

## LETTER

## The risk of marine bioinvasion caused by global shipping

H. Seebens,<sup>1</sup> M. T. Gastner<sup>2,3</sup> and B. Blasius<sup>1\*</sup>

### Abstract

The rate of biological invasions has strongly increased during the last decades, mostly due to the accelerated spread of species by increasing global trade and transport. Here, we combine the network of global cargo ship movements with port environmental conditions and biogeography to quantify the probability of new primary invasions through the release of ballast water. We find that invasion risks vary widely between coastal ecosystems and classify marine ecoregions according to their total invasion risk and the diversity of their invasion sources. Thereby, we identify high-risk invasion routes, hot spots of bioinvasion and major source regions from which bioinvasion is likely to occur. Our predictions agree with observations in the field and reveal that the invasion probability is highest for intermediate geographic distances between donor and recipient ports. Our findings suggest that network-based invasion models may serve as a basis for the development of effective, targeted bioinvasion management strategies.

### Keywords

Ballast water, bioinvasion hot spots, ecoregion, intermediate distance, invasion probability, shipping network.

*Ecology Letters* (2013) 16: 782–790

### INTRODUCTION

During the last few centuries thousands of species have dispersed beyond their native ranges and have transformed marine ecosystems around the world (Ruiz *et al.*, 1997; Mack *et al.*, 2000; Kolar & Lodge, 2001; Lockwood *et al.*, 2007). In the past decades the worldwide dispersal of species has increased by orders of magnitude, and this has contributed to some regions now being invaded by several new species per year (Cohen & Carlton, 1998; Reise *et al.*, 1998; Coles *et al.*, 1999; Hewitt *et al.*, 2004). The globalisation of maritime trade plays a key role in this accelerated spread of species because many of them are dispersed by cargo ships (Kolar & Lodge, 2002; Hulme, 2009). For stability reasons, cargo ships load large amounts of ballast water which may contain aquatic species entrained as stowaways over long distances. The release of ballast water is assumed to represent the world's largest invasion vector (Ruiz *et al.*, 1997). Even though invasive species are now recognised as one of the greatest threats to marine biodiversity (Sala *et al.*, 2000; Molnar *et al.*, 2008; McGeoch *et al.*, 2010), pathways for species dispersal remain poorly understood on a global scale (Mack *et al.*, 2000; Hulme, 2009) and only a few studies have explicitly incorporated data of shipping networks to predict invasion dynamics (Drake & Lodge, 2004; Tatem *et al.*, 2006; Brawley *et al.*, 2009; Keller *et al.*, 2010), not least due to the paucity of comprehensive information about shipping traffic.

Recently, there has been progress in the global tracking of ship journeys. The Automatic Identification System (AIS), mandatory on cargo ships engaged in international voyages, automatically reports approaching ships to coastal stations. Based on these data, Kaluza *et al.* (2010) constructed the network of global cargo ship movements providing information about the volume, frequency, origin and destination of shipping connections between ports worldwide. Using this analysis, prominent patterns in the global cargo shipping

traffic could be identified, such as the most central ports in the network, shipping sub-networks and communities of highly interconnected ports – indicating the ports and routes which can be central to the spread of non-native species (Kaluza *et al.*, 2010; Kölzsch & Blasius, 2011).

Invasion models purely based on shipping routes provide a first important step towards understanding patterns of ship-associated bioinvasion (Drake & Lodge, 2004; Tatem *et al.*, 2006; Kaluza *et al.*, 2010; Kölzsch & Blasius, 2011). However, the likelihood of new introductions depends not only on shipping intensity but also on other ecological and environmental factors that allow non-native species to persist (Keller *et al.*, 2010). To achieve a more comprehensive estimation of global marine bioinvasion potentials, we combine in this study AIS-based data about global cargo shipping with ballast water release protocols, port environmental conditions (i.e. temperature and salinity) and biogeography to develop a model of marine bioinvasion by worldwide shipping. Our modelling approach differs from other invasion models (Kolar & Lodge, 2002; Drake & Lodge, 2004; Tatem *et al.*, 2006; Keller *et al.*, 2010) in several important ways: First, the model simultaneously integrates propagule pressure (expressed by global shipping and ballast water releases), invasibility (expressed by environmental matching between source and destination sites) and community mismatch (expressed by biogeographic dissimilarity). Second, the model explicitly considers the full trajectory of marine vessels by considering all ports previously called by a ship as potential sources for non-native species. This is crucial for risk estimation as species in ballast water tanks easily can be introduced to multiple ports of call. Third, the model does not describe the spread of a particular known species, but instead estimates the likelihood for the invasion of a new, previously unknown species from a large pool of potential invaders.

Using this framework, our model quantifies the probability that a non-native species is successfully introduced and establishes a popu-

<sup>1</sup>Institute for Chemistry and Biology of the Marine Environment, Carl-von-Ossietzky University, Oldenburg, Germany

<sup>2</sup>Department of Mathematics, Imperial College London, London, UK

<sup>3</sup>Department of Engineering Mathematics, University of Bristol, Bristol, UK

\*Correspondence: E-mail:blasius@icbm.de

lation in a new environment for every port and shipping connection worldwide. We denote this probability as the 'risk of invasion', although we acknowledge that we do not consider the further spread of a non-native species after its first introduction nor its impact to the recipient ecosystem. The invasion risks obtained in this way show a large variation among port-to-port connections, mainly depending on shipping intensity, environmental match and geographic distances. To obtain a better insight into invasion flows at a global scale, we aggregate the calculated invasion risks at the level of large-scale marine ecoregions. This allows us to compare our model predictions to empirical data and to classify ecoregions according to their aggregated invasion risk and the evenness of relative invasion risks from different source regions associated with worldwide shipping activity. Based on our model results and the empirical data, we formulate an intermediate distance hypothesis stating that the likelihood of new introductions is maximal for geographic distances of about 8000–10 000 km between donor and recipient sites. Finally, we use our model framework to investigate scenarios for invasion mitigation by ballast water treatment.

## MATERIALS AND METHODS

### Data

#### *Ship movement data*

We use ship movement data recorded during 2007 and 2008 by the AIS, a transponder system installed on board of ships and in ports, at canals and other land sites. The system automatically reports the location of a ship and other information such as ship size, ship type and date when a ship enters the transmission range ( $\approx 30$  km) of another AIS transponder. The data are made available by Lloyd's Register Fairplay (<http://www.lrfairplay.com>) for commercial purposes and have been used by Kaluza *et al.* (2010) to reconstruct the network of global cargo ship movements. The data include both international and domestic shipping. However, AIS devices had not been installed in all ships and ports in 2008. The database represents 67% of the world fleet of ships measured in dead weight tonnages. The remaining 33% of the world vessel fleet mainly consists of small ships (i.e. 52% of these ships are smaller than 1000 gross tonnages) and thus are less important for shipping-associated primary invasions (see calculation of ballast water releases below). Following Kaluza *et al.* (2010), we used these data to construct a trajectory (i.e. a list of ports or canals visited by the ship and sorted by time, including arrival and departure dates) for each ship in the database. Contrary to Kaluza *et al.* (2010), we did not restrict our analysis to ships larger than 10 000 gross tonnages, but instead included ships of all sizes reported by AIS. In addition, we considered all ship movements in the time span 2007–2008 because we believe it to be representative of global traffic patterns, and therefore should provide a reasonable approximation of species transport across years. The resulting network includes 32 511 ships which made 2 892 523 voyages calling at 1469 distinct ports.

#### *Ballast water releases*

Ballast water releases were calculated using data provided by the National Ballast Information Clearinghouse (NBIC Online Database, 2012). NBIC collects ballast water discharge volumes for every ship calling a port in the USA since 2004, which constitutes the most comprehensive sampling available yet. In this study we com-

puted 717 250 releases during 2004-01-01–2012-07-12 with an average of 65 205 releases for each ship type. The relationship between ballast water releases and ship sizes for those ships common to our database (42%) shows a large variability and many ships did not release any ballast water during their stay in a port. The fraction  $\xi$  of port calls without ballast water discharges depended on ship type ranging from 42% (Bulk-Dry) to 88% (Refrigerated Cargo). Excluding the zero releases, we assumed a power-law relationship between released ballast water and ship sizes for all ship types to fit a mean ballast water release  $W$  specific to the type and size of a vessel (Figure S1). Consider now a ship sailing from port  $i$  to port  $j$  on route  $r$  with  $\delta_r$  intermediate stopovers. At each intermediate stopover, we assume that an average fraction  $\xi W_r/V_r$  of ballast water is exchanged with  $\xi$  being the fraction of zero releases and  $V_r$  the ballast tank volume of a ship. The average total amount of ballast water from port  $i$  that is released at port  $j$  is then described by  $B_r = \xi W_r (1 - \xi W_r/V_r)^{\delta_r}$ . The ballast tank volume of a ship is estimated from the ship's carrying capacity measured in dead weight tonnes. A list of mean ballast tank volumes for various ship size classes and vessel types, provided by ABS (2011), indicates that on average  $V_r$  is one quarter of the ships's carrying capacity.

#### *Data of port conditions*

For each port, surface water temperatures were taken from the World Ocean Atlas (WOA, <http://www.nodc.noaa.gov>) which contains 50-year averages at a  $1^\circ$  resolution. Surface water salinities were compiled from Lloyd's Register Fairplay (<http://www.port-guide.com>) providing port-specific water densities for most of the ports used in this study (69%). Water densities were transformed to salinities using water temperatures from the WOA according to McCutcheon *et al.* (1993). For the remaining ports, salinities were taken from the WOA. Port salinities are difficult to sample representatively as the density of the water can fluctuate distinctly in space and time. We therefore tested the robustness of our result by comparing our model predictions with the model outcome using salinity data of Keller *et al.* (2010) (Figure S2). Environmental data were averaged annually, except for simulating seasonal invasion risks where monthly temperature averages were taken from the World Ocean Atlas and interpolated to a daily resolution.

#### *Data of reported invasions*

For model validation we used field data of reported invasion events. The data describe either the number of reported introductions of marine species from various source regions into one of the following four sites: North Sea, Europe (Reise *et al.*, 1998), Pearl Harbor, Hawaii (Coles *et al.*, 1999), Port Phillip Bay, Australia (Hewitt *et al.*, 2004), and San Francisco Bay, USA (Cohen & Carlton, 1998); or they describe the number of reported invasions in various ecoregions (Molnar *et al.*, 2008). Only those studies were included providing sufficient information about the source regions of invaded marine species. If distinguished in the reference, only species likely to have been introduced by ships (North Sea, various ecosystems) or through ballast water releases (Port Phillip Bay) were considered.

## Model for marine bioinvasion

The model estimates the likelihood of new primary invasions for every port call of a vessel from a large pool of species at previous stopover sites. Consider a ship that sails from donor port  $i$  to a

recipient port  $j$  along a route  $r$  of successive stopover sites (see Figure S3 in Supporting Information). Our model accounts for the fact that bioinvasion is a multistage process: A species must pass three distinct invasion transitions, which act as successive filters, before it can successfully invade the recipient port along route  $r$ . Each transition is associated with an independent probability of failure, as different requirements must be met to pass from one stage to the next:

(i) The probability to be alien

$$P_{ij}(Alien) = \left(1 + \frac{\gamma}{d_{ij}}\right)^{-\beta} \quad (1)$$

describes the likelihood that a native species in donor port  $i$  is non-native in recipient port  $j$ . This probability is estimated by biogeographical dissimilarity (Tuomisto *et al.*, 2003; Thieltges *et al.*, 2009) which we assume to increase sigmoidally with the geographical distance  $d_{ij}$  between sites (Tuomisto *et al.*, 2003). Here,  $\beta > 0$  is a shape parameter and  $\gamma$  is a characteristic geographic scale on which species composition does not change (Spalding *et al.*, 2007). The probability  $P(Alien)$  is a crucial component of our model and accounts for the fact that the likelihood of new introductions increases with the dissimilarity between the donor and recipient communities. In particular, this term ensures that the invasion risks between two closely located ports are negligible which is a natural assumption as the vicinity of a port should contain almost only species that are already present at the port. Note that  $P(Alien)$  identifies the chance that a species is non-native, but does not differentiate whether or not this species has potential negative impact in the recipient ecosystem.

(ii) The probability of introduction

$$P_r(Intro) = \rho_r (1 - e^{-\lambda B_r}) e^{-\mu \Delta t_r} \quad (2)$$

depends on the survival of species entrained in ballast tanks, which decays exponentially with mortality rate  $\mu$  as a function of travel time  $\Delta t_r$  (Verling *et al.*, 2005; Wonham *et al.*, 2001, 2005), it increases with the total amount of ballast water  $B_r$  that originates from port  $i$  ( $\lambda$  is a characteristic constant), and it is reduced by the factor  $\rho_r$  due to possible ballast water treatment. This factor can be set specific to different ship movements, ports or ship types and thus allows modelling different ballast water treatment scenarios. If, for example, ballast water is treated equally at every port call of the ship and  $\rho$  describes the fraction of species left in the ballast water after a single treatment, the risk reduction over the full trajectory is given by  $\rho_r = \rho^{\delta_r + 1}$ .

(iii) The probability of establishment

$$P_{ij}(Estab) = \alpha e^{-\frac{1}{2} \left[ \left( \frac{\Delta T_{ij}}{\sigma_T} \right)^2 + \left( \frac{\Delta S_{ij}}{\sigma_S} \right)^2 \right]} \quad (3)$$

increases with the environmental similarity between donor and recipient ports, which is modelled as a Gaussian function of the differences in water temperature  $\Delta T_{ij}$  and salinity  $\Delta S_{ij}$ , standardised by  $\sigma_T$  and  $\sigma_S$ , as the two main environmental factors during establishment, and  $\alpha$  is the basic probability of establishment.

Assuming that these three probabilities are independent (Figure S4), their product determines the likelihood of a new primary invasion by ballast water release originating from one particular ship movement  $r$ . Here, 'primary invasion' refers to the establishment of new species introduced from their native geographic range.

Although, our model takes into account indirect connections between donor and recipient ports via intermediate stopovers along a shipping route, it neither considers secondary invasions due to stepping-stone processes (Floerl *et al.*, 2009), nor the further spread of species after establishment, nor their impact to recipient ecosystems.

The total invasion risk  $P_{ij}(Inv)$  from  $i$  to  $j$  (i.e. the probability of observing at least one primary invasion during the study period 2007–2008) is the complement of species failing to invade on all shipping routes  $r_{ij}$  in the network that are directly or indirectly connecting port  $i$  to  $j$  (Figure S3),

$$P_{ij}(Inv) = 1 - \prod_{r_{ij}} [1 - P_{ij}(Alien) P_r(Intro) P_{ij}(Estab)]. \quad (4)$$

The risk of a new invasion at port  $j$  is obtained by aggregating invasion risks over all shipping connections passing through  $j$ ,

$$P_j(Inv) = 1 - \prod_i [1 - P_{ij}(Inv)]. \quad (5)$$

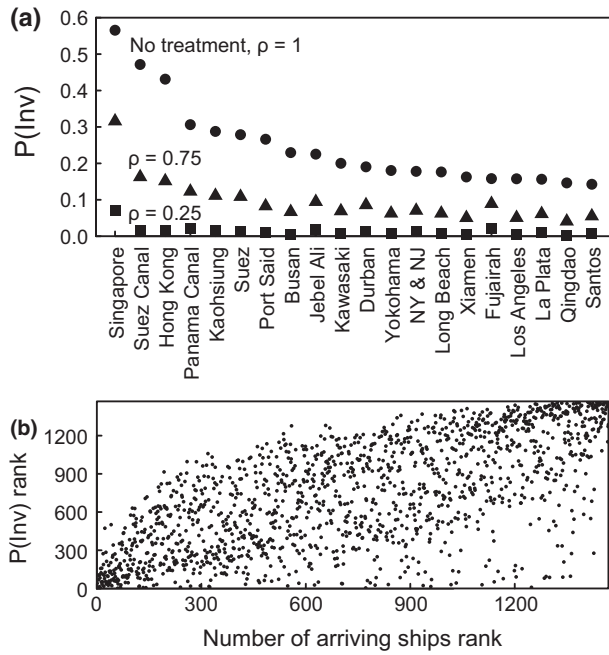
Aggregated invasion risks can be similarly obtained for different ships, ship types (Figure S5), ports (Figure S6), geographic regions or ecosystems. The expected number of new establishments in port  $j$  in a time period  $\tau$  can be calculated as  $N_j = -\log(1 - P_j(Inv)) \cdot \tau / \tau_0$ , with the duration of the study period  $\tau_0 = 2$  years. The model set-up also allows us to resolve the temporal dynamics of invasion risks such as risk development during the voyage of a ship (Figure S7) or the seasonal dynamics of ecoregions (Figure S8). Parameter values were taken as follows:  $\beta = 8$ ,  $\gamma = 1000$  km,  $\mu = 0.02$  per day,  $\lambda = 0.002 m^{-3}$ ,  $\alpha = 0.00015$ ,  $\sigma_T = 2^\circ\text{C}$  and  $\sigma_S = 10$  ppt. A sensitivity analysis is provided in the Supplementary Information (Figure S9). The model, its main assumptions, as well as various model extensions are described in detail in the Appendix 1 of the Supporting Information.

## RESULTS AND DISCUSSION

### Hotspots and highways of marine bioinvasion

The estimated invasion risks show great geographic heterogeneity and vary strongly among the different ports and shipping connections (Fig. 1 and Figure S6). The majority of ports are unlikely to receive new primary introductions via ballast water, and high invasion risks are concentrated on a small number of ports. These invasion hot spots are mainly located in (South) East Asia (eight ports among the top 20), the Middle East (five ports among the top 20), but also in the USA (three ports among the top 20), with Singapore, the Suez Canal (Egypt), Hong Kong and the Panama Canal being the sites of highest invasion probability (Fig. 1a and Table S1). All identified invasion hot spots are characterised by high shipping intensity and, in general, the invasion risk increases with the number of ship arrivals (Spearman's  $\rho = 0.68$ , Fig. 1b). On the other hand, a high shipping intensity does not necessarily lead to a high invasion risk. Most notably, North Sea ports do not rank among the top endangered ports despite their enormous maritime traffic. This result is contrary to previous studies which identified invasion hot spots solely from shipping intensity (Drake & Lodge, 2004). Traffic volume alone constitutes a poor predictor of bioinvasion rates (Verling *et al.*, 2005; Minton *et al.*, 2005) owing to the neglect of other critical factors (Figure S10). For instance, our model results change distinctly if port environmental conditions are





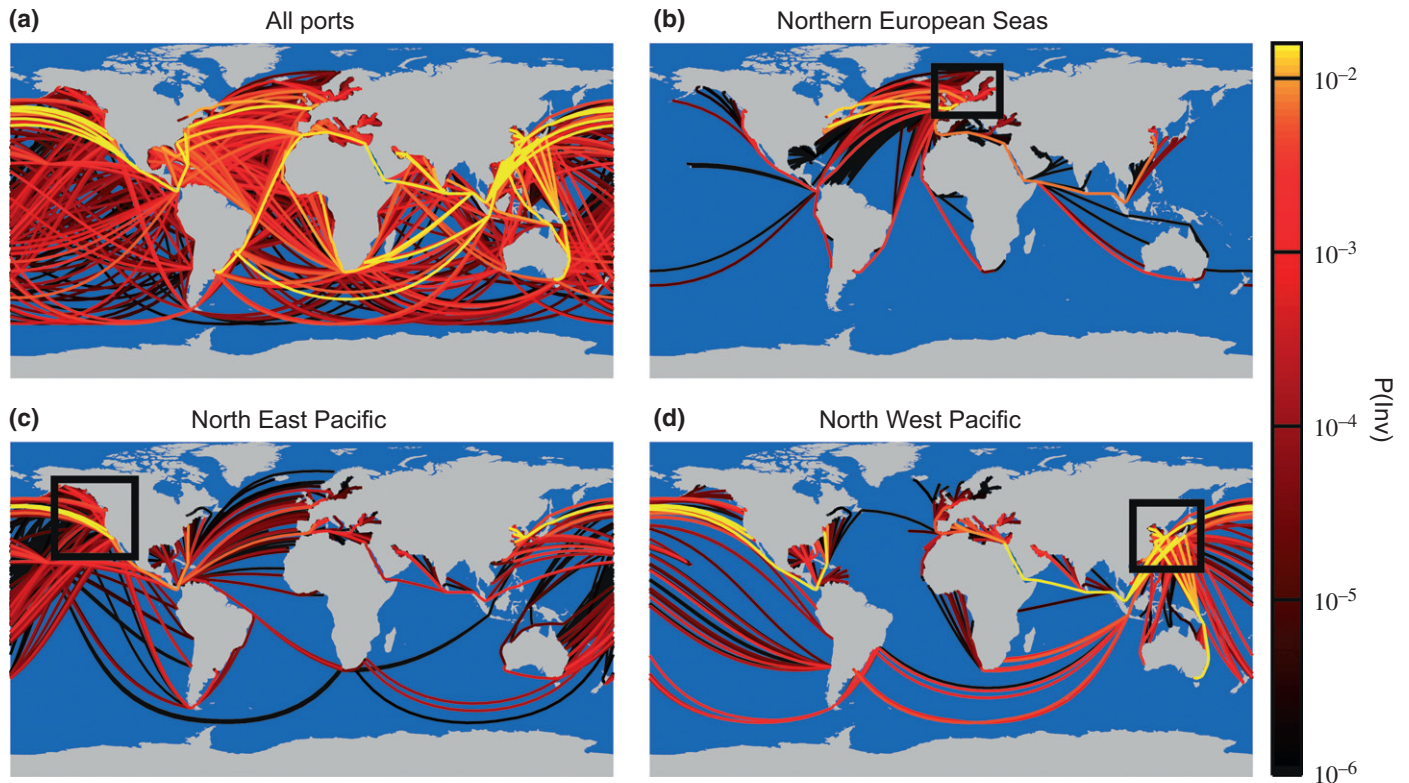
**Figure 1** Predicted invasion risks in rank order. (a) Estimated invasion probability  $P(Inv)$  of the top 20 highest risk ports for three different levels of ballast water treatment  $\rho$  (see also Table S1). (b) Rank plot showing the relationship between the number of arriving ships at a port and the port invasion risk (without ballast water treatment,  $\rho = 1$ ); each dot characterises a single port.

neglected by setting  $P_j(Estab)$  constant for all links (Spearman correlation of link invasion risks between the two models,  $\rho = 0.71$ ). Ignoring environmental heterogeneity resulted in a distinct increase in invasion risks of Central European ports such as Antwerp and Hamburg which then rank among the top 20 high-risk ports.

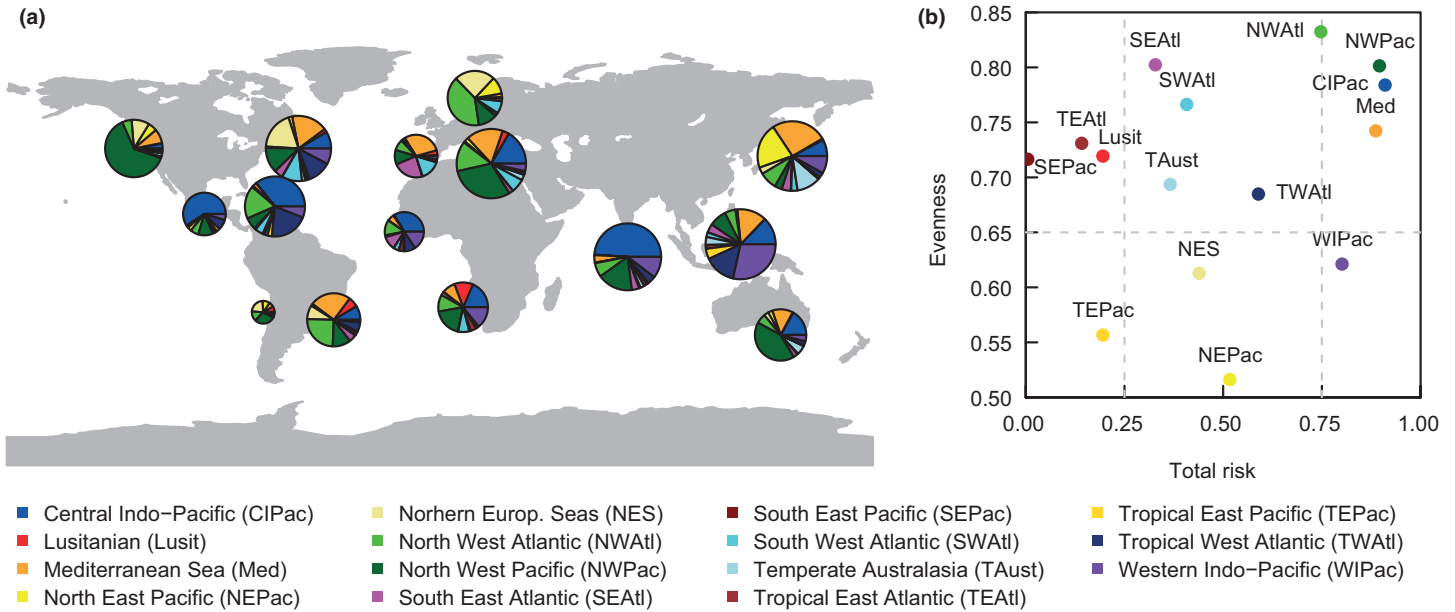
The major invasion pathways (Table S2) align with the frequently visited shipping routes Asia – Europe and Asia – North America (Kaluza *et al.*, 2010) (Fig. 2a). However, high invasion risks can also be observed on shipping connections at low latitudes as many frequently visited ports are located in tropical to subtropical areas with very similar climatic conditions.

**Invasion risk of ecoregions**

Even aggregated over larger scales, coastal ecosystems vary strongly in the distribution of source regions from which bioinvasion is likely to occur. Whereas for some regions (e.g. Northern European Seas, North-East Pacific) the major invasion pathways are concentrated on relatively few ‘highways’ of bioinvasion, there is a striking diversity of high-risk routes for others (e.g. North-West Pacific) (Fig. 2b–d). These differences result from a combination of traffic volume, distance to the most connected regions and environmental similarity. For instance, with respect to maritime traffic, the Northern European Seas are most strongly connected to tropical and subtropical ecosystems. However, among the adjacent water bodies only the North-West Atlantic provides sufficiently similar climatic conditions and thus dominates as the major source region of invasions (Fig. 2b).



**Figure 2** Major introduction pathways. (a) Invasion probability along every shipping connection between two ports in the network indicated in colour coding. (b–d) Same as in (a) but restricted to all connections with destination in one of the three target regions (black rectangles): (b) Northern European Seas, (c) North-East Pacific and (d) North-West Pacific. Shipping routes are calculated as the shortest geodesic paths on water (Kaluza *et al.*, 2010).



**Figure 3** Predicted invasion risks aggregated over  $n = 15$  coastal ecoregions (Table S3). (a) For each ecoregion  $j$ , the area of the pie chart indicates the aggregated risk of new invasions into that region  $P_j(Inv) = 1 - \prod_i (1 - P_{ij}(Inv))$ , with  $P_{ij}(Inv)$  being the mutual invasion risk between ecoregions. The size of each pie sector indicates the relative contribution  $p_{ij} = P_{ij}(Inv) / \sum_i P_{ij}(Inv)$  of source region  $i$  (indicated by colour coding) to the invasion risk of ecoregion  $j$ . (b) Characterisation of ecoregions by the total invasion risk  $P_j(Inv)$  and the evenness of risk composition using Pielou's index (Pielou, 1966),  $E_j = -\sum_i p_{ij} \log p_{ij} / \log(n)$ .

To identify global invasion flows, we estimated the mutual invasion risks between 15 coastal ecosystems, distinguished according to common concepts of marine biogeography (Spalding *et al.*, 2007) (Fig. 3a, Table S3). For each ecoregion we calculated the vulnerability, that is the total probability  $P_j(Inv)$  of new invasions into that region, and risk composition, that is the evenness  $E_j \in [0, 1]$  of relative invasion probabilities into that region from various source regions (Pielou, 1966). Thereby, the ecoregions can be classified into several categories (Fig. 3b) (see also the network visualisation of the mutual invasion risks, Figure S11). The most endangered marine ecosystems (Central Indo-Pacific, North-West Pacific, Mediterranean Sea, and North-West Atlantic) combine high invasion risk with high evenness, reflecting the central position in the worldwide flow of trade, together with the short environmental distance to most adjacent regions. Another class of ecoregions is characterised by a high invasion risk despite small evenness (Western Indo-Pacific and North-East Pacific). These regions are highly vulnerable, but receive species from only few source regions. In contrast, ecoregions with small values of total risk and evenness are only weakly connected to a small number of source regions and thus are relatively immune to invasions (Tropical East Pacific). Finally, some ecoregions combine a low invasion risk with high evenness (South-East Pacific, Tropical East Atlantic and Lusitanian). Ecosystems in this class have a central position in the shipping network, but due to a small shipping intensity the total invasion risk remains small.

### Seasonality of invasion risk

Many ports experience a striking seasonality in their environmental conditions and shipping intensity. Considering seasonally resolved temperature data in the model resulted in a distinct seasonal pattern of invasion risks in most ecosystems (Figure S8). Even tropical ecosystems, characterised by a low seasonal variation in environmental

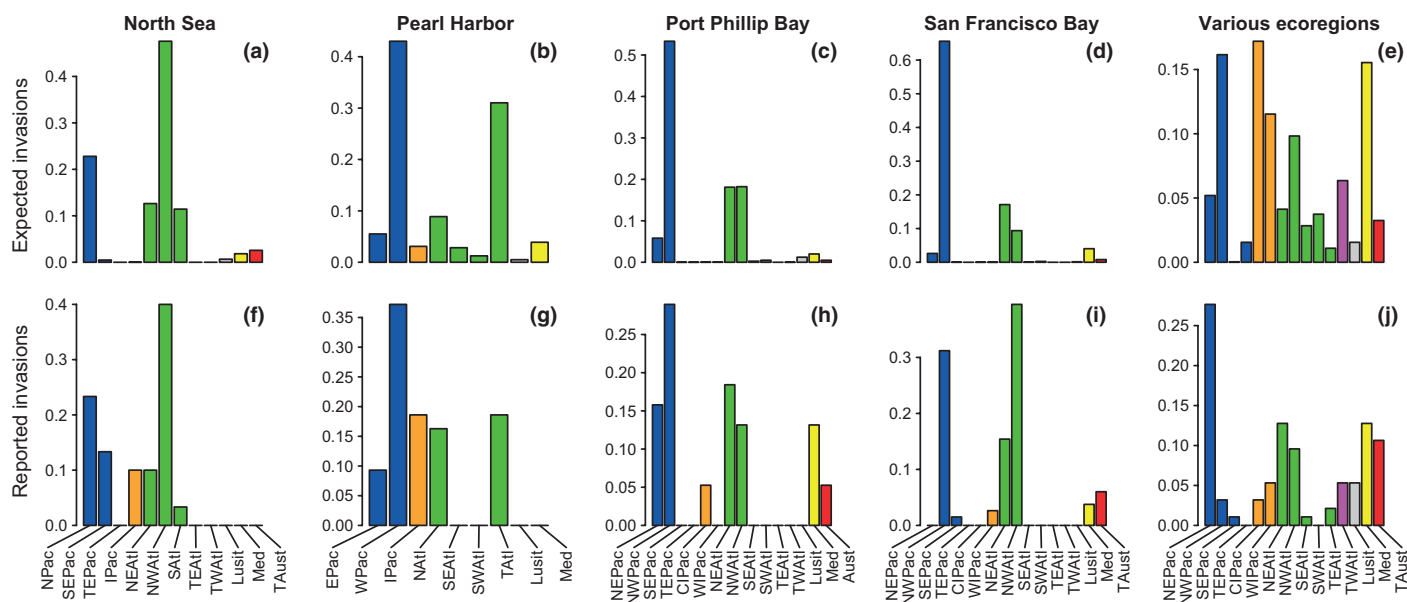
conditions, experience distinct intraannual changes in the risk of being invaded. This can be explained by seasonal variation in the match of environmental variables at tropical regions with those at the potential source areas in temperate regions. Notably, the seasonality of the invasion risk may vary strongly within the same climatic zone, for example the highest risk in North America was achieved at the west coast in winter but at the east coast during summer (Figure S8b).

### Ship types

The global shipping network consists of multiple layers each representing the movements of a different ship type with specific movement characteristics. For instance, oil tankers operate between a few ports, whereas bulk carriers move more randomly (Kaluza *et al.*, 2010). Consequently, invasion dynamics may be associated with a specific ship type which indeed can be seen in the variation in high-risk links among ship types (Figure S5). Container ships show a diverse pattern of high-risk routes connecting almost all continents, whereas oil tankers, bulk carriers and ro-ro-cargo ships exhibit only a few routes with a high invasion probability. Ship types may also vary in their ballast water performances (Verling *et al.*, 2005). Some ship types like container ships usually have large amounts of ballast water loaded, whereas others like oil tankers or bulk carriers need additional ballast mainly in one direction when they are empty. Consequently, propagule pressures can be specific to the type of the ship and the direction of travel (Minton *et al.*, 2005). However, ship-based data including the directionality of ballast water transport are not available on a global scale and could not be considered in this study.

### Comparison with field data

For model validation, we compared the expected number of invasions with recorded data on the frequency and source regions of



**Figure 4** Comparison of invasion numbers predicted by the model (a–e) with reported invasion events in four highly invaded sites (f–i) and various ecoregions (j). For each of the four selected sites, the relative contribution of different source regions for the expected (a–d) and the reported (f–i) number of invasions into that site is shown. For the various ecoregions, the number of predicted (e) and reported (j) invasions into different ecoregions is shown, independent of the source regions. All distributions were normalised to one. Deviations between model predictions and field data according to root mean squared error [RMSE]: 0.06 (North Sea), 0.08 (Pearl Harbor), 0.08 (Port Phillip Bay), 0.12 (San Francisco Bay) and 0.08 (various ecoregions). Abbreviations of ecoregions (colours indicate different oceans): Pac: Pacific, IPac: Indo-Pacific, Atl: Atlantic, Med: Mediterranean Sea, Lusit: Lusitanian, Aust: Australasia, N: North, E: East, S: South, W: West, T: Tropical and C: Central.

marine invasions in four of the most invaded marine regions of the world (Cohen & Carlton, 1998; Reise *et al.*, 1998; Coles *et al.*, 1999; Hewitt *et al.*, 2004) and with the recorded number of invasive species in various ecoregions (Molnar *et al.*, 2008) (Fig. 4). As the field studies differ in scale and sampling effort, we normalised the distributions of simulated and reported species numbers prior to the analysis, discarding the absolute value of invasion risks. We found only small deviations in the root mean squared error (RMSE) of reported data and model results (see also Figure S12): [RMSE = 0.06 (North Sea), RMSE = 0.08 (Pearl Harbor), RMSE = 0.08 (Port Phillip Bay), RMSE = 0.12 (San Francisco Bay) and RMSE = 0.08 (various ecoregions)]. This agreement between model predictions and field data indicates that our model is able to capture the relative contribution of different source regions to the observed non-native species in a particular site (e.g. the model correctly predicts the North-West Atlantic and the North Pacific as the major source regions for invasions into the North Sea) (Fig. 4a–d, f–i). On the other hand, the model predictions provide a balanced estimate of invasion risks across global sites (i.e. we do not overpredict certain source regions in relation to others) (Fig. 4e, j).

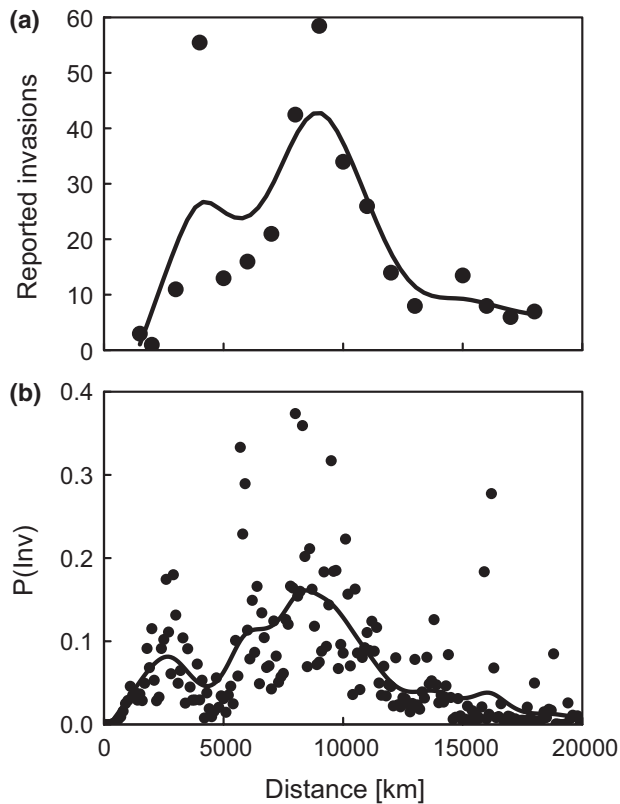
In a sensitivity analysis we found that, in general, the model is robust to the choice of parameter values; most parameter variations have only minor influence on the RMSE (Figure S9). The strongest parameter sensitivity is found in the shape of the distance dependency of the biogeographic dissimilarity. To retain the predictive power of the model, both parameters  $\beta$  and  $\gamma$  should be above critical levels ( $\beta > 6$ ,  $\gamma > 600$  km). Otherwise, the sigmoidal shape of the community dissimilarity is essentially eliminated, which is a crucial element of our model. Similarly, the mortality rate of organisms inoculated in a ballast tank should be sufficiently small

( $\mu \leq 0.02d^{-1}$ ) so that invasion risks on long connections are not suppressed too much.

Despite the observed agreement between model predictions and recorded invasion events, there remains a degree of uncertainty in the overall risk estimate. This is mainly due to the simplifications inherently associated with model building, but also to the limited availability of comprehensive data of invasion flows to calibrate the model. Data sampling, for example, is strongly biased towards some well-investigated sites in industrial countries, whereas almost no data exist for Asia which is among the regions most strongly impacted by invasion risks. Furthermore, for many established non-natives, the vector of introduction is uncertain. Thus many studies of marine invasions do not distinguish between introduction due to aquaculture, hull fouling or ballast water releases. This might explain, for example, some of the discrepancies between model predictions and the number of reported invasions in San Francisco Bay (Fig. 4).

### Intermediate distance hypothesis

Interestingly, the field data show that most introduced species originate from sites of intermediate geographic distances to destination ports (8000–10 000 km, Fig. 5a). This observation is confirmed with remarkable agreement by our model predictions (Fig. 5b) and can be explained as follows: At short distances, the probability for the introduction of a non-native species is small, resulting in a low invasion risk. At large distances, the shipping intensity is reduced and the chance to survive the transport is small, also resulting in a low invasion risk. Transportation along links of intermediate distance ensures both a high chance for the introduction of non-native species and their survival during transportation (Figure S13). Our numerical investigations demonstrate that this pattern of maximal

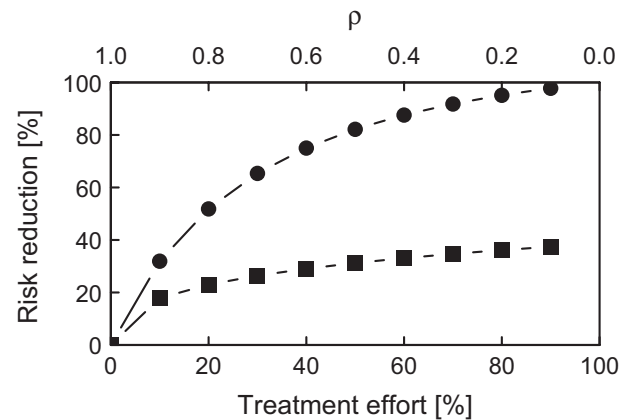


**Figure 5** Intermediate distance hypothesis: invasion risk is largest for intermediate geographic distances. Sum of reported established species at different sites (a) and predicted invasion probabilities  $P(\text{Inv})$  (b) as a function of the geographic distance  $d_{ij}$  between source and destination ports [dots, histogram in bins of 1000 km (a) and 100 km (b)] and spline fit (solid line).

invasion risk at intermediate distances is robust to a wide range of parameter settings, providing a first indication that it may be common at least in marine bioinvasion.

### Mitigation scenarios

Efforts to minimise the risk of bioinvasion through shipping are currently hotly debated. This discussion was boosted by a convention of the International Maritime Organization (IMO 2004) which postulates the treatment of ballast water on board of all ships. Adhering to the proposed standards, we tested the effect of reducing the invasion probability by a fraction  $\rho$  every time a ship enters a port (Drake & Lodge, 2004; Minton *et al.*, 2005). Although the uniform risk reduction for all ships is a very rough approximation of the currently debated ballast water treatments, our simulations reveal that already moderate efforts of ballast water treatment can yield substantial results (Fig. 1a). For example, overall invasion probabilities can be reduced by 56 or 82% if ballast water is treated according to a treatment effort of 25 or 50% at every port respectively (Fig. 6). This overproportional effect is explained by the fact that successive risk reductions at single ports multiply during the voyage of a ship. Consequently, considerable global risk reduction can be achieved by reducing risks at only a few selected high-risk ports (Fig. 6). A ballast water treatment effort of 25% at the 10 ports of highest invasion risk yields an average reduction in invasion



**Figure 6** Effects of ballast water treatment on the predicted invasion risk. Risk reduction is shown as a function of treatment effort [ $100(1-\rho)$ ], if ballast water is treated at every port call of a ship (dots) or only at the 10 sites of highest invasion risk (squares). The figure illustrates that risk reduction is not a linear function of the treatment effort and that even small treatments have a disproportionately high effect on the overall invasion risk. Treatment is modelled as a reduction in  $P(\text{Inv})$  by a factor  $\rho$  each time a ship entered a port (i.e. a value of  $\rho = 1$  indicates no treatment). Risk reduction denotes the risk difference in simulations without and with ballast water treatment relative to the untreated risk in percent.

risks on the whole network by 24.8%. This contrasts findings of other studies (Drake & Lodge, 2004) that did not consider the possibility of invasions from ports prior to the last port call.

### Model uncertainty

Any attempt to model invasion dynamics is impeded by two major difficulties. The first challenge is the limited availability of comprehensive data for model calibration and validation. With the currently available AIS records, information about global shipping traffic has a sound database. To test the dependence on the shipping data, we investigated how the model results change if the shipping database is artificially reduced from a 2-year time span to a single year. We found only minor changes: simulated invasion risks from data restricted to the year 2007 or 2008 correlated with the general invasion risks with Spearman's  $\rho = 0.98$  or  $\rho = 0.99$  respectively. To test for the influence of data about environmental variables, we performed model simulations based on two different data sets for port salinities and found a strong correlation in the simulated port invasion risks (Spearman's  $\rho = 0.99$ ) (Figure S2). From the perspective of sufficient data, our study suffers mostly from the lack of knowledge about ballast water management. Even a detailed database of ballast water release for a large collection of ports, as for example provided by NBIC, does not specify some crucial information, such as the origin of the ballast water, or the amount of ballast water transported to other countries or exchanged outside ports. In the model we resolved this by assuming a constant ballast water exchange at every port call, specific to ship type and ship size, which may give rise to substantial errors in the estimated exchanged ballast water volumes at the level of a single ship. However, the effect of these variations is the more suppressed the more ships travel on the same link because the invasion risk between a pair of ports is calculated from the average over many ship movements (on average



432 ship movements per link were recorded along the top 100 000 high-risk links between two ports connected either directly or via intermediate stopover sites). A more severe problem is the missing data about the directionality of ballast water movements; although the direction of shipping traffic is explicitly considered in our model, directional ballast water transfers (i.e. uptake vs. deposition of water) are not explicitly incorporated. Given the demonstrated global patterns of shipping bulk resources (e.g. grain, raw materials), this could be a systematic bias in our risk estimations; however, due to the lack of data, this limitation cannot be resolved yet.

The second problem in modelling invasion dynamics is the correct formulation of the model itself. To begin with, bioinvasion is inherently a complex process. Beyond the presence of suitable environmental conditions, the probability of establishment is contingent on other factors, such as trophic levels, niche occupancy, population dynamics or biotic resistance (Lockwood *et al.*, 2007). These factors are relevant influential factors determining the success of establishment and act as sources of uncertainty within the model. Moreover, many of these factors are not fully understood and are still controversially debated (Lockwood *et al.*, 2007) and are often not easy to translate into surrogates which can be used for modelling. For example, according to the biotic resistance hypothesis (Kennedy *et al.*, 2002), the biotic pressure imposed on newly introduced individuals may be approximated in models by the species richness in the recipient community; however, this hypothesis is still under debate, particularly in marine systems (Fridley *et al.*, 2007; Stachowicz *et al.*, 2007). Other factors are inherently difficult to incorporate into a model setting that is designed to quantify invasion dynamics from a large pool of species with *a priori* unknown traits. For example, our model framework does not lend itself to describe the probability that an established species is a true invader in the sense that it poses negative impacts on the recipient ecosystem.

To test the robustness of our results we formulated different model extensions that: (1) consider the influence of species richness in the recipient or in the donor ports, (2) include hull fouling as an additional invasion vector or (3) implement a simplified description of shipping traffic assuming a constant ballast water release for all ships (see Appendix S1). For each model variant we applied a non-linear optimisation algorithm to find the parameter set with the best fit to the reported invasion data (Table S4). In general, the considered model variants did not differ significantly in their predictive power and fitted the data almost equally well. This indicates that the quality of the model cannot easily be improved by incorporating additional invasion mechanisms. In particular, the model fit was not improved by including hull fouling as another bioinvasion vector, but also not impaired significantly by simplifying the model description of ballast water release. The best fit was obtained by the model variant which implements a mechanism for biotic resistance in the recipient community, by assuming that the probability of establishment  $P(Estab)$  is a decaying function of species richness. Using this mechanism, the invasion risk in eco-regions with large biodiversity (i.e. mostly the tropical regions) is suppressed. Most notably, this reduces invasion risks in the Central and Western Indo-Pacific (orange bars in Fig. 4e) which yields a better overall description of the field data (Fig. 4j). However, as the empirical evidence for the biotic resistance is weak (Fridley *et al.*, 2007; Lockwood *et al.*, 2007; Stachowicz *et al.*, 2007), we did not include this mechanisms in the main model version in this study.

## CONCLUSIONS

The ability to forecast locations at the greatest risk of new invasions remains one of the main challenges confronting invasion biologists (Mack *et al.*, 2000; Kolar & Lodge, 2001). Among the uncertainties are the lack of detailed information about ballast water releases on a global scale, the relevance and implementation of biotic interactions and the likelihood that an established species poses negative impacts in the recipient ecosystem. Our study highlights that a forecast of invasions can only be achieved by combining information concerning ballast water transport, environmental heterogeneity and biogeographic distributions. Our vector-based modelling approach provides new avenues in this direction; it allows predicting expected shifts in invasion risks caused by global climate change (e.g. rising water temperature), changing trading patterns or new shipping routes (e.g. opening of Arctic passages), new ballast water management options (e.g. 'mid-ocean-exchange') and it may serve as a basis for the establishment of effective and targeted mitigation programs.

## ACKNOWLEDGEMENTS

We thank A. Abelson, C. Feenders, S. Gollasch, H. Hillebrand, A. Kölzsch and H. Rosenthal for helpful comments and D. Tittensor for kindly providing data of species richness. The study was supported by the German VW-Foundation. M.T.G. gratefully acknowledges financial support by Imperial College London.

## REFERENCES

- ABS (American Bureau of Shipping). (2011). *Ballast Water Treatment Advisory*. Available at: <http://www.eagle.org>. Last accessed 23 October 2012.
- Brawley, S.H., Coyer, J.A., Blakeslee, A.M.H., Hoarau, G., Johnson, L.E. & Byers, J.E. *et al.* (2009). Historical invasions of the intertidal zone of Atlantic North America associated with distinctive patterns of trade and emigration. *Proc. Natl. Acad. Sci. USA*, 106, 8239–8244.
- Cohen, A.N. & Carlton, J.T. (1998). Accelerating invasion rate in a highly invaded estuary. *Science*, 279, 555–558.
- Coles, S.L., DeFelice, R.C., Eldredge, L.G. & Carlton, J.T. (1999). Historical and recent introductions of non-indigenous marine species into Pearl Harbor, Oahu, Hawaiian Islands. *Mar. Biol.*, 135, 147–158.
- Drake, J.M. & Lodge, D.M. (2004). Global hot spots of biological invasions: evaluating options for ballast-water management. *Proc. R. Soc. Lond. B*, 271, 575–580.
- Floerl, O., Inglis, G.J., Dey, K. & Smith, A. (2009). The importance of transport hubs in stepping-stone invasions. *J. Appl. Ecol.*, 46, 37–45.
- Fridley, J.D., Stachowicz, J.J., Naeem, S., Sax, D.F., Seabloom, E.W. & Smith, M.D. *et al.* (2007). The invasion paradox: reconciling pattern and process in species invasions. *Ecology*, 88, 3–17.
- Hewitt, C.L., Campbell, M.L., Thresher, R.E., Martin, R.B., Boyd, S., Cohen, B.F. *et al.* (2004). Introduced and cryptogenic species in Port Phillip Bay, Victoria, Australia. *Mar. Biol.*, 144, 183–202.
- Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.*, 46, 10–18.
- IMO. (2004). *International Convention for the Control and Management of Ships' Ballast Water and Sediments*. International Maritime Organization, London.
- Kaluza, P., Kölzsch, A., Gastner, M.T. & Blasius, B. (2010). The complex network of global cargo ship movements. *J. R. Soc. Interface*, 7, 1093–1103.
- Keller, R.P., Drake, J.M., Drew, M.B. & Lodge, D.M. (2010). Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Divers. Distrib.*, 17, 93–102.
- Kennedy, T.A., Naeem, S., Howe, K.M., Knops, J.M.H., Tilman, D. & Reich, P. (2002). Biodiversity as a barrier to ecological invasion. *Nature*, 417, 636–638.



- Kölzsch, A. & Blasius, B. (2011). Indications of marine bioinvasion from network theory. An analysis of the global cargo ship network. *Europ. Phys. J. B*, 84, 601–612.
- Kolar, C.S. & Lodge, D.M. (2001). Progress in invasion biology: predicting invaders. *Trends Ecol. Evol.*, 16, 199–204.
- Kolar, C.S. & Lodge, D.M. (2002). Ecological predictions and risk assessment for alien fishes in North America. *Science*, 298, 1233–1236.
- Lockwood, J.L., Hoopes, M.F. & Marchetti, M.P. (2007). *Invasion Ecology*. Wiley-Blackwell, Malden, MA, USA.
- Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.*, 10, 689–710.
- McCutcheon, S., Martin, J. & Barnwell, T.J. (1993). Water quality. In: *Handbook of Hydrology*. (ed Maidment, D.). McGraw-Hill Professional, New York, NY, PP. 11.1–11.73.
- McGeoch, M.A., Butchart, S.H.M., Spear, D., Marais, E., Kleynhans, E.J. & Symes, A. et al. (2010). Global indicators of biological invasion: species numbers, biodiversity impact and policy responses. *Divers. Distrib.*, 16, 95–108.
- Minton, M.S., Verling, E., Miller, A.W. & Ruiz, G.M. (2005). Reducing propagule supply and coastal invasions via ships: effects of emerging strategies. *Front. Ecol. Environ.*, 3, 304–308.
- Molnar, J.L., Gamboa, R.L., Revenga, C. & Spalding, M.D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Front. Ecol. Environ.*, 6, 485–492.
- National Ballast Information Clearinghouse (NBIC) Online Database. (2012). Electronic Publication, Smithsonian Environmental Research Center & United States Coast Guard, 2008. Available at: <http://invasions.si.edu/nbic/search.html>. Last accessed 16 October 2012.
- Pielou, E. (1966). The measurement of diversity in different types of biological collections. *J. Theor. Biol.*, 13, 131–144.
- Reise, K., Gollasch, S. & Wolff, W.J. (1998). Introduced marine species of the North Sea coasts. *Helgol. Mar. Res.*, 52, 219–234.
- Ruiz, G.M., Carlton, J.T., Grosholz, E.D. & Hines, A.H. (1997). Global invasions of marine and estuarine habitats by non-indigenous species: mechanisms, extent, and consequences. *Integr. Comp. Biol.*, 37, 621–632.
- Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J. & Dirzo, R., et al. (2000). Biodiversity – global biodiversity scenarios for the year 2100. *Science*, 287, 1770–1774.
- Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A. & Finlayson, M.A.X., et al. (2007). Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, 57, 573–583.
- Stachowicz, J.J., Bruno, J.F. & Duffy, J.E. (2007). Understanding the effects of marine biodiversity on communities and ecosystems. *BioScience*, 57, 573–583.
- Tatem, A.J., Hay, S.I. & Rogers, D.J. (2006). Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. USA*, 103, 6242–6247.
- Thieltges, D.W., Ferguson, M.A.D., Jones, C.S., Krakau, M., de Montaudouin, X., & Noble, L.R., et al. (2009). Distance decay of similarity among parasite communities of three marine invertebrate hosts. *Oecologia*, 160, 163–173.
- Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, 299, 241–244.
- Verling, E., Ruiz, G.M., Smith, L.D., Galil, B., Miller, A.W. & Murphy, K.R. (2005). Supply-side invasion ecology: characterizing propagule pressure in coastal ecosystems. *Proc. R. Soc. Lond. B*, 272, 1249–1256.
- Wonham, M.J., Lewis, M.A. & MacIsaac, H.J. (2005). Minimizing invasion risk by reducing propagule pressure: a model for ballast-water exchange. *Front. Ecol. Environ.*, 3, 473–478.
- Wonham, M.J., Walton, W.C., Ruiz, G.M., Frese, A.M. & Galil, B.S. (2001). Going to the source: role of the invasion pathway in determining potential invaders. *Mar. Ecol.-Prog. Ser.*, 215, 1–12.

## SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library ([www.ecologyletters.com](http://www.ecologyletters.com)).

Editor, Franck Courchamp

Manuscript received 28 January 2013

First decision made 21 February 2013

Manuscript accepted 7 March 2013