A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

NeoBiota 101: 73–89 (2025)
DOI: 10.3897/neobiota.101.157775

Research Article

The spread of Indo-Pacific origin fish species in the Mediterranean Sea is influenced by sea currents, habitat factors, and increasingly by shipping

Heinrich zu Dohna[1,2], Issam Lakkis[3], Michel Bariche[1]

1 Department of Biology, American University of Beirut, 11–0236, Beirut, Lebanon
2 Department of Biology, American University in Dubai, Dubai, UAE
3 Department of Mechanical Engineering, American University of Beirut, 11–0236, Beirut, Lebanon
Corresponding author: Heinrich zu Dohna (heinrich.dohna@gmail.com)

Abstract

The construction of the Suez Canal initiated a profound marine biogeographic rearrangement by connecting two marine regions that had been separated for about 16 million years. To date, more than 100 fish species native to the Indo-West Pacific Ocean have crossed the Suez Canal and become established in the Mediterranean Sea. The purpose of this study is to identify the major factors that influence how fish species of Indo-Pacific origin (FSIPO) spread within the Mediterranean. We analyzed Mediterranean records of 136 FSIPOs, divided the Mediterranean Sea into polygons, and examined how sea currents, shipping, environmental variables, and species characteristics affect the probability that an FSIPO record is found in a polygon. Our analysis showed that any particular FSIPO is more likely to colonize regions that receive currents from the entrance of the Suez Canal, as well as regions that are close to – or receive currents or cargo ships from – other regions already occupied by the same species. Higher salinity is linked to increased colonization rates, and colonization rates overall show a significant increase over time. In addition, colonization due to spatial proximity to occupied regions and due to shipping increased significantly over time. Comparisons between simulated and observed rates of spread showed that there is additional variation among fish species that our analysis did not capture.

Key words: Invasion biology, Lessepsian migration, spatio-temporal spread

OPEN ACCESS

Academic editor: Tammy Robinson-Smythe
Received: 3 May 2025
Accepted: 23 August 2025
Published: 16 September 2025

Citation: zu Dohna H, Lakkis I, Bariche M (2025) The spread of Indo-Pacific origin fish species in the Mediterranean Sea is influenced by sea currents, habitat factors, and increasingly by shipping. NeoBiota 101: 73–89. https://doi.org/10.3897/neobiota.101.157775

Introduction

Invasive species can be a major threat to marine biodiversity and ecosystem functioning (Molnar et al. 2008). Successful prevention and mitigation of disruptions caused by invasive species require a detailed understanding of the factors that facilitate their spread.

One of the most invaded marine regions in the world is the Mediterranean Sea (Giangrande et al. 2020). A key conduit for non-native species entering the Mediterranean is the Suez Canal (Galil 2009). Species that are not native to the Mediterranean and enter it through the Suez Canal are commonly referred to as Lessepsian species. Studying how Lessepsian species spread in the Mediterranean is important for two reasons. Lessepsian species have the potential to fundamentally affect the ecology of the Mediterranean Sea, and a better understanding of their mode of spread can help mitigate their impact. Secondly, the process of repeated invasions from the same starting point provides an opportunity to gain general insights into the factors affecting the spread of marine invasive species.

Patterns of spread of invasive species have been shown to be influenced by sea currents (Jaspers et al. 2018). International shipping is also a major conduit for the transportation of invasive species (Molnar et al. 2008; Seebens et al. 2013). Shipping data can predict the occurrence of invasive species in broad geographic regions on a global scale (Seebens et al. 2013, 2016). However, the role of shipping and sea currents in the regional spread of Lessepsian species is less well understood.

While shipping and sea currents determine where propagules of fish species of Indo-Pacific origin (FSIPO) can end up, the local habitat conditions determine whether the individuals can establish a population. It is not clear how environmental variables such as temperature and salinity influence the establishment success of these species. Previous studies that related the thermal conditions in an FSIPO's native range to the species' dispersal and establishment success in the Mediterranean yielded conflicting results. Depending on the details of the analysis methods, preference for warmer regions in the native range is either not related to establishment success in the Mediterranean (Arndt and Schembri 2015), associated with higher establishment success (Belmaker et al. 2013), or associated with lower establishment success (Lasram et al. 2008). These studies focused on the habitat preferences of Lessepsian species rather than on the effects of habitat on establishment probabilities. The challenge of the latter approach is to disentangle habitat variables from effects that influence propagule pressure, such as sea currents.

A comprehensive understanding of the factors influencing the spread of Lessepsian species must also consider the effects of species-specific characteristics. Finally, there is evidence that the rate of introduction of Lessepsian species has changed over time (Samaha et al. 2016). Therefore, it is essential not only to consider the static effects of variables that may influence the spread but also to analyze the interaction of these variables with time.

In this study, we analyzed spatio-temporal Mediterranean records of FSIPO to investigate how surface currents, shipping, environmental factors, and ecological characteristics of species influence the spread of FSIPO in the Mediterranean over time. We found that the spread of FSIPO is influenced by sea currents and shipping rates. Regions with higher salinity were more likely to be colonized, and the apparent colonization rates increased over time. There is also evidence that cargo shipping has become a more important mode of spread over the years.

Methods

Assigning Lessepsian records to polygons

The Mediterranean coastline was obtained from the R package *maps* and divided into polygons of 250 km length and 0.6° width. Additional separate polygons were created to cover the coastline of each island, leading to a total of 54 polygons (Suppl. material 1: fig. S1). Spatio-temporal records of FSIPO in the Mediterranean were obtained from the ORMEF database (Azzurro et al. 2022). All fish species that were not of Indo-Pacific origin were removed from the ORMEF dataset. In the following text, we use the term "FSIPO" rather than "Lessepsian species," because the term "Lessepsian species" implies that a species entered through the Suez Canal. While it is likely that most FSIPOs in the Mediterranean entered through the Suez Canal, either by active dispersal or transport on a cargo vessel, our analysis is agnostic to the mode of entry. Based on the latitude and longitude of each FSIPO record, a distance to all polygons was calculated using the function *st_distance* from the R package *sf* (Pebesma 2018). According to this function, the distance from a point to a polygon is calculated as the distance from the point to the closest point in the polygon. Points that fall within a polygon have distance zero to that polygon. Records were assigned to the closest polygon. In total, 79% of records fell within polygons and therefore had zero distance to their closest polygon. The first record per polygon per species was retained. Twenty-five records were more than 100 km from the closest polygon and were excluded from the analysis, leading to 688 records of 136 species. For each of these species, information was obtained from FishBase (Froese and Pauly 2024) about their maximum total length, trophic position, carrying capacity, minimum and maximum depth, maximum age, and minimum, mean, and maximum temperature. The data were obtained in bulk using the R package *rfishbase* (Boettiger et al. 2012).

Polygon covariates

Mean and maximum sea surface temperature (SST) and salinity data were obtained from Bio-ORACLE (https://www.bio-oracle.org/), and average mean and maximum SST and salinity values were calculated per polygon. To contrast the conditions in the Mediterranean Sea with the northern Red Sea, average SST and salinity values were calculated for the Gulf of Suez as well. An FSIPO can be detected either through systematic scientific surveys or because it has been caught by commercial or sport fishing and identified by scientists. Fishing activity in a region might be proportional to population density, and the presence of trained scientists might be proportional to a region's GDP. To include variables that could influence species detection, the population size of major cities along the Mediterranean coast and their per capita GDP were obtained from Wikipedia and the CIA World Factbook (https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html). In addition, the locations of marine research institutes were obtained from the CIESM website. To obtain a polygon-specific average population density, population sizes of cities were inversely weighted by distance from the city and then averaged over polygons. The Mediterranean basin was divided into a raster of 0.2° × 0.2° pixels. The distance of each pixel to all cities was calculated, and a pixel-specific population density was determined by multiplying the population size times the per capita GDP of each city by a weight that declines exponentially with distance to the city (decay rate 0.01 km^{-1}) and summing the weighted contributions of all cities. The pixel-specific population densities were then averaged over each polygon. The same approach was used for calculating the densities of marine research institutes. Preliminary analysis showed that the estimated effects were qualitatively the same for different decay rates. A polygon's probability of receiving FSIPO is likely to be influenced by cargo vessels. Cargo vessel densities per polygon were estimated from publicly available data (obtained from https://www.emodnet-humanactivities.eu/download-data.php). These data contain densities of cargo vessels from 2017–2020 (Suppl. material 1: fig. S2). Densities were averaged over the 4 years and then averaged per polygon.

Probabilities to receive particles according to surface currents and distance

A Lagrangian particle tracking algorithm (El Mohtar et al. 2018) was used to calculate for each polygon the probability of receiving particles from each other polygon. In this study, the forward tracking algorithm employed daily velocity fields for 11,650 days spanning the period 1987–2018 from the physical reanalysis dataset provided by the E.U. Copernicus Monitoring Environment Marine Service (CMEMS, Simoncelli et al. 2014). The reanalysis data, with a horizontal grid resolution of 1/16° (ca. 6–7 km), are predicted by a hydrodynamics model supplied by the Nucleus for European Modelling of the Ocean (NEMO), with a variational data assimilation scheme (OceanVAR) for temperature and salinity vertical profiles and satellite sea-level anomaly along-track data. For this study, each day 1000 particles were placed at random positions within each polygon, and the movement of these particles was then simulated for 30 days. The probability that polygon i received particles from polygon j (element c_{ji} in matrix $\mathbf{C}$) was calculated as the proportion of particles that started in polygon j and ended up in polygon i at day 30. The probability matrix of sea current–based transport between pairs of polygons was calculated based on 629,100,000 (54 polygons × 1000 particles per polygon × 11,650 days) simulated particles. Distance-based equivalents of these probabilities (denoted by matrix $\mathbf{D}$) were calculated by fitting an exponential decay function through the plot of c_{ji} against the distance between the centroids of polygons i and j.

Shipping rates between polygons

The same data that were used to calculate cargo vessel densities were used to estimate between-polygon cargo shipping rates. First, densities were averaged over the 4 years. Then coastal points were defined as all points within an interior buffer with a width of 0.07° of a polygon of the Mediterranean Sea. Port points were identified as coastal points with a density of cargo vessels above 500. The coastal points, along with the port points, served as potential connection points for shipping routes. A graph was created by connecting all pairs of points that could be connected by straight lines without intersecting a coastline. Any shipping path between two ports was estimated as the shortest path along this graph of connection points. Additional connection points and barriers were introduced manually to improve the matching between estimated shipping routes and the lines of high vessel densities in the vessel density map (Suppl. material 1: fig. S2). To estimate the shipping rates between pairs of ports, the shipping density per line segment was calculated as the median shipping density along all pixels intersecting the line segment of a shipping path, divided by the number of shipping paths passing through the respective line segment. The shipping densities per segment were then averaged for all segments along a shipping path between two ports. This average served as an estimate of the shipping rate between two ports. The shipping rate between two polygons was calculated by summing the shipping densities of all shipping paths that connect the two polygons. The matrix of shipping rates between all polygons is denoted by $\mathbf{R}$.

Colonization through sea currents

To simplify the terminology in the following text, the occurrence of a first record in a polygon is referred to as a colonization event. Furthermore, a polygon is considered occupied by a particular species in year *t* if this species was recorded in the respective polygon in year *t* or any year prior to *t*. To determine how well sea currents predict initial colonization events, all particles that emanated from the polygon containing the Suez Canal (Polygon 30 in Suppl. material 1: fig. S1) in the sea current simulations were used. The proportion of these particles ending up in all polygons was compared to the proportion of FSIPO with their first Mediterranean record in the same polygons.

Fitting a dispersal model

Parameters of dispersal models were fitted by discretizing the time from the first to the last record in the OMREF dataset into yearly time steps and fitting logistic regressions. Only the 136 species with Indo-Pacific origin were considered. Each year, each polygon could potentially be colonized by any of the 136 species not previously recorded in that polygon, resulting in 888,624 observations in the regression model. The response variable in the regression model was a binary indicator of a colonization event in a particular polygon and year. According to the fitted models, each polygon could be colonized from other occupied polygons or from the Suez Canal. Both of these colonization rates could be based on distance, sea currents, or shipping rate, leading to six variables related to the propagule pressure experienced by a polygon. A seventh variable related to propagule pressure is the density of cargo vessels in a polygon. These variables are referred to as propagule pressure variables. The first six propagule pressure variables were calculated based on the three matrices **C**, **D,** and **R**, which correspond to sea currents, distance, and shipping between polygons, respectively. The element in the i^{th} row and j^{th} column of these matrices describes the probability that a propagule emanating from polygon *i* ends up in polygon *j*, according to the process described by the respective matrix. To account for propagule pressure from occupied polygons, the matrices describing sea currents, distance, and shipping between polygons were multiplied by binary variables indicating whether or not any source polygon was occupied in a given year. The effects of sea currents, distance, and shipping on the rates at which polygons receive propagules from the Suez Canal were approximated by the row vectors in **C**, **D,** and **R** that correspond to the polygon containing the Suez Canal (Polygon 30 in Suppl. material 1: fig. S1). The model also included the polygon-specific variables described in the section "*Polygon covariates*" (mean and maximum sea surface temperature (SST), salinity, human population density, marine research station density, proximity to the Suez Canal, and mean density of cargo vessels) and nine species-specific variables obtained from Fishbase (maximum total length, trophic position, carrying capacity, minimum and maximum depth, maximum age, minimum, mean, and maximum temperature).

There is evidence for an increased arrival rate of FSIPOs, even when accounting for imperfect detection of FSIPOs (Samaha et al. 2016). This increase could be due to multiple factors, such as the widening of the canal, increased shipping, increased surveillance, or other factors. To identify whether a specific factor interacts with time, the full regression included, along with each of the variables mentioned above, an interaction term of each variable with time. In a first round, the full model was fitted with a lasso penalty and a penalty parameter lambda that minimizes deviance in 10-fold cross-validation. The fitting was performed using the R package *glmnet* (Friedman et al. 2010). According to the full model, p_{ikt}, the probability that polygon i is colonized by species k in year t is given by

$$logit(p_{ikt}) = \beta_0 + (\beta_c \vec{c}_{*i} + \beta_d \vec{d}_{*i} + \beta_r \vec{r}_{*i}) \cdot \vec{x}_{kt} + \beta_{cs} c_{30i} + \beta_{ds} d_{30i} + \beta_{rs} r_{30i} + \beta_v v_i + \vec{\beta}_u \cdot \vec{u}_i$$
$$+ \vec{\beta}_z \cdot \vec{z}_k + t[(\beta_{c2} \vec{c}_{*i} + \beta_{d2} \vec{d}_{*i} + \beta_{r2} \vec{r}_{*i}) \cdot \vec{x}_{kt} + \beta_{cs2} c_{30i} + \beta_{ds2} d_{30i} + \beta_{rs2} r_{30i} + \beta_{v2} v_i + \vec{\beta}_u \cdot$$
$$\vec{u}_i + \vec{\beta}_z \cdot \vec{z}_k]$$

where $\vec{c}_{*i}$, $\vec{d}_{*i}$, and $\vec{r}_{*i}$ denote, respectively, the i^{th} column vector in the matrices **C**, **D,** and **R,** the coefficients β_c, β_d, and β_r denote, respectively, the coefficients for the effect of sea currents, distance, and shipping from polygons occupied by the same species in year t-1 on the probability of a first record in polygon i in year t. The vector $\vec{x}_{kt}$ denotes the polygon occupancy by species k at time step t-1, the dot denotes the dot product $\vec{x}_{kt}$ and is equal to one if the first record of species k in polygon i occurred before t and zero otherwise. c_{30i}, d_{30i}, and r_{30i} denote entries in the matrices **C**, **D,** and **R,** in row 30 (corresponding to the polygon containing the Suez Canal) and column i, denotes the density of cargo vessels in polygon i and β_v the respective regression coefficient, denotes the vector of environmental variables (SST, salinity, and human population density) of polygon i, $\vec{\beta}_u$ the vector of regression coefficients for the environmental variables, $\vec{z}_k$ denotes the vector of characteristics of species k (maximum total length, trophic position, carrying capacity, minimum and maximum depth, maximum age, minimum, mean, and maximum temperature) and $\vec{\beta}_z$ the vector of regression coefficients for these species characteristics. The regression coefficients with the subscript 2 correspond to the coefficients for the interaction of each variable with time.

The penalized regression can identify important predictors from a large set of potential predictors, but it does not provide unbiased estimates or p-values for regression coefficients. The variables that yielded non-zero coefficients in the penalized regression were therefore included in three more regressions without penalty terms. First, a regression with only additive terms and a random species effect (using the R function *glmer*) was performed. Variables that yielded significant coefficients in that model were used to fit a regression model with additive terms and a random species effect and a model where the propagule pressure variables interact with the environmental variables in a multiplicative way. The rationale for the multiplicative interaction is that environmental variables can only modify the colonization potential. If, for example, a polygon receives no colonizers from the Suez Canal or other polygons, then environmental variables cannot make up for the lack of colonizers. The structure of the multiplicative model was as follows:

$$logit(p_{ikt}) = \beta_0 + [(\beta_c \vec{c}_{*i} + \beta_d \vec{d}_{*i} + \beta_r \vec{r}_{*i}) \cdot \vec{x}_{kt} + \beta_{cs} c_{30i} + \beta_{ds} d_{30i} + \beta_{rs} r_{30i} + \beta_v v_i]$$

$$[1 + \vec{\beta}_u \cdot \vec{u}_i + \vec{\beta}_z \cdot \vec{z}_k + t(\vec{\beta}_{u2} \cdot \vec{u}_i + \vec{\beta}_{z2} \cdot \vec{z}_k)] + t[(\beta_{c2}\, \vec{c}_{*i} + \beta_{d2}\, \vec{d}_{*i} + \beta_{r2} \vec{r}_{*i}) \cdot \vec{x}_{kt} + \beta_{cs2} c_{30i}$$

$$+ \beta_{ds2} d_{30i} + \beta_{rs2} r_{30i} + \beta_{v2} v_i][1 + \vec{\beta}_u \cdot \vec{u}_i + \vec{\beta}_z \cdot \vec{z}_k]$$

The regression coefficients for propagule pressure variables and their interactions with time appear twice in this model, in the additive and multiplicative terms. This model was fitted through non-linear minimization using the R function *nlim*. To improve the numerical stability of the different fitting procedures, all variables that did not involve dispersal between polygons were standardized to have a minimum and standard deviation of one. The variables concerning the effects between polygons were not standardized to facilitate the use of the estimated coefficients for the simulation model (see below).

Invasive species research distinguishes between different stages of the invasion process, such as transport, establishment, and secondary spread (Blackburn et al. 2011; Daly et al. 2023). The regression model above uses a response variable that indicates whether or not a published first record of a particular FSIPO occurred in a given polygon in a given year. Long-range fish dispersal usually happens in the larval stage, but new species are recorded primarily in the adult stage. For a first record in a new polygon, the FSIPO had to be transported there as larvae, settle, and reach adulthood. Furthermore, it is unlikely that the first dispersers were detected. A new record in a polygon therefore implies transport and establishment on the scale of the polygon. On the scale of the entire Mediterranean, regression coefficients for propagule pressure from the Suez Canal (β_{cs}, β_{ds}, β_{rs}) quantify primary transport, whereas regression coefficients for propagule pressure from occupied polygons (β_c, β_d, β_r) quantify secondary spread within the Mediterranean. The regression coefficients associated with the temperature and salinity in a polygon indicate the contributions of local conditions to establishment.

Quantifying the effect of shipping

The estimated regression coefficients of the best-fitting model were used to quantify the effects of shipping on the spread of FSIPOs. For each colonization event, the odds ratio was calculated to determine whether this event occurred due to shipping. Specifically, the ratio of the estimated colonization odds with shipping variables over the colonization odds without shipping variables was calculated. The model included six potential shipping variables, namely the density of cargo vessels per polygon, the estimated between-polygon shipping rates, the estimated shipping rates from the Suez Canal, and the interaction of these variables with time. The variables that were associated with a significant regression coefficient in the best-fitting model were included in the calculation. The shipping rate at which a particular FSIPO arrives at a particular polygon in a given year was calculated by multiplying the estimated between-polygon shipping rate with an indicator variable for the occupation of source polygons by an FSIPO (see regression equations above). Hence, the exposure of a polygon to invasion by a particular FSIPO due to shipping changed from year to year. Therefore, the odds ratio for shipping was calculated separately for each colonization event. These odds ratios were then averaged per polygon to identify regions that are particularly prone to FSIPO spread through cargo shipping.

Simulating dispersal

A standard analysis of the regression model fit does not indicate how well the fitted model predicts the invasion dynamics. The coefficients of the best-fitting logistic regression model were therefore used to simulate the spread of FSIPO. Let $\vec{x}_t$ denote the vector of polygon occupancy at time step t, where x_{it} is equal to one if polygon i becomes colonized before t and zero otherwise. The simulation starts in the year 1896 with $x_{i0} = 0$ for all polygons. At each step, the colonization probabilities of each polygon are calculated according to the estimated regression coefficients, the current t and $\vec{x}_{t-1}$. New x_{it} are sampled according to a Bernoulli process. The process was run in 1000 separate simulations for 124 time steps. Two measures were used to compare the simulations with the observed data. The proportion of simulation runs in which a particular polygon was occupied at the end of the run was compared to the proportion of species recorded in that polygon. The frequency distribution of total polygons occupied in a simulation was compared to the observed distribution of polygons occupied among FSIPOs. The proportion of simulations that resulted in a single occupied polygon was adjusted to match the observed proportion of FSIPOs that occur in a single polygon. This adjustment was done to account for the fact that invasions that have occurred in a single polygon might be undercounted because of imperfect detection. The comparisons between observed and simulated invasions indicate how well the data match the assumption implicit in the regression model that each FSIPO is an independent realization of the same underlying invasion process that is governed by the parameters fitted in the logistic regression. To distinguish these simulations from the simulated particles based on sea currents, we will refer to the simulations described in this section as “invasion simulation” and the particle simulation as “sea current simulation.”

Results

General patterns

Human population density varies widely along the Mediterranean coast, with the highest densities in the northwestern Mediterranean and the Levantine coast and the lowest density along the coasts of Libya and eastern Egypt (Suppl. material 1: fig. S3). The mean SST ranges from 17 °C in the northern parts of the Mediterranean to 23 °C in the southeast (Suppl. material 1: fig. S3). Within the Mediterranean, salinity increases from west to east (Suppl. material 1: fig. S3). The highest mean and maximum SST values in the Mediterranean exceed those in the Gulf of Suez (Suppl. material 1: fig. S4). Both the mean and maximum salinity are higher in the Gulf of Suez than anywhere in the Mediterranean Sea (Suppl. material 1: fig. S4). The largest numbers of FSIPOs were recorded along the Levantine coast (Fig. 1A).

Effects of sea currents

The spatial patterns of the first Mediterranean records of FSIPOs broadly follow the patterns of simulated particles that emanated from the polygon containing the Suez Canal (Fig. 1B–D). According to the sea current simulations, average coastal sea currents generally flow in a counter-clockwise direction in the Mediterranean (Fig. 1E). While there is a general correspondence between the number

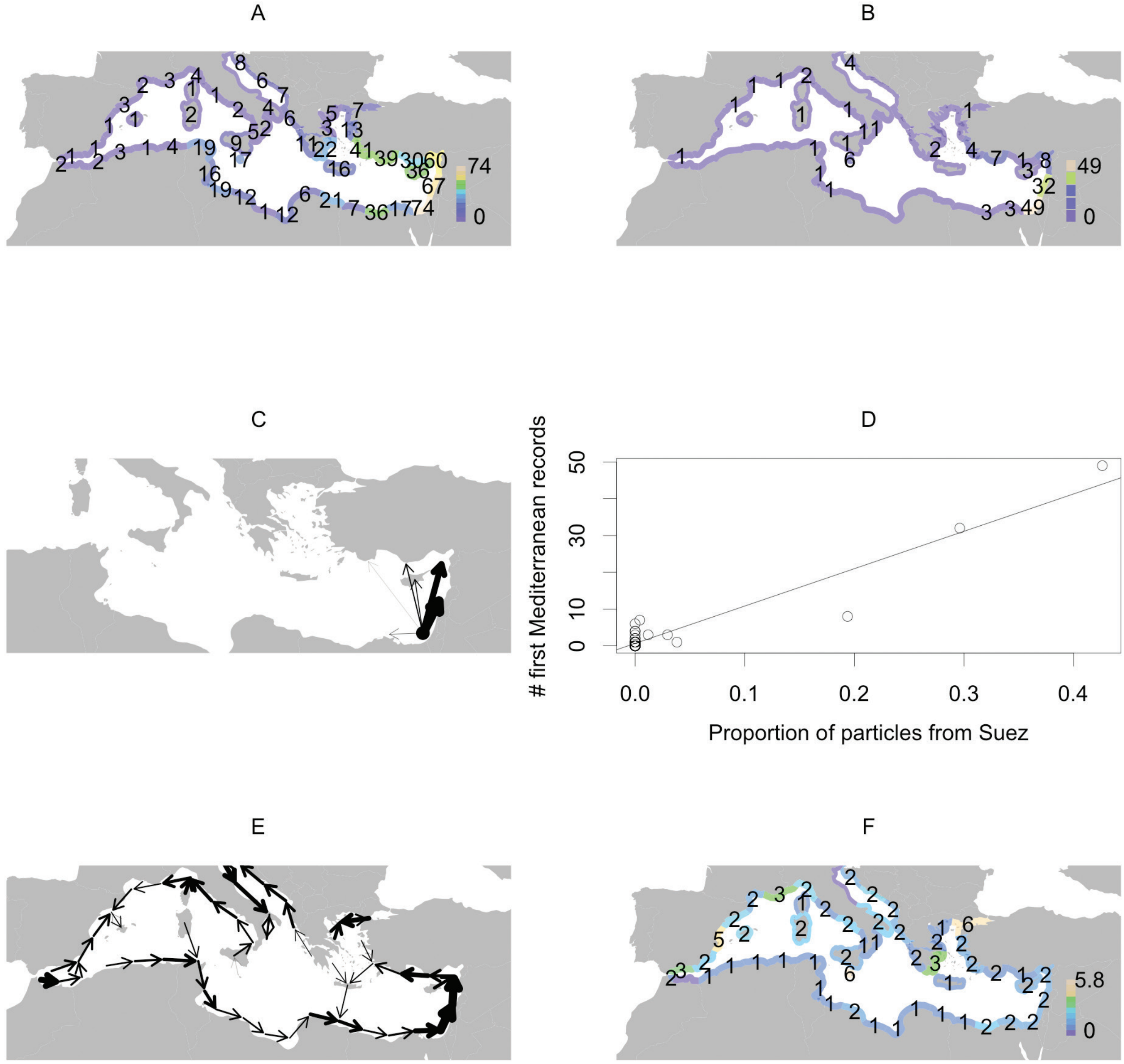

Figure 1. Patterns among polygons in the Mediterranean. **A.** Total number of FSIPOs recorded per polygon. **B.** Number of FSIPOs with the first Mediterranean record in the respective polygon. **C.** Arrows indicate the number of simulated particles flowing from the polygon containing the Suez Canal to all other polygons. Arrow width is proportional to the number of particles. **D.** Number of species with the first Mediterranean record in a polygon vs. the proportion of particles from the polygon containing the Suez Canal that end up in the respective polygon. **E.** Arrows point from the centroid of each source polygon to the centroid of the target polygon that received the highest proportion of simulated particles among all particles emanating from the source polygon. Arrow width is proportional to the number of particles flowing from source to target polygons. **F.** Rounded average odds ratio per polygon for colonization through shipping.

of simulated particles a polygon received from the Suez Canal and the number of FSIPOs whose first Mediterranean record was found in that polygon (Fig. 1D), there are also some notable discrepancies. For example, there are 16 polygons that did not receive simulated particles from the Suez Canal but have at least one species with a Mediterranean-wide first record (Fig. 1B). The most extreme case is Malta, which is far from any polygon that obtained simulated particles from the Suez Canal (Fig. 1C), yet the first Mediterranean records of six species were found there (Fig. 1B). Malta also shows the highest odds ratio for colonization due to shipping (Fig. 1F).

Predictors of FSIPO records

The penalized logistic regression model selected 12 variables that predict the occurrence of an FSIPO record in a polygon. Nine of these predictor variables were statistically significant when included in the additive logistic regression model without a penalty term (Table 1). Six of the nine significant predictor variables were propagule pressure variables, and two of these interacted significantly with time. The variables that did not interact with time were (i) the probability of receiving simulated particles from the Suez Canal, (ii) the spatial proximity to the Suez Canal, (iii) the probability of receiving simulated particles from other polygons occupied by the same species, and (iv) the shipping rates from other polygons occupied by the same species. All of these variables are associated with significant positive regression coefficients. The two propagule pressure variables that positively interacted with time were (i) the spatial proximity to polygons occupied by the same FSIPO and (ii) the density of cargo vessels in a polygon (Table 1). Furthermore, colonization probability in a polygon increased with average sea surface temperature, maximum salinity, and time (Table 1). A second model was fitted in which propagule pressure variables interact multiplicatively with other variables. The effect of all variables except for temperature was qualitatively similar in this model (Table 1). The raw coefficients and the calculation of the net variable effects are explained in Suppl. material 1: table S1. The additive model fitted the data substantially better than the multiplicative one (ΔAIC = 312.5).

Simulated FSIPO spread

When the spread of FSIPOs was simulated based on regression coefficients, the number of times a species was recorded in a polygon in the simulation was significantly correlated with the actual total number of species recorded in a polygon ($R = 0.72$, $p = 6.48 \times 10^{-10}$). The total number of occupied polygons varies more strongly between different fish species than between different simulation runs (Fig. 2).

Table 1. Overview of regression coefficients, mean net effects*, and associated p-values for coefficients that are significant in the additive logistic regression model.

	Additive Model		Multiplicative Model	
Variable	Coefficient	P-value	Mean net effect*	P-value
Sea currents from Suez Canal	0.17	<0.0001	0.13	<0.0001
Proximity to Suez Canal	0.39	<0.0001	0.70	<0.0001
Sea currents from polygons occupied by the same FSIPO	2.50	<0.0001	2.44	<0.0001
Shipping from polygons occupied by the same FSIPO	0.015	0.00061	0.017	0.00038
Mean sea surface temperature	0.23	0.0069	-0.23	0.0074
Maximum salinity	0.36	<0.0001	0.27	<0.0001
Year	1.36	<0.0001	0.69	<0.0001
Proximity to polygons occupied by the same FSIPO × Year	1.85	<0.0001	1.78	<0.0001
Cargo ship density × Year	0.070	<0.0001	0.082	<0.0001

* The interpretation of regression coefficients is not straightforward for the multiplicative model. The raw regression coefficients and an explanation of how the mean net effects were calculated are provided in Suppl. material 1.

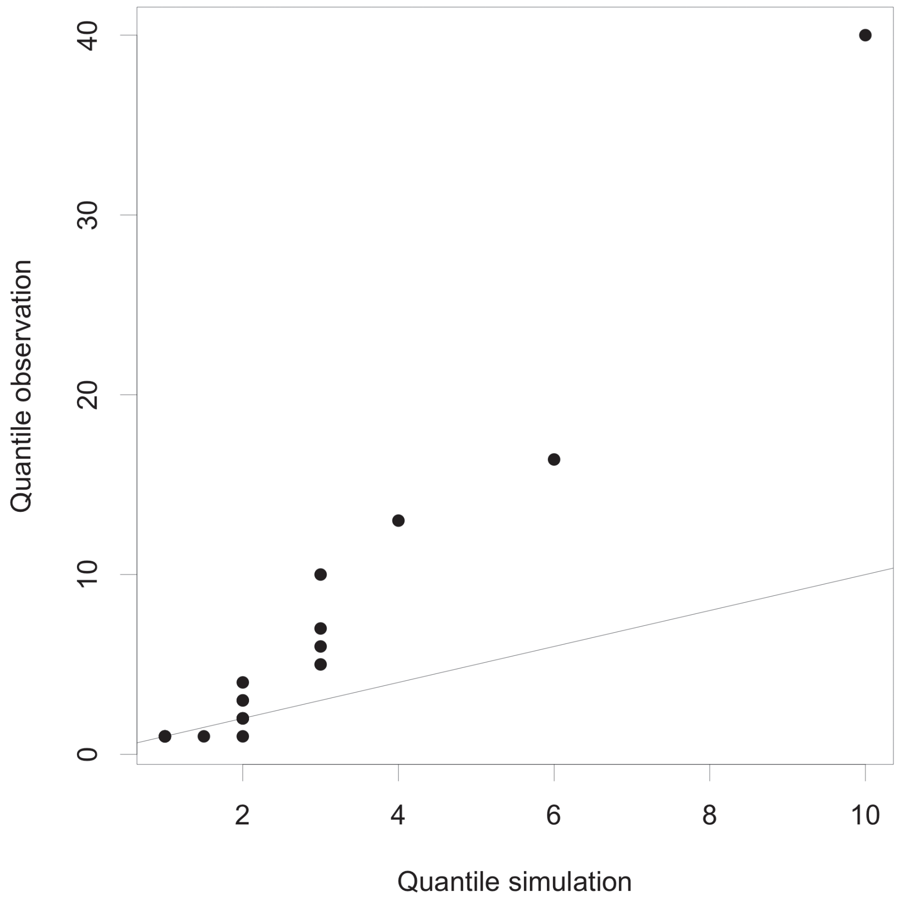

Figure 2. Quantiles of the number of polygons occupied by each species vs. quantiles of polygons occupied at the end of each simulation.

Discussion

The Mediterranean coastline was divided into polygons, and different logistic regression models were fitted to predict first records of FSIPOs per polygon. These models show that the probability of being colonized by a particular FSIPO is higher in regions that are close to or receive currents from the entrance of the Suez Canal, as well as regions that are close to or receive currents or cargo ships from regions already occupied by the same FSIPO. In all models, higher salinity was associated with higher colonization rates, and colonization rates increased significantly over time. The effects of cargo shipping density and proximity to occupied polygons also increased over time. The effects of temperature were model-dependent. Simulations based on parameters estimated by logistic regression models generally agreed with the observed patterns, particularly during the early stage of the colonization process. However, the total number of occupied polygons varied more strongly among different FSIPOs than between the simulations. The average odds ratio for colonization due to shipping was particularly high for Malta.

Transport into the Mediterranean

The first stage of the invasion process is the transport of individuals into a new area (Blackburn et al. 2011; Daly et al. 2023). Our regression models show that an area's exposure to initial FSIPO transport into the Mediterranean is determined by its geographic proximity to the Suez Canal and by the amount of sea currents it receives from the area around the Suez Canal. The importance of sea currents from the Suez Canal for initial transport is also evident from the strong correspondence between the number of first Mediterranean records in a polygon and the number of particles it receives according to simulated sea currents (Fig. 1D). Our regression model also showed that a higher density of cargo ships in a polygon is associated with increased FSIPO colonization. This variable is agnostic to the source of colonizers, and it is plausible that it partially covers initial transport into the Mediterranean. The conjecture that cargo shipping contributes to initial transport into the Mediterranean is supported by spatial patterns. Malta, for example, received six first Mediterranean FSIPO records even though it is far from the Suez Canal and does not receive any currents (Fig. 1B). Malta is also a central hub for the Mediterranean shipping network (Ducruet and Notteboom 2021).

Establishment

The next stage of the invasion process is establishment (Blackburn et al. 2011; Daly et al. 2023). Establishment is determined by local conditions, such as temperature and salinity. In both regression models, higher maximum salinity values of polygons were correlated with a polygon's higher propensity to become colonized. The qualitative results for temperature and salinity were not dependent on whether the other variable was included in the model, indicating that the results were not driven by the correlation between temperature and salinity. Given that most individuals of FSIPOs might have originated from the Gulf of Suez, which has higher salinity values than any region in the Mediterranean (Suppl. material 1: figs S4, S5), one would expect that these individuals are adapted to high-salinity environments. The results of our study are consistent with this expectation. Salinity has recently been recognized as an important driver of where FSIPOs can establish in the Mediterranean (D'Amen et al. 2023).

The effect of temperature was less straightforward. Temperature had a positive effect on colonization in the additive model but a negative effect in the multiplicative model. This apparent contradiction can be explained by the fact that colder areas in the northern and western Mediterranean are distant from the eastern Mediterranean and therefore out of reach for colonizers from the Suez Canal or from occupied regions in the eastern Mediterranean. In the additive model, the low colonization probability of the northern and western Mediterranean is partially attributed to its lower temperature. In the multiplicative model, the effect of temperature only mattered in regions that could receive colonizers. For most species and time points, these were the regions close to the Suez Canal or in the eastern Mediterranean. The multiplicative model suggests that within regions close to the Suez Canal or close to areas occupied by many FSIPOs, warmer areas are less likely to be colonized than colder areas. This is counterintuitive since the native range of FSIPOs typically encompasses tropical and subtropical regions, whereas the Mediterranean is mostly temperate (D'Amen et al. 2022). However, our analysis shows that the mean and maximum SST in the warmest regions of the Mediterranean exceed the values observed in the Gulf of Suez (Suppl. material 1: fig. S4). Hence, the warmest regions in the Mediterranean, such as the coast of Egypt east of the Suez Canal and the southern Levantine coast (Suppl. material 1: fig. S5), might exceed the temperature tolerance of individuals that crossed from the Gulf of Suez.

Many previous studies have analyzed the importance of temperature for the spread of FSIPOs, either by relating geographic and temporal differences in temperature to species abundance (Givan et al. 2018) or by analyzing the correlation between temperature in FSIPOs' native ranges and their establishment success in the Mediterranean (Lasram et al. 2008; Belmaker et al. 2013; Arndt and Schembri 2015). The results of these studies are contradictory. There is evidence that temperature is a strong determinant of fish abundance (Givan et al. 2018). Nevertheless, depending on how establishment success and temperature in the native range are measured, establishment success in the Mediterranean is either associated with lower temperature in the native range (Lasram et al. 2008), higher temperature in the native range (Belmaker et al. 2013), or unrelated to temperature in the native range (Arndt and Schembri 2015). The result of our multiplicative model is most in line with an earlier study that found that among Red Sea species, higher maximum SST in their native range correlates positively with the propensity to become established in the Mediterranean (Belmaker et al. 2013). The authors conjectured that maximum summer temperatures in the eastern Mediterranean can exceed the maxima in the native range limit of Lessepsian species. Our results support this conjecture.

The fact that in our analysis different models produced opposite effects of temperature on establishment underscores that proper accounting for variation in propagule pressure is necessary for an accurate estimation of the effects of local conditions on establishment.

Secondary spread

The next stage of the invasion process is secondary spread, i.e., the spread from areas where the invasive species first established to new areas (Blackburn et al. 2011; Daly et al. 2023). According to our fitted models, the spread of FSIPOs from occupied to unoccupied polygons in the Mediterranean is facilitated by spatial proximity, sea currents, and cargo shipping. These results hold for both the additive and multiplicative regression models. Our results also suggest temporal trends in secondary spread. Geographic proximity to occupied polygons interacts significantly with time, i.e., the importance of geographic proximity increases over time. This increase might be due to the fact that local populations continue to grow, so that colonization from nearby polygons becomes a more important aspect of spread within the Mediterranean.

Our results also demonstrate the importance of cargo shipping for the secondary spread of FSIPOs in the Mediterranean. A polygon's probability of becoming colonized by an FSIPO is significantly associated with the cargo shipping it receives from polygons already occupied by the same FSIPO. The regression models further show that a polygon's colonization probability increases with the product of time and the polygon's density of cargo ships. A polygon's density of cargo ships is a measure of how many ships an area receives and therefore encompasses the role of cargo shipping for both initial transport and secondary spread. The positive interaction with time indicates that the impact of cargo shipping on FSIPO spread has increased over time.

Our analysis also confirmed the special role of Malta in the spread of FSIPOs through cargo shipping. Malta is a central hub for the Mediterranean shipping network, where containers are moved between vessels (Ducruet and Notteboom 2021), making it a potential hotspot for the release of ballast water. Previous studies have identified Malta as a center for alien species introduced into the Mediterranean by cargo ships (Nunes et al. 2014; Otero et al. 2013). Whether an FSIPO has been introduced by cargo shipping is usually inferred from the location of the first record (Katsanevakis et al. 2013). An FSIPO whose first Mediterranean record occurred in Malta instead of the Levant is deemed to be introduced by cargo ships rather than sea currents. Our approach made no such assumption. Instead, we estimated the relationship between cargo vessel density and colonization probability while accounting for sea currents and spatial proximity. These estimates were then used to calculate, for each colonization event, the odds ratio that the colonization occurred due to cargo shipping. This alternative approach provides a quantitative confirmation of Malta's importance for the spread of FSIPOs through cargo shipping. Since our results do not rely on inferences based on the first Mediterranean record, they also apply to secondary spread of FSIPOs through shipping, even when the initial introduction into the Mediterranean happened through other means.

Model fit and simulation

Two types of regression models were fitted: one in which separate additive coefficients were fitted for all variables, and one in which coefficients were fitted for the multiplicative interactions between propagule pressure variables and environmental variables. The multiplicative interaction model should be a better description of reality, since environmental variables that facilitate establishment by FSIPOs only matter when colonizing individuals arrive from the Suez Canal or other occupied areas. Nevertheless, the additive model fitted the data substantially better. The most likely reason for the better fit of the additive model is that the propagule pressure variables in the model did not fully describe the actual spatial variation in propagule pressure. In particular, our method for estimating dispersal through ballast tanks was very crude. Actual data on vessel movement and the uptake and release of ballast water would provide much more accurate predictors of dispersal through vessels. However, these data are of commercial interest and were not available for our study.

Coefficients of the fitted additive regression model were used to simulate how FSIPOs spread through the Mediterranean. Overall, there was good agreement between simulated and observed patterns of spread, with a notable exception. There was a much larger variance in the total number of occupied polygons among the observed than the simulated invasions. Both simulations and observed invasions had a median of 2 occupied polygons, but there were pronounced differences between observed and simulated invasions at the larger quantiles (Fig. 2). Among observed invasions, the species at the 99th percentile, the bluespotted cornetfish (*Fistularia commersonii*), was recorded in 40 polygons, whereas the 99th percentile of simulated invasions occupied only 10 polygons (Fig. 2). Since the simulations were based on the results of the regression model, the fact that the number of occupied polygons varies less among the simulated than the observed invasions suggests that there is variation among FSIPOs that the regression model did not account for. The regression model did not capture between-species variation despite the inclusion of an additive random species effect in the regression and simulation models. Hence, the observed between-species variation is driven by non-additive interactions with other variables. These could be any number of ecological differences related to dispersal behavior and/or pre-adaptation to conditions in the Mediterranean Sea. In general, a regression model with a limited number of observations and predictor variables can elucidate general patterns underlying invasion dynamics of multiple species but is unlikely to fully account for the idiosyncrasies inherent in ecological systems that determine between-species variation in invasion success. In other words, the type of model employed in this study can show general trends but cannot predict outliers.

Conclusion

Our analysis of FSIPOs shows that sea currents play an important role early in the invasion process but cannot explain how species cross into the western Mediterranean. The estimated effects of cargo shipping on FSIPO spread demonstrate that Malta plays a key role in species crossing into the western Mediterranean. More granular data on cargo shipping are necessary to better understand the link between shipping and the spread of FSIPOs.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

H.D. and M.B. were supported by the American University of Beirut – University Research Board (URB) Faculty Research Grants Program 23939.

Author contributions

Conceptualization: MB, HD, IL. Data curation: MB. Formal analysis: HD, IL.

Author ORCIDs

Heinrich zu Dohna https://orcid.org/0000-0003-0474-9628
Michel Bariche https://orcid.org/0000-0001-6831-4311

Data availability

Data and code used for this publication have been uploaded to Zenodo: https://zenodo.org/records/15025754.

References

Arndt E, Schembri PJ (2015) Common traits associated with establishment and spread of Lessepsian fishes in the Mediterranean Sea. Marine Biology 162: 2141–2153. https://doi.org/10.1007/s00227-015-2744-3

Azzurro E, Smeraldo S, D'Amen M (2022) ORMEF: Occurrence Records of Mediterranean Exotic Fishes database. https://www.seanoe.org/data/00730/84182/

Belmaker J, Parravicini V, Kulbicki M (2013) Ecological traits and environmental affinity explain Red Sea fish introduction into the Mediterranean. Global Change Biology 19: 1373–1382. https://doi.org/10.1111/gcb.12132

Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023

Boettiger C, Lang DT, Wainwright PC (2012) rfishbase: Exploring, manipulating and visualizing FishBase data from R. Journal of Fish Biology 81: 2030–2039. https://doi.org/10.1111/j.1095-8649.2012.03464.x

D'Amen M, Smeraldo S, Di Franco A, Azzurro E (2022) The spread of Lessepsian fish does not track native temperature conditions. ICES Journal of Marine Science 79: 1864–1873. https://doi.org/10.1093/icesjms/fsac121

D'Amen M, Smeraldo S, Azzurro E (2023) Salinity, not only temperature, drives tropical fish invasions in the Mediterranean Sea, and surface-only variables explain it better. Coral Reefs 42: 467–472. https://doi.org/10.1007/s00338-023-02357-9

Daly EZ, Chabrerie O, Massol F, Facon B, Hess MC, Tasiemski A, Grandjean F, Chauvat M, Viard F, Forey E, Folcher E, Buisson E, Boivin T, Baltora-Rosset S, Ulmer R, Gibert P, Thiébaut G, Pantel JH, Heger T, Richardson DM, Renault D (2023) A synthesis of biological invasion hypotheses associated with the introduction–naturalisation–invasion continuum. Oikos 2023: e09645. https://doi.org/10.1111/oik.09645

Ducruet C, Notteboom T (2021) The global maritime network in container shipping: Liner service design, network configuration and port centrality. In: Dong-Wook S, Photos P (Eds) Maritime logistics: a guide to contemporary shipping and port management, 148–166.

El Mohtar S, Hoteit I, Knio O, Issa L, Lakkis I (2018) Lagrangian tracking in stochastic fields with application to an ensemble of velocity fields in the Red Sea. Ocean Modelling 131: 1–14. https://doi.org/10.1016/j.ocemod.2018.08.008

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33: 1–22. https://doi.org/10.18637/jss.v033.i01

Froese R, Pauly D (Eds.) 2024. FishBase. World Wide Web electronic publication. www.fishbase.org [06/2024]

Galil BS (2009) Taking stock: Inventory of alien species in the Mediterranean Sea. Biological Invasions 11: 359–372. https://doi.org/10.1007/s10530-008-9253-y

Giangrande A, Pierri C, Del Pasqua M, Gravili C, Gambi MC, Gravina MF (2020) The Mediterranean in check: Biological invasions in a changing sea. Marine Ecology (Berlin) 41: e12583. https://doi.org/10.1111/maec.12583

Givan O, Edelist D, Sonin O, Belmaker J (2018) Thermal affinity as the dominant factor changing Mediterranean fish abundances. Global Change Biology 24: e80–e89. https://doi.org/10.1111/gcb.13835

Jaspers C, Huwer B, Antajan E, Hosia A, Hinrichsen HH, Biastoch A, Angel D, Asmus R, Augustin C, Bagheri S, Beggs SE, Balsby TJS, Boersma M, Bonnet D, Christensen JT, Dänhardt A, Delpy F, Falkenhaug T, Finenko G, Fleming NEC, Fuentes V, Galil B, Gittenberger A, Griffin DC, Haslob H, Javidpour J, Kamburska L, Kube S, Langenberg VT, Lehtiniemi M, Lombard F, Malzahn A, Marambio M, Mihneva V, Møller LF, Niermann U, Okyar MI, Özdemir ZB, Pitois S, Reusch TBH, Robbens J, Stefanova K, Thibault D, van der Veer HW, Vansteenbrugge L, van Walraven L, Woźniczka A (2018) Ocean current connectivity propelling the secondary spread of a marine invasive comb jelly across western Eurasia. Global Ecology and Biogeography 27: 814–827. https://doi.org/10.1111/geb.12742

Katsanevakis S, Zenetos A, Belchior C, Cardoso AC (2013) Invading European Seas: Assessing pathways of introduction of marine aliens. Ocean and Coastal Management 76: 64–74. https://doi.org/10.1016/j.ocecoaman.2013.02.024

Lasram FBR, Tomasini JA, Guilhaumon F, Romdhane MS, Do Chi T, Mouillot D (2008) Ecological correlates of dispersal success of Lessepsian fishes. Marine Ecology Progress Series 363: 273–286. https://doi.org/10.3354/meps07474

Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment 6: 485–492. https://doi.org/10.1890/070064

Nunes AL, Katsanevakis S, Zenetos A, Cardoso AC (2014) Gateways to alien invasions in the European seas. Aquatic Invasions 9: 133–144. https://doi.org/10.3391/ai.2014.9.2.02

Otero M, Cebrian E, Francour P, Galil B, Savini D (2013). Monitoring marine invasive species in Mediterranean marine protected areas (MPAs): a strategy and practical guide for managers. IUCN, Malaga, 136.

Pebesma E (2018) Simple features for R: Standardized support for spatial vector data. The R Journal 10: 439–446. https://doi.org/10.32614/RJ-2018-009

Samaha C, zu Dohna H, Bariche M (2016) Analysis of Red Sea fish species' introductions into the Mediterranean reveals shifts in introduction patterns. Journal of Biogeography 43: 1797–1807. https://doi.org/10.1111/jbi.12793

Seebens H, Gastner MT, Blasius B (2013) The risk of marine bioinvasion caused by global shipping. Ecology Letters 16: 782–790. https://doi.org/10.1111/ele.12111

Seebens H, Schwartz N, Schupp PJ, Blasius B (2016) Predicting the spread of marine species introduced by global shipping. Proceedings of the National Academy of Sciences of the United States of America 113: 5646–5651. https://doi.org/10.1073/pnas.1524427113

Simoncelli S, Fratianni C, Pinardi N, Grandi A, Drudi M, Oddo P, Dobricic S (2014) Mediterranean Sea Physical Reanalysis (CMEMS MED-Physics) (Version 1) [Data set]. Copernicus Monitoring Environment Marine Service (CMEMS). https://doi.org/10.25423/MEDSEA_REANALYSIS_PHYS_006_004

Supplementary material 1

Supplementary figures and tables

Authors: Heinrich zu Dohna, Issam Lakkis, Michel Bariche

Data type: docx

Link: https://doi.org/10.3897/neobiota.101.157775.suppl1