from combined 18S amplicon sequence data with QIIME version 22019-04 (Bolyen et al., 2019), after using BBTOOLS version 37.44 (Bushnell, 2014) to include Adelaide and Singapore sequence data from Grey et al. (2018). To do so, reads were initially imported, paired and filtered using QIIME's import functions with default parameters. Read pairs were then subjected to an additional round of adapter trimming using CUTADAPT version 1.18 (Martin, 2011), denoised using DADA2 version 1.6.0 (Callahan et al., 2016) and an Expected Error value of 9 (Edgar & Flyvbjerg, 2015). Prior to merging, reads were trimmed to 220 nucleotides. The resulting unique ASVs were filtered using QIIME's taxonomic classifier Vsearch (Rognes et al., 2016) to obtain eukaryotic ASVs as follows. We appended our positive control reference sequences to the QIIME release of the SILVA database version 132 (Pruesse et al., 2007), and we then ran an in silico experiment to

optimize the assignment settings of Vsearch. Specifically, we optimized the Vsearch classifier to unequivocally identify the positive control species and to have maximum correspondence with taxonomy assignments obtained using BLAST+ version 2.8.-alpha (Camacho et al., 2009) and the March 2018 NCBI nt/nr database (Benson et al., 2018). We then used the optimized Vsearch classifier and the SILVA database version 132 to identify and retain all eukaryotic ASVs ($-p\text{-per-identity} = 0.875$, $-p\text{-min-consensus} = 0.5$, $-p\text{-query-cov} = 0.9$).

To account for contamination in the field and laboratory, we inspected and subtracted the small number of ASVs present in the negative controls (see blanks indicated in Figure S5d) from all subsequent datasets. Sequence run read outputs and read counts at each bioinformatics stage are shown in Figure S5.

We generated two data sets based on different criteria for read depth based on ASV accumulation by read curves (Figure S6). One data set included 19 ports ("deep rarefaction"; without the ports of Chicago, Nanaimo and Vancouver), and one contained 21 ports ("shallow rarefaction"; without the port of Chicago). Regression modelling results between these two data sets were comparable, so we present the results from the deeply rarefied data set in the main text and provide shallowly rarefied data set results in the Supporting Information.

2.6 | Estimation of biological dissimilarity between port pairs

For both deep and shallow rarefaction data sets, we calculated two unweighted β diversity metrics between port pairs from 18S ASVs: Jaccard (Jaccard, 1912) and Unifrac dissimilarity (Lozupone et al., 2011). While the Jaccard metric measures the total fraction of unshared ASVs, the Unifrac metric estimates the fraction of total unshared branch lengths when the port pairs are placed on a given phylogenetic tree. Thus, different port pairs with the same fraction of unshared ASVs will always have the same Jaccard dissimilarity. However, their Unifrac dissimilarity will differ depending on the overall relatedness of the ASVs. The less related the unshared ASVs between ports are, the bigger the fraction of unshared branch lengths is, and larger the Unifrac dissimilarity. Ship-borne species spread can, therefore, have a different impact on these two metrics. For example, introducing a single species from one port to another will probably have a limited effect on the Jaccard dissimilarity between them, but it might have a relatively large impact in the Unifrac dissimilarity if the introduced species is phylogenetically distinct from existing species in the recipient port.

We calculated Jaccard dissimilarities between all port pairs using the *vegdist* function in the VEGAN 2.6-4 package (Oksanen et al., 2022) after pooling ASV read counts from five samples per port. We calculated Unifrac dissimilarities for all sample pairs using QIIME version 22019-04 and then averaged five Unifrac sample-pairwise dissimilarities per port to obtain the average port-pairwise dissimilarities. This average pairwise dissimilarity approach follows the recommendation of Marion et al. (2017) to reduce bias when comparing communities of different sizes. To calculate Unifrac dissimilarities,

we first obtained reference trees by aligning eukaryote ASVs using MAFFT version 7.3922018/3/3 (Kato & Standley, 2013) with automatic algorithm and parameter selection. This initial alignment was then post-processed to retain only consenting columns supported by at least 50% of ASVs and allowing only columns with <10% gaps across the entire alignment, resulting in alignments 252 bp in length (original alignments were ~380 bp, Figure S4). Finally, reference trees were obtained using FASTTREE 2.1.8 (Price et al., 2010) with default settings, and unweighted Unifrac dissimilarities were calculated between all port pairs. As expected, Unifrac and Jaccard dissimilarities were positively correlated, with Unifrac estimates having a wider range than the Jaccard estimates (Figure S7, $R^2 = .622$, $p < .001$).

2.7 | Estimates of ship-borne species transport risks

We used 8 years of shipping data, spanning 1997–2018 and obtained from Lloyd's List Intelligence, an Informa Group Company (LLI), to create two higher-order networks from each of which we derived two estimates of species transport risk between ports (Figure 1). The two networks differed only in how they estimated initial edge weights: one where the estimated edge weights were based solely on the number of voyages between two ports ("traffic-risk") and the other where edge weights were estimated with a function that weighted each voyage by its ballast transport risk ("ballast-risk") (Figure 1). Ballast transport risks were estimated for each voyage v between ports i and j using a formula introduced by Seebens et al. (2013):

$$P(\text{transport})_{ij}^v = \left(1 - e^{-\lambda D_{ij}^v}\right) e^{-\mu \Delta t_{ij}^v} \quad (1)$$

$$D_U = \alpha + \beta_1 ENV + \beta_2 BGR + \beta_3 SHP + \beta_4 (ENV * BGR) + \beta_5 (ENV * SHP) + \beta_6 (BGR * SHP) + \mu (PortO)\gamma + \mu (PortD)\gamma + \epsilon \quad (3)$$

where D_{ij}^v is the amount of ballast water discharged at the destination port, λ is the species introduction potential per volume of discharge, μ is the daily mortality rate of species in ballast water and Δt_{ij}^v voyage duration in days. Following Xu et al. (2014), the amount of ballast water transported between ports D_{ij}^v is estimated based on ship type and ship gross weight tonnage, $\mu = 0.02$, and $\lambda = 3.22 \times 10^{-6}$.

Since many ships do not discharge water at their first port of call, we created higher-order networks for each model to account for this non-Markovian transport process prior to estimating final route transport risks. For both traffic and ballast transport risks, significantly repeated multiport voyages, called higher-order dependencies, were determined using the algorithm developed by Saebi, Xu, Kaplan, et al. (2020). For each higher-order dependency identified, we then calculated the total risk of direct species transport along the path and then used the algorithm of Xu et al. (2016) to create the higher-order network of species transport. Since several nodes in this higher-order network can map to a single physical port, we then built the physical adjacency matrix in which each edge weight (species transport risk) is calculated by averaging over all higher-order

network edges that correspond to that pair of ports. Finally, we normalized the edge weights by dividing each edge weight by the maximum value in the network to obtain the final direct route species transport risks for networks built using traffic-risks or ballast-risk edge weights (Figure 1). While these estimates of direct traffic and direct ballast risks account for higher-order species transport between two ports connected by shipping (which are based on repeated ship movements), they do not include stepping-stone species transport risks (between ports that are not directly connected by ship traffic).

To test for potentially significant effects of stepping-stone risk on patterns of global ship-borne species spread, we calculated the stepping-stone risk as the Jaccard similarity score for each pair of ports. This score is equal to the number of shared directly connected ports normalized by the union of directly connected ports:

$$J(i, j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|} \quad (2)$$

where N_i and N_j are the set of directly connected ports for ports i and j , respectively.

2.8 | Dissimilarity regression modelling

A critical first step in understanding the role of shipping traffic in the homogenization of marine biota is the selection of an appropriate set of models covering reasonable hypotheses. Marine species distributions are known to be influenced by temperature, salinity and biogeography. Therefore, we considered this additive dissimilarity regression model:

in which D_U is either the Jaccard or Unifrac dissimilarity between two ports, the ports of origin ($PortO$) and destination ($PortD$) are random effects, shipping transport risk (SHP) and environmental dissimilarity (ENV) between two ports are fixed factors, biogeographical dissimilarity between two ports (BGR) is a binary fixed effect that reflects if the ports of origin and destination are in the same marine realm as defined by Costello et al. (2017) or not, and a "*" indicates an interaction effect. Environmental dissimilarity (ENV) was estimated as Euclidean distance between mean annual temperature, mean annual salinity and annual temperature range (Keller et al., 2011). Shipping transport risk (SHP) was one of the four risks described in the previous section: direct traffic-based risk (SHP_{dt}), direct ballast risk (SHP_{db}), stepping-stone traffic risk (SHP_{st}) and stepping-stone ballast risk (SHP_{sb}) (Figure 1).

To test for effects of alternative DNA extraction and library preparation methods used for Adelaide and Singapore samples, we took two approaches. First, we created and analysed models with a random "methods" effect, which was a binary term reflecting whether port pair samples were processed with identical laboratory

methods or not. Second, we ran the Jaccard model with and without the two ports (Adelaide and Singapore) with methods different from the other ports.

These complex models initially considered all additive effects and potential effects of all the two-way interactions. To avoid over- or under-fitting, we used a backward selection to identify optimal model structure. Statistical modelling and model selection were done separately for the Jaccard and Unifrac dissimilarity responses using the R PROGRAMMING LANGUAGE VERSION 3.4.3 (R Core Development Team, 2019). Because the observed Jaccard dissimilarities between port-pairs were highly biased towards values near one (skewness = -3.20, kurtosis = 15.85) we first fitted to a One-Inflated Beta model (family BEOI) and obtained the final generalized additive model for location, scale and shape (GMLSS) using mu (μ) selection as implemented in the “stepGAIC” function of the package GAMLSS (Rigby & Stasinopoulos, 2005). To account for any potential collinearity (concurvity) effect on the models' interactions, we performed all analyses using centred and untransformed response variables. For the Unifrac dissimilarities, final Gaussian models were obtained using stepwise regression using the “glmselect” function and the LME4 package version 1.1-16 (Bates et al., 2015).

Finally, we performed backward-selection on a dissimilarity regression model without BGR (biogeography) as an independent variable for two reasons. First, a priori we knew that ports have similar habitat structure globally and that local habitat conditions may override the distinctiveness of biogeographical realms at this scale (Costello et al., 2017). Second, post hoc we found the distribution of taxa that we sampled was different from that which Costello et al. (2017) used to define biogeographical realms (Figure S8).

2.9 | NIS identification

Although not the primary focus of this study, species identification with binomial nomenclature is still a critical component of shipborne NIS research and management. Therefore, we assigned taxonomy to ASVs with BLAST version 2.9 (Camacho et al., 2009), to the April 2019 NCBI nr/nt reference database (Benson et al., 2018), excluding subject sequences solely identified as “environmental sample” and using a minimum e-value of 10^{-5} . For each query-subject pair, we kept the top five high-scoring alignments (>98.5% identity) and excluded ambiguous assignments where more than one species matched equally well to the query sequence. Unambiguously identified species were then matched to the World Register of Introduced Marine Species, WRiMS (Ahyong et al., 2022) and the species ports' presence data were contrasted with the known distributions of these species according to the World Register of Marine Species, WoRMS (WoRMS Editorial Board, 2019). The R packages PHYLOSEQ 1.22.3 (McMurdie & Holmes, 2013) and TIDYVERSE 1.2.1 (Wickham, 2017) were used to import ASV data and BLAST results. We used TAXONOMIZR 0.2.2 (Sherrill-Mix, 2019) to obtain taxonomic information of the best BLAST hits.

3 | RESULTS

The locations of the 19 ports yielding deep rarefaction sequence data spanned seven of the 30 marine biogeographical realms defined by Costello et al. (2017), with the majority situated in only three realms (North Pacific, Caribbean and Gulf of Mexico, and Northeast Atlantic, Figure 2a; see Figure S9 for similar Unifrac dissimilarity plot). The 11,975 eukaryotic ASVs (Callahan et al., 2016) we obtained represented taxa that are predominantly planktonic as well as taxa that are predominantly benthic (Figure S10), although it would be impossible to confidently assign any specific ASV to either category. Port-pairwise Jaccard dissimilarities ranged from 0.78 to 1 (mean: 0.98), while Unifrac dissimilarities ranged from 0.48 to 0.93 (mean: 0.77) (Figure 2b; Figure S7). Port-pairwise values of our four shipping risk estimates correlated highly with each other ($R \geq .78$ and $p < .001$ for all correlations, Figure S11), while stepping-stone traffic and ballast risks were both slightly negatively correlated with environmental dissimilarity ($R = -.18$ and $p = .02$, Figure S12). Because results for Jaccard and Unifrac models were very similar and supported the same conclusions, for simplicity we present below some results for only Jaccard (and provide parallel results for Unifrac models in the Supporting Information).

For both Jaccard and Unifrac metrics calculated from deeply and shallowly rarefied data sets, dissimilarity regression modelling results supported our prediction that increasing environmental dissimilarity and decreasing shipping transport risks would increase community dissimilarity among ports. All backward-selected models identified significant effects of environmental dissimilarity on community dissimilarity between ports (Table 1; see Table S2 for selection statistics); port pairs with similar environments tended to have smaller Unifrac ($r = .368$, $p < .0001$) and smaller Jaccard ($r = .432$, $p < .001$) metrics (i.e., more similar communities; Figure S13). Our prediction that shipping would reduce community dissimilarity was also supported: all four Jaccard and three of the Unifrac Akaike information criterion (AIC)-selected models included shipping as an explanatory variable (Table 1). Our prediction that shared biogeography would reduce community dissimilarity was not strongly supported: port communities within the same biogeographical realm tended to have lower community dissimilarities than those in different realms (Figure S14), but this effect was only statistically significant in the direct risk models. When checking for multicollinearity between predictor variables that could complicate model fitting and interpretation, we found weak negative correlations between environmental dissimilarities and shipping risk in both stepping-stone models (Figure S12, $R = -.18$, $p = .02$), but otherwise predictor variables were not significantly correlated (see Figures S12 and S15).

Consistent with our prediction, models using stepping-stone connection risks outperformed models using direct connection risks. Stepping-stone models were, on average, 8 AIC units lower for the Jaccard metric, and 10 AIC units lower for the Unifrac metric (Table 1). Contrary to our expectations, including ballast either did not improve the stepping-stone model (Jaccard) or worsened it by 1 AIC unit (Unifrac) (Table 1). Including ballast improved only

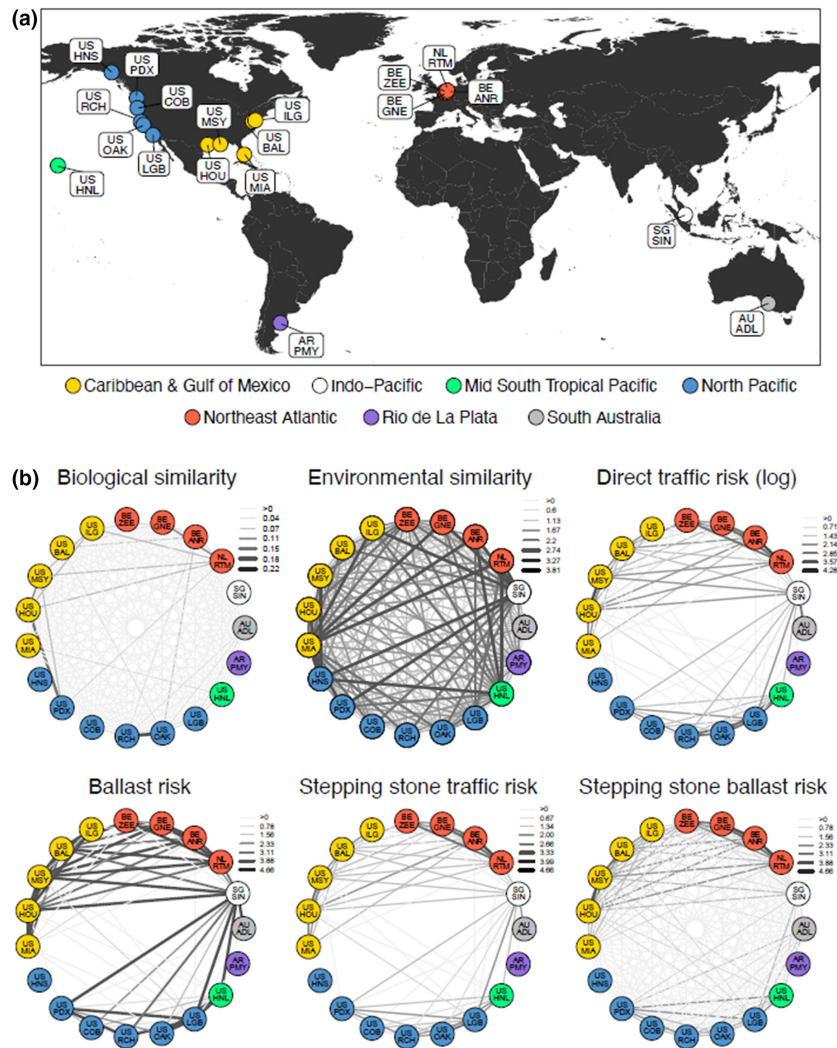


FIGURE 2 Sampling extent and variable summary for the analysis of between-port Jaccard dissimilarities. (a) Sampled ports, coloured by biogeographical realms (Costello et al., 2017), in the “deep rarefaction” data set included Adelaide (AU-ADL), Antwerp (BE-ANR), Baltimore (US-BAL), Coos Bay (US-COB), Ghent (BE-GNE), Haines (US-HNS), Honolulu (US-HNL), Houston (US-HOU), Long Beach (US-LGB), Miami (US-MIA), New Orleans (US-MSY), Oakland (US-OAK), Portland (US-PDX), Puerto Madryn (AR-PMY), Richmond (US-RCH), Rotterdam (NL-RTM), Singapore (SG-SIN), Wilmington (US-ILG) and Zeebrugge (BE-ZEE). Port codes follow the United Nations Code for Trade and Transport Locations. (b) Summaries of network-like relationships between ports expressing biological similarity, environmental similarity and different shipping-risk estimates (see Figure 1). To allow for direct comparison with the ship-borne risk estimate patterns, we inverted biological and environmental dissimilarities by subtracting each dissimilarity value from either the theoretical (1 for Jaccard dissimilarities) or observed (4.36 for environmental dissimilarity) maximum. Line colour and width are scaled according to each metric, where thick black lines represent the highest similarities or strongest links and thin light-grey lines the lowest similarities or weakest links.

the model predicting Jaccard dissimilarity with direct traffic risks (Table 1).

The simplest model predicting Jaccard dissimilarity between ports included environmental dissimilarity, biogeographical realm, stepping-stone traffic risk, and two-way interactions between stepping-stone traffic risk and environment and between stepping-stone traffic risk and biogeographical realm (Table 1, first row, with AIC tied for lowest among-Jaccard models). The simplest model predicting Unifrac dissimilarity between ports included only environmental dissimilarity, stepping-stone traffic risk and the interaction of these two variables (Table 1, second row, with lowest AIC). For both Jaccard dissimilarity (Figure 3) and Unifrac dissimilarity (Figure S16),

stepping-stone traffic risk decreases biological dissimilarity, particularly on routes whose origin and destination are environmentally similar. This interaction is consistent with our expectation that introduced organisms are unlikely to survive in vastly different environments, regardless of how many times they are introduced by ships.

Jaccard, but not Unifrac, dissimilarity between ports is also partially predicted by the interaction between stepping-stone traffic risk and biogeographical realm. For Jaccard dissimilarity, stepping-stone traffic risk reduces dissimilarity, as expected, when the two ports are in different biogeographical realms. However, unexpectedly, the opposite is true if the two ports are in the same biogeographical realm (Figure S17); we would have predicted a negative, but smaller, effect

TABLE 1 Dissimilarity regression model summaries.

Ship risk metric	AIC-selected model	AIC
Stepping-stone traffic	JACCARD = ENV + BGR + SHP _{st} + ENV*SHP _{st} + BGR*SHP _{st} + (ORG) + (DEST)	-1089
	UNIFRAC = ENV + SHP _{st} + ENV*SHP _{st} + (ORG) + (DEST)	-465
Stepping-stone ballast	JACCARD = ENV + BGR + SHP _{sb} + ENV*SHP _{sb} + BGR*SHP _{sb} + (ORG) + (DEST)	-1089
	UNIFRAC = ENV + SHP _{sb} + ENV*SHP _{sb} + (ORG) + (DEST)	-464
Direct traffic	JACCARD = ENV + BGR + SHP _{dt} + BGR*SHP _{dt} + (ORG) + (DEST)	-1077
	UNIFRAC = ENV + BGR + ENV*BGR + (ORG) + (DEST)	-457
Direct ballast	JACCARD = ENV + BGR + SHP _{db} + BGR*SHP _{db} + (ORG) + (DEST)	-1085
	UNIFRAC = ENV + BGR + SHP _{db} + BGR*ENV + BGR*SHP _{db} + (ORG) + (DEST)	-454

Note: Parameters and Akaike information criterion (AIC) score for the four dissimilarity regression models for Jaccard and Unifrac dissimilarities. Initial model parameters included environmental similarity (ENV), shipping risk (SHP), biogeographical realm (BGR) and all pairwise interactions indicated by a "*". Ports of origin (ORG) and destination (DEST) were modelled as random factors. For each biological dissimilarity metric (Jaccard and Unifrac), backwards selection was used to determine the best minimum adequate model; the best performing models with the simplest formulation of shipping risk are highlighted in grey.

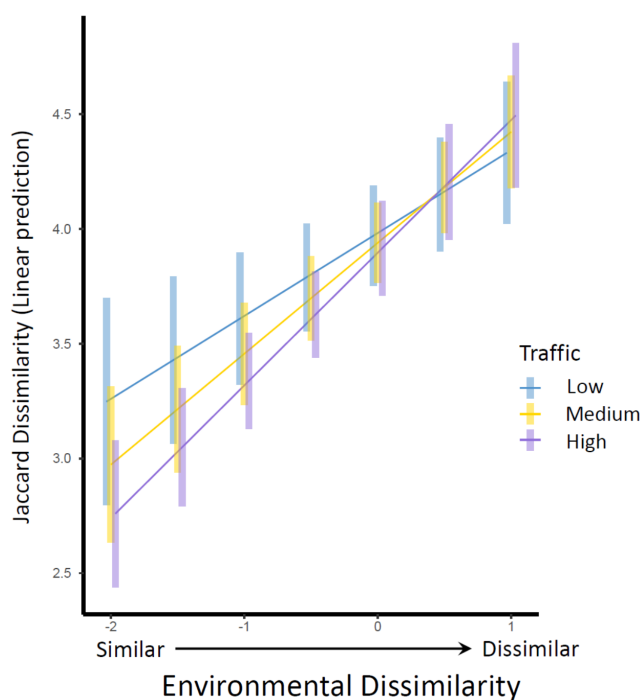


FIGURE 3 Interactive effect of stepping-stone traffic risk and environmental dissimilarity on Jaccard dissimilarity. Biological dissimilarity between ports is best explained by the interaction of stepping-stone traffic risks and the environmental dissimilarity between the ports of origin and destination (Table 1). Medium = mean; Low = mean - 1SD; High = mean + 1SD. Bars represent the $\pm 95\%$ confidence interval.

relative to connectivity between different biogeographical realms, or no effect at all if biogeographical realms are internally homogeneous. When biogeographical realm was not included as a dependent variable, results were little changed: biological dissimilarity was best predicted by environment, stepping-stone traffic risk and the interaction between those two variables (Table S3).

As expected, among the independent variables, environmental dissimilarity had the largest effect on both Jaccard and Unifrac

TABLE 2 Parameter estimates ($\pm SE$) of the models best predicting biological dissimilarity (lowest AIC) and having the simplest shipping metric (stepping-stone traffic risk) (Table 1, grey rows).

Dissimilarity metric	Parameter	Estimate
Jaccard	Intercept	3.687 (± 0.020)
	ENV	0.557 (± 0.075)
	SHP _{st}	0.263 (± 0.132)
	BRG	0.262 (± 0.204)
	ENV*BRG	-0.588 (± 0.016)
Unifrac	Intercept	0.767 (± 0.011)
	ENV	0.030 (± 0.005)
	SHP _{st}	-0.003 (± 0.005)
	ENV*SHP _{st}	0.013 (± 0.004)

Note: Model parameters included environmental similarity (ENV), traffic risk (SHP_{st}) and biogeographical realm (BGR); * indicates an interaction effect.

dissimilarity, and coefficients associated with shipping risk were smaller than those associated with environmental similarity (Table 2).

Our dissimilarity regression modelling results were robust to sequencing depth and laboratory methods. Analogous analyses adding two more ports (Nanaimo, Vancouver) at a shallower sequencing depth yielded similar results, except with biogeographical realms also emerging as an additional explanatory variable for Unifrac dissimilarity (Table S4). Likewise, considering different laboratory methods as a random effect yielded results similar to those without the random effect, although some Unifrac regressions with laboratory methods as a random effect resulted in singular fits, which indicate overparameterization (Table S2). Finally, dropping Adelaide and Singapore (the two ports with different laboratory methods) from the analysis also did not meaningfully change results (Table S5).

The taxa we detected overlapped with those on which Costello et al. (2017) based their biogeographical realm boundaries, but the

distribution among taxa was very different (Figure S8), confirming one of several reasons to suspect that the biogeographical realms as defined by Costello et al. (2017) may not be as relevant to our results as initially hypothesized.

Although species identification was not the focus of this study, from all the unique ASVs identified (Table S6), we inspected the subset of ASVs with taxonomic annotations matching species in the WRiMS (Ahyong et al., 2022) to explore known NIS distributions across our data set. We identified 273 ASVs of which 79 (28.9%) matched equally well to more than one nominal species, indicating possible reference database errors or a lack of taxonomic resolution with this primer set. The remaining 194 taxonomically unambiguous ASVs represented 11 classes and 57 unique species (Table S7), with the solitary tunicate *Ciona savignyi* (41 ASVs) and reef-forming polychaete *Ficopomatus enigmaticus* (42 ASVs) comprising over 40% of all reads for these ASVs. Two of these 57 species were found in non-native areas where they had not been previously reported (Table S7): *Pseudocalanus elongatus* (Copepoda) at Coos Bay, and *Botrylloides leachii* (Ascidacea) at Richmond and Oakland. One species, *Oithona davisae* (Copepoda), was found slightly outside of its reported native range in Singapore. All other species were found within their previously reported native and introduced ranges.

4 | DISCUSSION

Our analysis of community dissimilarities among ports quantified the relative effects of shipping and the environment on biological homogenization at intercontinental scales. Our hypotheses that shipping would reduce community dissimilarity between ports and that environmental dissimilarity would have the relatively largest effect on community dissimilarity were supported for both Jaccard and Unifrac dissimilarity metrics (Tables 1 and 2; Figure 3). The significant interaction between shipping and environment suggests that high-volume shipping routes connecting environmentally similar ports pose the highest risk for species spread. These results clearly support holistic consideration of shipping, environment and their interaction for predicting and managing ship-borne NIS. The environment–shipping interaction is particularly important to consider when predicting route risks under future climate change, which will probably alter port environments and native ranges, subsequently altering predictions about which port-pair routes to prioritize for management.

The lack of a strong effect of biogeographical realms on port community dissimilarities was unexpected, and could result from multiple, nonmutually exclusive causes. First our ports represent substantial geographical bias (Figure 1), with a relatively low number of port-pairs occupying different versus the same biogeographical realms. Second, ports represent novel and globally similar habitats which support communities that may be more similar to each other than to surrounding habitats (Bulleri & Airoldi, 2005; Forrest et al., 2013; Lambert & Lambert, 2003; Megina et al., 2016; Piola & Johnston, 2008), blurring the signals of the biogeographical regions in which they are embedded. Indeed, Costello et al. (2017) noted that habitat effects are strong and

may obscure biogeographical realm effects. Third, the biogeographical delineations that we used (Costello et al., 2017) are the best available but have limitations with respect to the goals of our study. They are based on data that do not adequately represent all taxa and may not be appropriate for the taxa we sampled (e.g., Figure S8 shows that our 18S data set detected relatively more nonmetazoan eukaryotes than considered by Costello et al., 2017). Indeed, it is possible that better delimitations of biogeographical realms will result from future extensive eDNA sampling. Finally, the limited temporal scope of our study (one sampling event per port) may not reflect β -diversity patterns over longer timescales (e.g., months to years) that might more reliably reflect biogeographical patterns. Additional studies including more ports from more realms, globally extensive eDNA sampling inside and outside of ports, and sampling throughout the year will be needed to test these hypotheses. In hindsight, it is not surprising that biogeographical realm was a weak predictor of between-port dissimilarity in our study.

Comparison of the four different models of ship-borne species spread risk partially supported our hypothesis that more complex metrics for shipping risks would perform better: stepping-stone models performed better than direct traffic models, but models incorporating ballast risks did not perform better than simpler risk metrics based on voyage counts alone. While the potential importance of stepping-stone connections was not surprising given earlier work (Apte et al., 2000; Floerl et al., 2009), the superior performance of models not incorporating ballast was unexpected. Although it is generally accepted that differences among ship types and related differences in ballast discharge result in unequal risks of species transport (Davidson et al., 2018; Drake & Lodge, 2004), the addition of ship type and voyage length in our ballast risk models did not significantly improve predictions of biological dissimilarity between ports. While it is appealing to think that the simpler metrics not incorporating ballast adequately capture risk, it is also possible that other ballast-related factors not included in our ballast parameter or species spread from other vectors (e.g., biofouling, aquaculture) not considered here obscure the ballast discharge effect. The development and evaluation of additional global ship-borne species spread models with more detailed risk estimates are needed to determine if this is the case.

While our study was not designed to identify specific NIS, we found that eDNA metabarcoding of even a few samples is useful for passive NIS surveillance at an intercontinental scale. For species monitoring, the current drawbacks of eDNA applications are their inaccuracy in assigning taxonomy to some sequences when reference data are deficient, and their susceptibility to biased abundance estimates (Balvočiūtė & Huson, 2017; Kelly et al., 2019). Regardless, from just five samples per port we derived sequence variants highly similar or identical to many known NIS and discovered potential range expansions of three species (*Pseudocalanus elongatus*, *Botrylloides leachii* and *Oithona davisae*), and identified population genetic diversity among ports for two species (*Ciona savignyi*, *Ficopomatus enigmaticus*). As other studies have shown, increased sampling effort across water depths would further increase the number of NIS detected (e.g., Koziol et al., 2019). Further development of eDNA techniques may soon enable population genetic inquiries (Sigsgaard et al., 2016; Andrés

et al., 2021) and possibly abundance estimates (Spear et al., 2021) in and among global ports. We anticipate that with growing genetic reference libraries, eDNA metabarcoding will become increasingly useful for species identification and monitoring in ports.

While eDNA metabarcoding offers many logistical benefits over traditional methods for large-scale, standardized biodiversity surveys, we did encounter difficulties in this study. First, controlling for tidal and weather variables in dynamic coastal habitats globally is logistically difficult given that some regions have prolonged rainy seasons and the interaction between diurnal and tidal cycles vary so that in some areas certain tidal periods occur only at night. Time-series surveys at each port would allow for estimating and controlling for some of these nontarget effects (and would also help capture a larger proportion of the biodiversity) while increasing effort and costs. Recent advances in automating eDNA sampling (e.g., Formel et al., 2021), laboratory processing (Buchner et al., 2021) and bioinformatics (e.g., Mousavi-Derazmahalleh et al., 2021) may alleviate these current constraints.

We also encountered difficulties shipping materials and samples between countries. While our sampling kits were small and did not contain any toxic or flammable substances, they were sometimes delayed or even refused by customs offices of multiple countries. Furthermore, confusion over how the Nagoya Protocol pertains to eDNA samples prevented our ability to obtain a material transfer agreement between some of our team's organizations (see Deplazes-Zemp et al., 2018 for more on this issue), dramatically reducing the number of ports and biogeographical scope of this study. We recommend that future global eDNA metabarcoding surveys carefully consider these issues prior to survey design, and that research funders provide support adequate for the administrative requirements of permitting. A global effort, in the context of the Convention on Biological Diversity, to clarify the legal and ethical considerations of international transport of eDNA is urgently needed (Lodge, 2022). One of the great strengths of eDNA metabarcoding is, in our opinion, its potential to rapidly and comprehensively survey regions that have been historically under-sampled (Czechowski et al., 2023). It is, however, essential that we do so in a legal, ethical and equitable manner.

Overall, this study shows that dissimilarity analysis of eDNA metabarcoding can reveal global biodiversity patterns and help us better understand their drivers. As expected, we found a strong effect of environmental similarity and a smaller but still significant effect of shipping risk on global port β -diversity patterns. This knowledge could help predict and mitigate ship-borne NIS spread in a world depending on shipping for the transport of goods in a rapidly changing climate. Currently, global ballast water policy takes a "one-size-fits-all" approach relying on onboard ballast treatment systems, and global biofouling policy remains under development (Tamburri et al., 2021). Consistent with our other studies (Wan et al., 2021; Wang et al., 2020, 2022; Wang, Saebi, et al., 2021; Wang, Silberman, & Corbett, 2021), the results reported here suggest that prioritizing "risky" routes with targeted management and biomonitoring, or establishing land-based treatment systems at hub ports, would be a more efficient way to reduce global ship-borne species spread. Further development and

refinement of global ship-borne species spread models with eDNA metabarcoding and dissimilarity analysis, as we have begun here, would enable such risk-based policies.

Although our study's biased biogeographical scope limits our inferences, our results illustrate that environmental dissimilarity, shipping and their interaction reduce biological dissimilarity among commercial port habitats. This finding supports the use of routed-based models of the risks of ship-borne species spread in evaluation of policy and management. Future advances in methods, techniques and genetic reference libraries, together with collective efforts to overcome current barriers to globally coordinated surveys, will increase the ability of eDNA metabarcoding to inform the patterns, processes and conservation of global biodiversity.

AUTHOR CONTRIBUTIONS

Conceptualization: EKG, MP, DML. Methodology: PC, JA, EKG, DML, MS. Investigation: PC, JA, KA, EKG, DML, MS. Visualization: JA, KA, PC, EKG. Supervision: JA, DML, NC. Writing—original draft: PC, JA, EKG, DML. Writing—review & editing: all authors.

ACKNOWLEDGEMENTS

For helpful discussions and/or logistical collaboration we thank Kimberly Howland, Anna Lacoursiere, Yiyuan Li, and Aibin Zhan, Demetrio Boltovskoy, Pedro Brandão, Juul Lorre, Dias F. Pinto, Kristy Deiner, Vanessa Hodes, Jill Remeysen, Lindsay Schaffner, Evangelina Schwindt, Peter Avis and Giles Hooker.

FUNDING INFORMATION

This work was funded by NSF Coastal SEES Award no. 1748389 awarded to Lodge, Grey, Corbett and Chawla. Grey was also supported by National Science Foundation award no. OIA-1849227 to Maine EPSCoR at the University of Maine.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

All data, code and materials used in the analyses are freely available electronically. Sequences are available in the NCBI Sequence Read Archive (BioProject ID PRJNA930753). Code used for analyses presented in this paper, as well as intermediate analyses, is available at <https://doi.org/10.5281/zenodo.7602244> as a stable release of the corresponding GitHub repository available at https://github.com/macrobiothus/ships_and_bugs.

BENEFIT SHARING

The research described in this publication complies with relevant national laws implementing the Convention on Biological Diversity and Nagoya Protocol agreements. The researchers who participate in the port eDNA survey and analyses were included as co-authors, data have been shared with all co-authors prior to publication, and research data and findings are being made publicly available to enhance our collective global capacity to mitigate the harms of ship-borne invasions.

ORCID

Paul Czechowski  <https://orcid.org/0000-0001-7894-4042>

Erin Grey  <https://orcid.org/0000-0001-5883-0013>

Rein Brys  <https://orcid.org/0000-0002-0688-3268>

REFERENCES

- Ahyong, S., Costello, M. J., Galil, B. S., Gollasch, S., Hutchings, P., Katsanevakis, S., Lejeune, C., Marchini, A., Occhipinti, A., Pagad, S., Poore, G. C. B., Rius, M., Robinson, T. B., Sterrer, W., Turon, X., Willan, R. C., & Zhan, A. (2022). World Register of Introduced Marine Species (WRiMS). <https://doi.org/10.14284/347>
- Andres, K. J., Sethi, S. A., Lodge, D. M., & Andrés, J. (2021). Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. *Molecular Ecology*, 30(3), 685–697.
- Apte, S., Holland, B. S., Godwin, L. S., & Gardner, J. (2000). Jumping ship: A stepping stone event mediating transfer of a non-indigenous species via a potentially unsuitable environment. *Biological Invasions*, 2(1), 75–79.
- Bailey, S. A. (2015). An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. *Aquatic Ecosystem Health & Management*, 18(3), 261–268.
- Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT – How do these taxonomies compare? *BMC Genomics*, 18(S2), 114. <https://doi.org/10.1186/s12864-017-3501-4>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bax, N., Williamson, A., Agüero, M., Gonzalez, E., & Geeves, W. (2003). Marine invasive alien species: A threat to global biodiversity. *Marine Policy*, 27(4), 313–323. [https://doi.org/10.1016/S0308-597X\(03\)00041-1](https://doi.org/10.1016/S0308-597X(03)00041-1)
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, 46(D1), D41–D47. <https://doi.org/10.1093/nar/gkx1094>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS One*, 12(9), e0183347. <https://doi.org/10.1371/journal.pone.0183347>
- Buchner, D., Macher, T. H., Beermann, A. J., Werner, M. T., & Leese, F. (2021). Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. *Environmental Science and Ecotechnology*, 8, 100122.
- Bulleri, F., & Airoidi, L. (2005). Artificial marine structures facilitate the spread of a non-indigenous green alga, *Codium fragile* ssp. *Tomentosoides*, in the North Adriatic Sea. *Journal of Applied Ecology*, 42(6), 1063–1072. <https://doi.org/10.1111/j.1365-2664.2005.01096.x>
- Bushnell, B. (2014). BBTtools software package v37.44. Available from <https://sourceforge.net/projects/bbmap/>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Capinha, C., Essl, F., Seebens, H., Moser, D., & Pereira, H. M. (2015). The dispersal of alien species redefines biogeography in the Anthropocene. *Science*, 348(6240), 1248–1251.
- Carlton, J. (1996). Marine bioinvasions: The alteration of marine ecosystems by nonindigenous species. *Oceanography*, 9(1), 36–43. <https://doi.org/10.5670/oceanog.1996.25>
- Carlton, J. T., & Geller, J. B. (1993). Ecological roulette: The global transport of nonindigenous marine organisms. *Science*, 261(5117), 78–82. <https://doi.org/10.1126/science.261.5117.78>
- Clarke, L. J., & Deagle, B. E. (2020). Eukaryote plankton assemblages in the southern Kerguelen Axis region: Ecological drivers differ between size fractions. *Deep Sea Research Part II: Topical Studies in Oceanography*, 174, 104538.
- Costello, M. J., Tsai, P., Wong, P. S., Cheung, A. K. L., Basher, Z., & Chaudhary, C. (2017). Marine biogeographic realms and species endemicity. *Nature Communications*, 8(1), 1057. <https://doi.org/10.1038/s41467-017-01121-2>
- Coutant, O., Cantera, I., Cilleros, K., Dejean, T., Valentini, A., Murienne, J., & Brosse, S. (2021). Detecting fish assemblages with environmental DNA: Does protocol matter? Testing eDNA metabarcoding method robustness. *Environmental DNA*, 3(3), 619–630.
- Cuthbert, R. N., Pattison, Z., Taylor, N. G., Verbrugge, L., Diagne, C., Ahmed, D. A., Leroy, B., Angelo, E., Briski, E., Capinha, C., Catford, J. A., Dalu, T., Essl, F., Gozlan, R. E., Haubrock, P. J., Kourantidou, M., Kramer, A. M., Renault, D., Wasserman, R. J., & Courchamp, F. (2021). Global economic costs of aquatic invasive alien species. *Science of the Total Environment*, 775, 145238.
- Czechowski, P., de Lange, M., Knapp, M., Terauds, A., & Stevens, M. I. (2023). Antarctic biodiversity predictions through substrate qualities and environmental DNA. *Frontiers in Ecology and the Environment*, 20(10), 550–557.
- Darling, J. A., Martinson, J., Lohan, K. M. P., Carney, K. J., Pilgrim, E., Banerji, A., Holzer, K. K., & Ruiz, G. M. (2020). Metabarcoding quantifies differences in accumulation of ballast water borne biodiversity among three port systems in the United States. *Science of the Total Environment*, 749, 141456.
- Davidson, I. C., Scianni, C., Minton, M. S., & Ruiz, G. M. (2018). A history of ship specialization and consequences for marine invasions, management and policy. *Journal of Applied Ecology*, 55(4), 1799–1811. <https://doi.org/10.1111/1365-2664.13114>
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872–5895.
- Deplazes-Zemp, A., Abiven, S., Schaber, P., Schaeppman, M., Schaeppman-Strub, G., Schmid, B., Shimizu, K. K., & Altermatt, F. (2018). The Nagoya protocol could backfire on the global south. *Nature Ecology & Evolution*, 2(6), 917–919.
- Drake, J. M., & Lodge, D. M. (2004). Global hot spots of biological invasions: Evaluating options for ballast-water management. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1539), 575–580. <https://doi.org/10.1098/rspb.2003.2629>
- Dully, V., Balliet, H., Frühe, L., Däumer, M., Thielen, A., Gallie, S., Berrill, I., & Stoeck, T. (2021). Robustness, sensitivity and reproducibility of eDNA metabarcoding as an environmental biomonitoring tool in coastal salmon aquaculture—an inter-laboratory study. *Ecological Indicators*, 121, 107049.
- Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31(21), 3476–3482. <https://doi.org/10.1093/bioinformatics/btv401>
- Ferrier, S., Drielsma, M., Manion, G., & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in Northeast New South Wales. II. Community-Level Modelling. *Biodiversity and Conservation*, 11(12), 2309–2338.

- Floerl, O., Inglis, G. J., Dey, K., & Smith, A. (2009). The importance of transport hubs in stepping-stone invasions. *Journal of Applied Ecology*, 46(1), 37–45.
- Formel, N., Enochs, I. C., Sinigalliano, C., Anderson, S. R., & Thompson, L. R. (2021). Subsurface automated samplers for eDNA (SAsE) for biological monitoring and research. *HardwareX*, 10, e00239.
- Forrest, B. M., Fletcher, L. M., Atalah, J., Piola, R. F., & Hopkins, G. A. (2013). Predation limits spread of *Didemnum vexillum* into natural habitats from refuges on anthropogenic structures. *PLoS One*, 8(12), 1–12. <https://doi.org/10.1371/journal.pone.0082229>
- Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A., Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A. A., Renshaw, M. A., & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, 8(1), 1–10.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One*, 9(2), e87624. <https://doi.org/10.1371/journal.pone.0087624>
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, 11(2), 37–50.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Keller, R. P., Drake, J. M., Drew, M. B., & Lodge, D. M. (2011). Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Diversity and Distributions*, 17(1), 93–102. <https://doi.org/10.1111/j.1472-4642.2010.00696.x>
- Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*, 9(1), 12133. <https://doi.org/10.1038/s41598-019-48546-x>
- Kozioł, A., Stat, M., Simpson, T., Jarman, S., DiBattista, J. D., Harvey, E. S., Marnane, M., McDonald, J., & Bunce, M. (2019). Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources*, 19(2), 366–376.
- Lambert, C., & Lambert, G. (2003). Persistence and differential distribution of nonindigenous ascidians in harbors of the Southern California bight. *Marine Ecology Progress Series*, 259(1917), 145–161. <https://doi.org/10.3354/meps259145>
- Lodge, D. M. (1993). Biological invasions: Lessons for ecology. *Trends in Ecology & Evolution*, 8(4), 133–137.
- Lodge, D. M. (2022). Policy action needed to unlock eDNA potential. *Frontiers in Ecology and the Environment*, 20(8), 448–449.
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: An effective distance metric for microbial community comparison. *The ISME Journal*, 5(2), 169–172.
- Marion, Z. H., Fordyce, J. A., & Fitzpatrick, B. M. (2017). Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover. *Ecology*, 98(4), 933–939.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17(1), 10–12.
- McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Megina, C., González-Duarte, M. M., & López-González, P. J. (2016). Benthic assemblages, biodiversity and invasiveness in marinas and commercial harbours: An investigation using a bioindicator group. *Biofouling*, 32(4), 465–475.
- Mokany, K., Ware, C., Woolley, S. N., Ferrier, S., & Fitzpatrick, M. C. (2022). A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, 31, 802–821.
- Mousavi-Derazmahalleh, M., Stott, A., Lines, R., Peverley, G., Nester, G., Simpson, T., Zawierta, M., De La Pierre, M., Bunce, M., & Christophersen, C. T. (2021). eDNAFlow, an automated, reproducible and scalable workflow for analysis of environmental DNA sequences exploiting Nextflow and singularity. *Molecular Ecology Resources*, 21(5), 1697–1704.
- Oksanen, J., Simpson, G. L., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Weedon, J. (2022). Vegan: Community ecology package. R Package Version 2.6-4. <https://CRAN.R-project.org/package=vegan>
- Pearman, J. K., von Ammon, U., Laroche, O., Zaiko, A., Wood, S. A., Zubia, M., Planes, S., & Pochon, X. (2021). Metabarcoding as a tool to enhance marine surveillance of nonindigenous species in tropical harbors: A case study in Tahiti. *Environmental DNA*, 3(1), 173–189.
- Piola, R. F., & Johnston, E. L. (2008). Pollution reduces native diversity and increases invader dominance in marine hard-substrate communities. *Diversity and Distributions*, 14(2), 329–342. <https://doi.org/10.1111/j.1472-4642.2007.00430.x>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), 7188–7196. <https://doi.org/10.1093/nar/gkm864>
- R Core Development Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Rey, A., Carney, K. J., Quinones, L. E., Pagenkopp Lohan, K. M., Ruiz, G. M., Basurko, O. C., & Rodríguez-Ezpeleta, N. (2019). Environmental DNA metabarcoding: A promising tool for ballast water monitoring. *Environmental Science & Technology*, 53(20), 11849–11859.
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, 54, 507–554.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
- Ruiz, G. M., Fofonoff, P. W., Carlton, J. T., Wonham, M. J., & Hines, A. H. (2000). Invasion of coastal marine communities in North America: Apparent patterns, processes, and biases. *Annual Review of Ecology and Systematics*, 31(1), 481–531.
- Saebi, M., Xu, J., Grey, E. K., Lodge, D. M., Corbett, J. J., & Chawla, N. (2020). Higher-order patterns of aquatic species spread through the global shipping network. *PLoS One*, 15, 1–24. <https://doi.org/10.1371/journal.pone.0220353>
- Saebi, M., Xu, J., Kaplan, L. M., Ribeiro, B., & Chawla, N. V. (2020). Efficient modeling of higher-order dependencies in networks: From algorithm to application for anomaly detection. *EPJ Data Science*, 9(1), 15. <https://doi.org/10.1140/epjds/s13688-020-00233-y>
- Sax, D. F., Stachowicz, J. J., Brown, J. H., Bruno, J. F., Dawson, M. N., Gaines, S. D., Grosberg, R. K., Hastings, A., Holt, R. D., Mayfield, M. M., O'Connor, M. I., & Rice, W. R. (2007). Ecological and evolutionary insights from species invasions. *Trends in Ecology & Evolution*, 22(9), 465–471.
- Seebens, H., Gastner, M. T., & Blasius, B. (2013). The risk of marine bioinvasion caused by global shipping. *Ecology Letters*, 16(6), 782–790. <https://doi.org/10.1111/ele.12111>
- Sherrill-Mix, S. (2019). Taxonomizr: Functions to work with NCBI accessions and taxonomy.
- Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Jaidah, M. A., Orlando, L., Willerslev, E., Møller, P. R., & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1(1), 1–5.

- Spear, M. J., Embke, H. S., Krysan, P. J., & Vander Zanden, M. J. (2021). Application of eDNA as a tool for assessing fish population abundance. *Environmental DNA*, 3(1), 83–91.
- Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S., Sigsgaard, E. E., & Hellström, M. (2017). Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: Advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), 635–645.
- Tamburri, M. N., Georgiades, E. T., Scianni, C., First, M. R., Ruiz, G. M., & Junemann, C. E. (2021). Technical considerations for development of policy and approvals for In-water cleaning of ship biofouling. *Frontiers in Marine Science*, 8, 804766.
- Tuomisto, H., & Ruokolainen, K. (2006). Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology*, 87(11), 2697–2708.
- United Nations Conference on Trade and Development (UNCTAD). (2021). In P. Stalker (Ed.), *Review of maritime transport*. United Nations Publications.
- Wan, Z., Shi, Z., Nie, A., Chen, J., & Wang, Z. (2021). Risk assessment of marine invasive species in Chinese ports introduced by the global shipping network. *Marine Pollution Bulletin*, 173, 112950.
- Wang, Z., Countryman, A. M., Corbett, J. J., & Saebi, M. (2022). Economic and environmental impacts of ballast water management on Small Island developing states and least developed countries. *Journal of Environmental Management*, 301, 113779.
- Wang, Z., Nong, D., Countryman, A. M., Corbett, J. J., & Warziniack, T. (2020). Potential impacts of ballast water regulations on international trade, shipping patterns, and the global economy: An integrated transportation and economic modeling assessment. *Journal of Environmental Management*, 275, 110892.
- Wang, Z., Saebi, M., Corbett, J. J., Grey, E. K., & Curasi, S. R. (2021). Integrated biological risk and cost model analysis supports a geopolitical shift in ballast water management. *Environmental Science & Technology*, 55(19), 12791–12800.
- Wang, Z., Silberman, J. A., & Corbett, J. J. (2021). Container vessels diversion pattern to trans-Arctic shipping routes and GHG emission abatement potential. *Maritime Policy & Management*, 48(4), 543–562.
- Wickham, H. (2017). Tidyverse: Easily install and load the “Tidyverse” (R package version 1.2.1).
- Wonham, M. J., Byers, J. E., Grosholz, E. D., & Leung, B. (2013). Modeling the relationship between propagule pressure and invasion risk to inform policy and management. *Ecological Applications*, 23(7), 1691–1706.
- WoRMS Editorial Board. (2019). World register of marine species. Available from <https://www.marinespecies.org> at VLIZ.
- Xu, J., Wickramaratne, T. L., & Chawla, N. V. (2016). Representing higher-order dependencies in networks. *Science Advances*, 2(5), e1600028.
- Xu, J., Wickramaratne, T. L., Chawla, N. V., Grey, E. K., Steinhäuser, K., Keller, R. P., Drake, J. M., & Lodge, D. M. (2014). Improving management of aquatic invasions by integrating shipping network, ecological, and environmental data: Data mining for social good. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1699–1708). Association for Computing Machinery.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Andrés, J., Czechowski, P., Grey, E., Saebi, M., Andres, K., Brown, C., Chawla, N., Corbett, J. J., Brys, R., Cassey, P., Correa, N., Deveney, M. R., Egan, S. P., Fisher, J. P., vanden Hooff, R., Knapp, C. R., Leong, S. C. Y., Neilson, B. J., Paolucci, E. M. ... Lodge, D. M. (2023). Environment and shipping drive environmental DNA beta-diversity among commercial ports. *Molecular Ecology*, 32, 6696–6709. <https://doi.org/10.1111/mec.16888>