

# Analysis of Red Sea fish species' introductions into the Mediterranean reveals shifts in introduction patterns

Carol Samaha, Heinrich zu Dohna\* and Michel Bariche

Department of Biology, American University  
of Beirut, 11–0236 Beirut, Lebanon

## ABSTRACT

**Aim** The Mediterranean Sea is one of the most threatened marine systems in the world. One of the major threats is the introduction of alien species from the Red Sea through the Suez Canal into the Mediterranean. The aim of our study was to address two interrelated questions, namely which traits of Red Sea fish species are associated with the species' introduction into the Mediterranean Sea and how the introduction patterns changed over time.

## Location

### Red Sea and Mediterranean Sea

**Methods** We used Cox regressions to identify traits of Red Sea species that are associated with the rate at which new alien Red Sea species are recorded in the Mediterranean (hazard rate) and to identify groups of species with different temporal trends of their baseline hazard rates. We fitted latent-variable models to determine whether different trends in baseline hazard rates can be attributed to trends in detection or introduction.

**Results** Our results showed that the highest hazard rate occurred among pelagic species, species living over soft bottoms, species present in small families and species recorded in the Red Sea close to the Suez Canal. We also found that alien species could be separated into three groups with different temporal trends in baseline hazard rates. The different trends were due to changes in introduction rate rather than detection. Two groups with historically low introduction rate showed an increase in introduction rate over time.

**Main conclusions** Our results provide novel biogeographical explanations for introduction patterns that were previously attributed to the effects of sea surface temperature and interspecific competition. The trends uncovered by our analysis indicate that the profile of introduced species is changing with potentially profound consequences for the Mediterranean Sea.

## Keywords

climate change, Cox regression, fish, introduction rate, invasion, Lessepsian migration, Mediterranean Sea

\*Correspondence: Heinrich zu Dohna,  
Department of Biology, American University of  
Beirut, 11–0236, Beirut, Lebanon.  
E-mail: hb54@aub.edu.lb

## INTRODUCTION

The Mediterranean Sea is considered a marine biodiversity hotspot and one of the most invaded marine regions in the world (Coll *et al.*, 2010; Costello *et al.*, 2010; Edelist *et al.*, 2013; Katsanevakis *et al.*, 2013). Due to its enclosed geography and long history of human settlement on its shores, the Mediterranean Sea is exposed to various anthropogenic threats. One such threat is the introduction of alien species

from the Red Sea through the Suez Canal, a phenomenon that is referred to as 'Lessepsian migration'. By the year 2013, 90 exotic fish species of Indo-Pacific origin had been recorded in the Mediterranean since the construction of the Suez Canal (Golani *et al.*, 2013). There is evidence that some Lessepsian fishes have displaced native Mediterranean species (e.g. Bariche *et al.*, 2004; Edelist *et al.*, 2013; Galil *et al.*, 2014) and have fundamentally altered ecosystems in the Mediterranean (Sala *et al.*, 2011; Vergés *et al.*, 2014). In light

of the current and potential future impact of Lessepsian migrants, an accurate assessment of the risks posed by the Suez Canal is a pressing task for marine ecologists.

Studies addressing the risk posed by alien species aimed at predicting either which species are likely to cross from the Red Sea into the Mediterranean (Belmaker *et al.*, 2013) or which species can become established after they crossed into the Mediterranean (Arndt & Schembri, 2015). Belmaker *et al.* (2013) analysed which traits among reef-associated fishes from the Indo-Pacific are significant predictors for the probability of introduction into the Mediterranean. Their analysis confirmed previous observations [e.g. that species forming large schooling groups have higher introduction probabilities (Golani, 2010)] and produced new results (e.g. that species occurring at higher maximum temperatures in the Indo-Pacific have higher introduction probabilities). Based on the latter result, the authors concluded that high sea surface temperature in the Mediterranean constitutes a barrier for introduction of species occurring at lower temperatures in the Indo-Pacific and that global warming would, therefore, be unlikely to accelerate species' introductions from the Red Sea into the Mediterranean.

The detection of traits associated with introduction probabilities provides important information about ecological mechanisms and future risks of such introductions. However, there are several challenges for the detection of these traits. One challenge is that there are numerous correlated traits that can be associated with introduction risk and any particular one could be a strong or poor predictor of introduction risk, depending on other traits included in the analysis. Another challenge is the ecological and geographical definition of the source pool, i.e. the set of species that could potentially become introduced. For example, if the geographical delineation of the source pool is too wide, statistically significant predictors of introduction could simply indicate traits associated with an occurrence close to the Suez Canal, rather than traits associated with the introduction itself. A third challenge is that the introduction risk could change differently over time for different species. In that case, a simple introduction probability, as used in earlier studies, does not adequately represent future introduction risk.

The Suez Canal has been widened and deepened multiple times since its initial construction (Rilov & Galil, 2009; Galil *et al.*, 2014). It has been expanded further in 2015 by the digging of a new 35-km-long shipping lane ([www.suezcanal.gov.eg](http://www.suezcanal.gov.eg), last accessed 2 March 2016). To assess the risk posed by this expansion, it is important to know whether previous expansions led to an increase in the invasion rate. A key component in answering this question is to determine whether there is evidence that the invasion rate increased over time in the past. Several authors have noted an increase in the rate at which species from the Indo-Pacific are recorded in the Mediterranean (Galil, 2008b; Golani, 2010; Raitzos *et al.*, 2010; Edelist *et al.*, 2013). However, an increase in the rate at which alien species are recorded does not necessarily indicate an increase in the introduction rate.

Such an increase could occur if detection of alien species improves over time due to increased surveillance. Furthermore, even if the introduction rate and the per-species detection probability stay constant, the rate at which new alien species are recorded can increase over time. This is because alien species are usually not detected immediately upon introduction and unrecorded species can accumulate and contribute to a rise in the detection rate (Costello & Solow, 2003; Solow & Costello, 2004). In fact, a latent-variable model that accounted for the possibility of undetected introductions produced no evidence for an increase in the rate at which Indo-Pacific species are introduced into the Mediterranean Sea (Belmaker *et al.*, 2009).

The purpose of this study is to combine an analysis of temporal trends in introduction risk with an analysis of species traits associated with introduction, while addressing some of the challenges of detecting those traits. Based on careful *a priori* considerations, we delineated a source pool that included only potential migrants and compiled a large set of predictor variables for the source pool species. We employed a Cox-proportional hazard model to identify predictors of introduction risk. We tested the proportionality assumption to reveal whether there are groups of species with different temporal trends in introduction risk. We fitted latent-variable models to determine whether trends in introduction rates were due to changing introduction or detection rates.

## MATERIALS AND METHODS

### Species included in the analysis

Our analysis considered all bony fish species that were reported from the Red Sea by Golani & Bogorodsky (2010). Based on species accounts present on [www.fishbase.org](http://www.fishbase.org) (Froese & Pauly, 2013; last accessed 30 September 2013), we updated synonyms and excluded from the list of Golani & Bogorodsky (2010) species that are also native to the Mediterranean Sea, occur in deep water (> 100 m upper range depth) or are globally distributed (see Table S1.1 in Supporting Information, Fig. 1a). The resulting list contained a total of 960 fish species. Since any of these species could theoretically move to the Mediterranean via the Suez Canal, they will be referred to as the 'source pool' (see Table S1.1). We used the online CIESM Atlas of Exotic Species in the Mediterranean (Golani *et al.*, 2013; last accessed 5 May 2014) to determine which source pool species have at least one record in the Mediterranean. We did not consider records of *Acanthopagrus bifasciatus*, *Terapon theraps* and the first record of *Pomadasys stridens* as the locations of these records in the Mediterranean clearly suggest a presence not related to a passage through the Suez Canal (Lipej *et al.*, 2008; Golani *et al.*, 2013; Ben Souissi *et al.*, 2014). Fishes that were recorded in the CIESM's list (Golani *et al.*, 2013) but were absent from the source pool (Golani & Bogorodsky, 2010) were not included in the analysis to ensure that the definition of the source pool (species that could pass through

the Suez Canal) was independent of the measured outcome (occurrence in the Mediterranean). We identified 78 species from the source pool that have records in the Mediterranean (Golani *et al.*, 2013). They will be referred to as 'introduced species' (see Table S1.2).

### Predictor variables included in the dataset

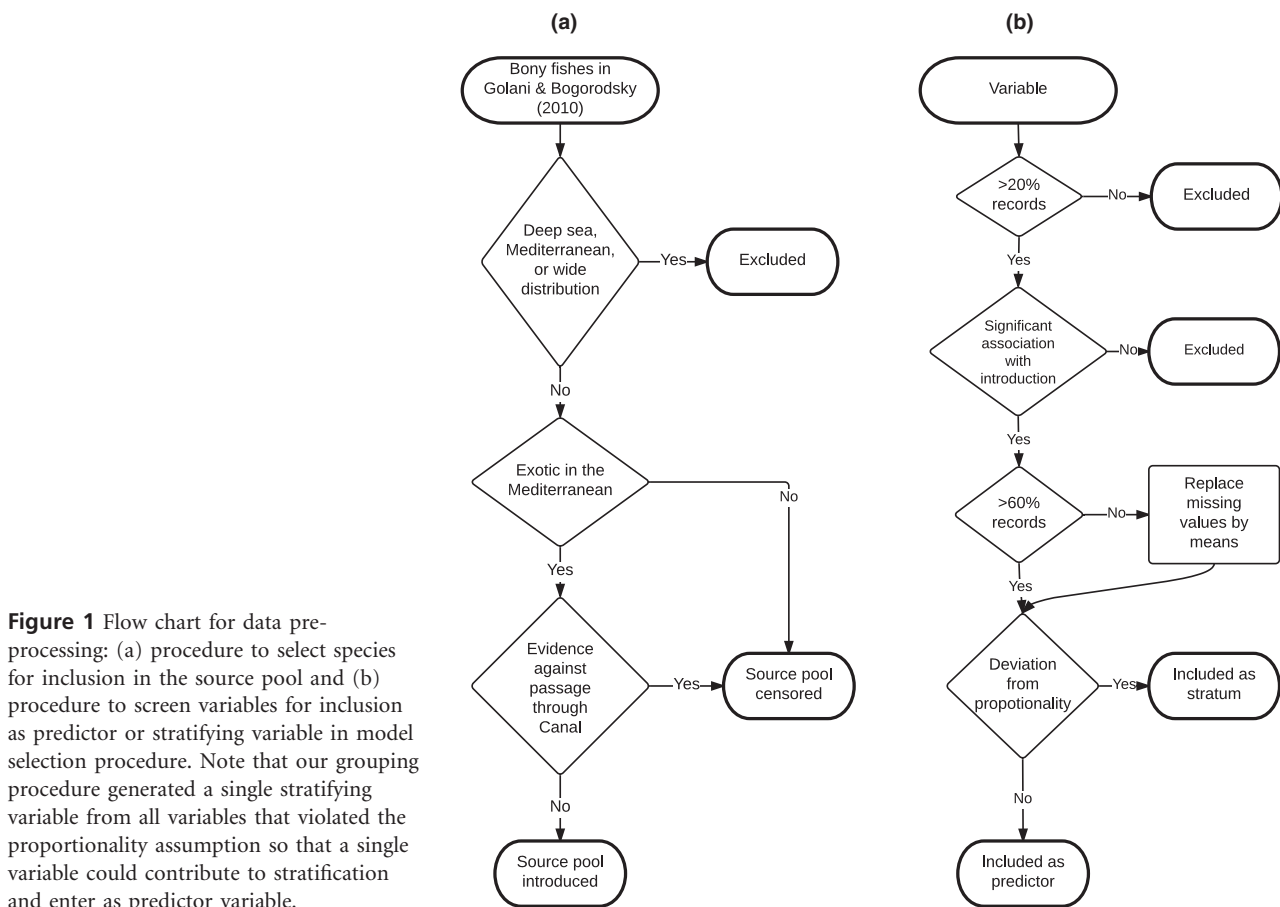
Biological information for all species within the source pool was compiled from a dataset acquired from [www.fishbase.org](http://www.fishbase.org) (Froese & Pauly, 2013). For each species, the dataset contained fields with information on taxonomy, ecology, behaviour, life history, morphology and an additional field with comments. We extracted from these comments further information and incorporated it manually into the main fields. The number of species present in the Red Sea within each family was included as a separate field. Additional information was obtained from the literature for six variables with missing records. They included information on distribution, longevity, length, activity, type of bottom and food items. We retained for the study only fields with records available for at least 20% of the species in the source pool (Fig. 1b).

In addition, we downloaded from [www.fishbase.org](http://www.fishbase.org) (Froese & Pauly, 2013) all available sampling location data in the Red Sea and Indian Ocean (Fig. 2). We obtained the data from the fishbase website by selecting 'Indian Ocean' in the

section 'Information by Ecosystem' and by selecting 'Point data' (downloaded on 11 April 2015). These data contained the latitude, longitude and fish species observed per location for 1191 locations. These sampling data were not exhaustive, i.e. some source pool species were not recorded in any sampling locations from the Red Sea and some species were not recorded in any sampling locations from the entire Indian Ocean. We obtained for each of these sampling locations the yearly range and monthly maximum sea surface temperature (*SSTrange* and *SSTmax*, respectively) from the Bio-ORACLE database (Tyberghein *et al.*, 2012, downloaded on 24 April 2015). We calculated for each species from the source pool, with at least one record in the location data, the mean *SSTrange* and maximum *SSTmax* among all its sampling locations and included the mean *SSTrange* and maximum *SSTmax* as predictor variables. Furthermore, we determined for each species the sampling location that is closest to the Suez Canal and included the shortest 'great-circle distance' from the closest location to the canal as predictor variable (variable 'minimum distance to canal'). The resulting dataset had a total of 33 predictor variables (see Table S1.3).

### Cox regressions

A Cox-proportional hazard model (Cox, 1972) was used to estimate the effects of predictor variables on the rate at



**Figure 1** Flow chart for data pre-processing: (a) procedure to select species for inclusion in the source pool and (b) procedure to screen variables for inclusion as predictor or stratifying variable in model selection procedure. Note that our grouping procedure generated a single stratifying variable from all variables that violated the proportionality assumption so that a single variable could contribute to stratification and enter as predictor variable.

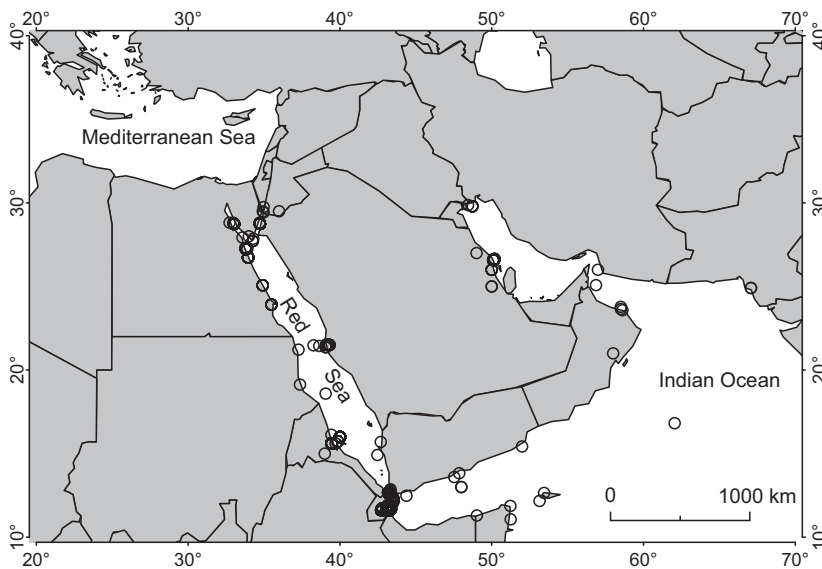
which new introduced species are recorded in the Mediterranean Sea. Following standard survival analysis terminology, we will refer to this rate as the ‘hazard rate’. The events described by this hazard rate are the first recordings of new alien species from the source pool. It is important to note that this hazard rate is not the same as the introduction rate, as it is influenced by the rate at which species are introduced and the rate at which introduced species are detected. We will use the term ‘true introduction rate’ when we refer to the rate at which species are introduced. Our analysis considered the year of first record in the Mediterranean as an event time. All source pool species with no record of introduction were treated as right-censored observations with the year 2013 as censoring time (Fig. 1a). The year 2013 was chosen as it coincided with the last update of the CIESM online database at the time of our analysis (Golani *et al.*, 2013). The number of years was counted since the opening of the Suez Canal in 1869 (i.e. year 2013 was treated as year 144). We clustered species by family in each Cox regression to account for phylogenetic non-independence of different fish species.

The analysis contained three parts, the screening of potential predictor variables, the selection of the best combination of predictor variables that passed the screening, and an analysis of the temporal change in the hazard rate and true introduction rate. The screening itself was carried out in several steps (Fig. 1b). We tested each of the 33 variables individually for an association with the hazard rate using a Cox regression model and a likelihood ratio test for significant model fit. We adjusted the *P*-values according to the adaptive Benjamini–Hochberg correction (Benjamini & Hochberg, 2000) to keep the false discovery rate among the screened variables below 5%. We tested all variables that passed the screening for a violation of the proportionality assumption. The proportionality assumption states that the proportional differences in hazard rates between subjects with different covariates stay constant over time. Screened

variables entered the model selection procedure as stratifying variables if they violated the proportionality assumption and as predictor variables otherwise. When a screened predictor variable contained missing values for more than 40% of the source pool species, the missing values for this variable were replaced by its mean value among all species that contained a record for this variable. We performed a Cox regression for all possible combinations of screened predictor variables and selected the combination with the lowest Akaike’s information criterion (AIC) value as the best model (Burnham & Anderson, 2002). We tested the goodness-of-fit of the best model following the method by Parzen & Lipsitz (1999) using eight rate classes. We calculated the correlations between all variables that passed the screening but did not test whether correlations differed significantly from zero as different species could not be considered independent observations. In addition to the standard Cox regression, we fitted a Cox regression with a ridge penalty (R package ‘penalized’; Goeman, 2010) using all predictor variables of the best model to analyse the robustness of the regression coefficients.

### Temporal change in the introduction rate

To analyse the temporal change in the hazard rate, we assigned species to groups with similar baseline hazard rates. First, we tested all variables that passed the screening for a deviation from the proportionality assumption on a global level (Grambsch & Therneau, 1994). Then, we applied the following grouping procedure to each variable that showed a significant deviation: we tested for the deviation from proportionality among all pairs of factor levels within a variable (Grambsch & Therneau, 1994) and pooled pairs of factor levels whose baseline hazard rate did not deviate significantly from proportionality, starting with the pair whose comparison produced the largest *P*-value. After each pooling, we recalculated the *P*-values for the comparisons of the



**Figure 2** Study area with point locations. The map is based on a rectangular projection with equivalent longitude and latitude scales at the centre of the map. Our analysis of point locations included all locations from the Red Sea and Indian Ocean that contained source pool species. The map shows, therefore, only a subset of the point locations used in the analysis.

new groups. The procedure stopped once there was a significant deviation from proportionality between all pairs of groups. After we applied this procedure to all predictor variables that showed a deviation from proportionality, we combined the newly created groups from different variables and created a single new variable with a factor level for each group combination. We repeated the grouping procedure for this newly created variable and thereby obtained a single grouping variable that defined groups with different baseline hazard rates.

We fitted different versions of the model developed by Solow & Costello (2004) to determine the most likely reason for different baseline hazard rates. Different species could have different baseline hazard rates because they experienced a different change in their true introduction rate, a different change in their detection rate or a different change in both. According to the model by Solow & Costello (2004), the true introduction rate  $\mu_t$  changes with time  $t$  according to

$$\mu_t = e^{\beta_0 + \beta_1 t}$$

The introductions described by this rate are not observed directly. The probability of a species becoming detected in year  $t$  given that it was introduced in year  $s$  is given by the following:

$$\pi_{st} = \frac{\exp(\gamma_0 + \gamma_1 t + \gamma_2 e^{t-s})}{1 + \exp(\gamma_0 + \gamma_1 t + \gamma_2 e^{t-s})}$$

The parameter  $\beta_1$  estimates how the true introduction rate changed with time, the parameter  $\gamma_1$  estimates how the detection probability changed with time and the parameter  $\gamma_2$  estimates how the detection probability changed with time since introduction.

We fitted 16 models of different complexity and used the AIC value to rank them. The trend parameters for introduction ( $\beta_1$ ) or detection ( $\gamma_1$  and  $\gamma_2$ ) were either constrained to zero or unconstrained, leading to four possible combinations of trend parameter constraints. Similarly, the parameters of detection and introduction could either be the same for all groups with different baseline hazard rates or be different for each group, leading to four possible combinations of how parameters differed between groups. Together these two sets of four combinations led to a total of 16 different models.

We estimated the maximum likelihood values for parameters for all 16 models through numerical optimization procedures. We used zero as starting values for the parameters  $\beta_1$ ,  $\gamma_0$ ,  $\gamma_1$  and  $\gamma_2$  and  $\log(N/T)$  for  $\beta_0$ , where  $N$  denotes the total number of observed introduced species and  $T$  the total number of years. The starting value for  $\beta_0$  was chosen so that  $\exp(\beta_0) = N/T$ , i.e. according to the starting values there is no temporal trend in introduction and the yearly introduction rate equals the total number of introduced species divided by the length of the observation period. The observation period started with the first Mediterranean record in 1902 and ended in 2013, the last year CIESM was updated at the time of our analysis. We calculated the AIC values for all models

to rank them. In addition, we performed likelihood ratio tests to determine whether individual trend parameters of the highest ranked model differed from zero. All calculations were performed using the statistical programming language R 3.2.1 (R Core Team, 2013).

## RESULTS

### Variable screening

Of 33 screened traits of fishes living in the Red Sea, six were significantly associated with introductions, when controlling for a false discovery rate of 5% (see Fig. S2.1). These were schooling, demersal-pelagic distinction, type of bottom, number of species per family, minimum distance to canal and *SSTmax* (see Table S1.3). The baseline hazard rates of species associated with different categories of the variables 'schooling' and 'type of bottom' deviated significantly from the proportionality assumption. Our grouping procedure separated species into three groups with distinct baseline hazard rates. In the subsequent analyses, the Cox-proportional hazard model was stratified according to these three groups with separate baseline introduction rates. The variables 'schooling', *SSTmax* and 'minimum distance to canal' contained each missing records for more than 40% of the species. To include these variables in the model selection procedure, we replaced missing values by means. For all species whose schooling value was missing but available for at least one member of the same family, we replaced the missing value by the mean schooling value of all species of the same families. For all other species, we replaced the missing 'schooling' values by the overall mean. We replaced the missing values for *SSTmax* and 'minimum distance to canal' by their overall means among all other species. To evaluate the effects of these replacements, we ran single-variable Cox regressions before and after replacing the missing values. The replacement of missing values by means changed the estimated regression coefficients by 12–21% of their pre-replacement values but did not affect the signs of the estimated coefficients or the significance levels of the  $P$ -values.

### Selection of the best model

The best combination of predictor variables according to the AIC excluded *SSTmax* and schooling and contained thus four of the six traits that were associated with introductions (Table 1). According to the best model, the hazard rate was higher among species living close to the surface, among species living over soft bottoms, among species found within less diverse families and among species with a low minimum distance to the canal (Table 1, Fig. 3). The goodness-of-fit test showed no significant deviation between the predicted and observed hazard rates. There were moderate correlations among the predictor variables (Table 2). However, a penalized ridge regression analysis showed that, despite these

correlations, the sign and magnitude of the estimated coefficients were robust (see Fig. S2.2).

### Change in introduction rate

Our grouping procedure separated species into three groups with distinct baseline hazard rates: group I included all species that had the type of bottom category 'none' and any schooling level (i.e. all pelagic species); group II included all species with the type of bottom category 'hard' and schooling level 1 (i.e. solitary or species living in pairs on or close to hard bottoms); and group III included all remaining species (i.e. all species living on or close to soft bottoms or alternatively on both types of bottoms and non-solitary species living on or close to hard bottoms). The best fitting model for the change in introduction rate contained a constant detection probability ( $\gamma_1 = \gamma_2 = 0$ ) that was the same for each group and a true introduction rate that changed differently over time for each group (Table 3). The maximum likelihood parameter estimates for this model were  $\gamma_0 = 11.5$  for all groups,  $\beta_0 = -2.0$ ,  $\beta_1 = 0.004$  for group I,  $\beta_0 = -14$ ,  $\beta_1 = 0.13$  for group II and  $\beta_0 = -2.3$ ,  $\beta_1 = 0.02$  for group III. The value for  $\gamma_0$  implies a yearly detection probability of > 99.9%. The difference in AIC value to any model with constant true introduction rate is > 10, providing strong evidence for a temporal increase in the true introduction rate (Table 3). The trend parameter  $\beta_1$  differed significantly from zero for groups II and III (for both groups  $P < 10^{-5}$ , likelihood ratio test) but not for group I ( $P = 0.52$ , likelihood ratio test). Hence, there is evidence that the introduction rate increased over time for species of groups II and III (i.e. all demersal species) but not for group I (all pelagic species) (Fig. 4). Furthermore,  $\beta_1$  was more than six times higher for group II (solitary or species living in pairs on or close to hard bottoms) than for group III (other demersal species),

**Table 1** Estimated coefficients and *P*-values of predictor variables that are part of the best model explaining the introduction rate of fish from the Red Sea to the Mediterranean Sea.

Variable	Encoding	Estimated coefficient	<i>P</i>
Number of species per family	Integer	-0.01	0.039
Demersal – pelagic*	1 = surface 2 = midwater 3 = close to bottom 4 = on the bottom	-0.98	0.0003
Type of bottom	Hard Both Soft None (pelagic)	-1.1 – 0.46 –	0.001 – 0.17 –
Minimum distance to the canal	Real number	-0.0002	0.0007

\*Species belonging to multiple categories were assigned the mean value of these categories.

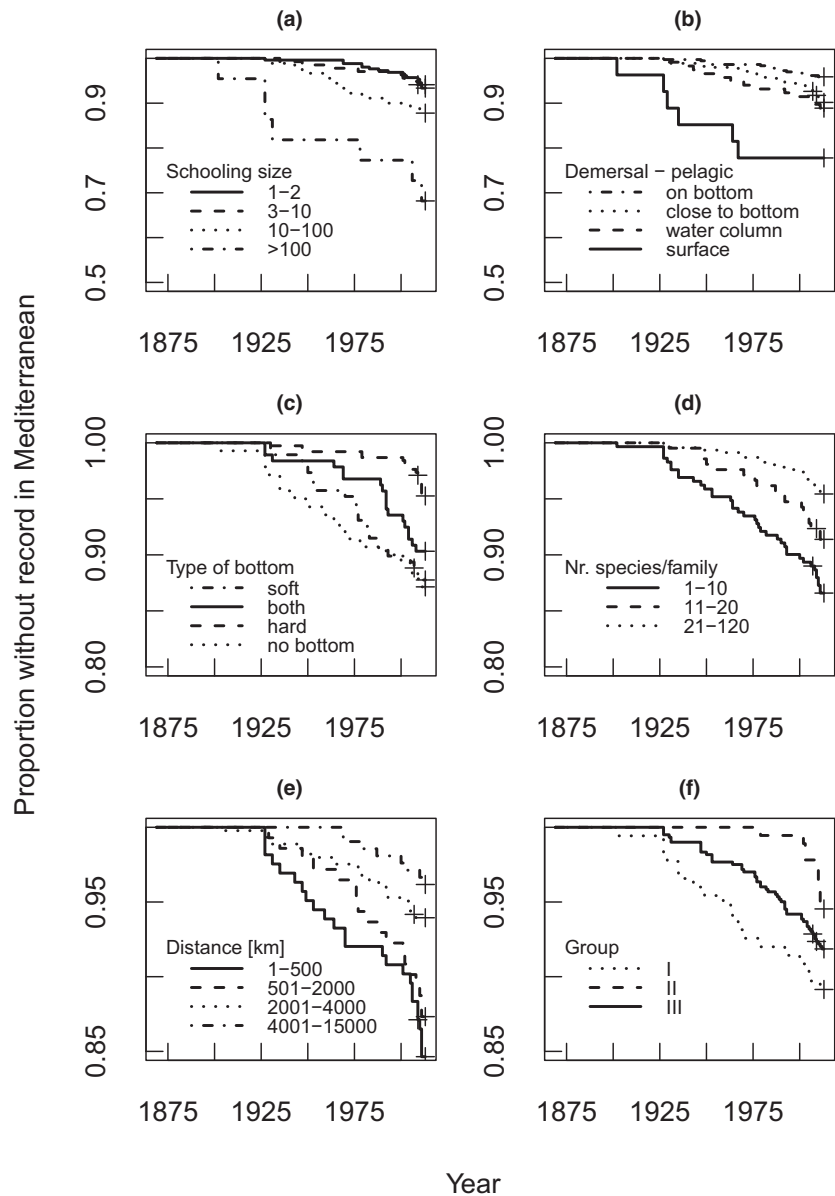
indicating a strong recent surge in the introduction rate of group II species (Fig. 4).

### DISCUSSION

We performed Cox regressions to determine which attributes of fishes present in the Red Sea are associated with the rate at which new alien species from the Red Sea are recorded in the Mediterranean (hazard rate). We also analysed whether the true introduction rate changed over time. Our results showed that the highest hazard rate occurred among pelagic species, species living over soft bottoms, species present in small families and species recorded close to the Suez Canal. We also found that alien species could be separated into three groups, each displaying a different temporal trend in their true introduction rate. Two groups showed an increase in true introduction rate over time. The group that showed the strongest increase in introduction rate was solitary fishes living on hard bottoms.

Our analysis has several limitations. The dataset we analysed contained many missing records. We removed 27 variables because these variables had records for 20% of the species or fewer. Some of these removed variables could have an influence on the introduction rate. We, furthermore, replaced missing values by mean values. Even though we do not have evidence that this replacement affected our conclusions, more complete data would undoubtedly improve the estimated effects of the variables on the introduction rate. Another area of possible improvement is our method of accounting for phylogenetic dependence. We accounted for phylogenetic dependence by introducing random family effects. A similar method has been used in a previous study (Belmaker *et al.*, 2013). There are methods to account for phylogenetic effects more thoroughly in regression models (e.g. Ives & Garland, 2010). However, we are not aware of such methods in the context of Cox regressions and applying these methods would require sequence data for all species in the source pool.

Our analysis – and previous studies, e.g. Belmaker *et al.* (2013) – assumed that the introductions of different species were independent of each other. This assumption can be justified as we used the date of first record in the Mediterranean as event time. Hence, our analysis focused on early stages of the introduction process, a stage that is less likely to be influenced by the presence of other species (Azzurro *et al.*, 2014). There is no overlap between the significant predictors that we identified in our study and the significant predictors of a species' establishment success after it had crossed into the Mediterranean (Arndt & Schembri, 2015). This is despite the fact that all significant predictors used in Arndt & Schembri (2015) were also used in our study. This lack of overlap suggests that the different stages of the introduction process pose different ecological challenges. Hence, the predictors identified in our study are more related to the species ability to pass through the Suez Canal than to its establishment success.



**Figure 3** Proportion of source pool (Red Sea) fish species with no record from the Mediterranean Sea since the opening of the Suez Canal. Species were grouped by (a) schooling size, (b) habitat in relation to the bottom, (c) type of bottom, (d) number of species within each family, (e) closest distance to Suez Canal among records from sampling locations in the Indian Ocean and (f) groups with different baseline hazard rate.

A novel result of our study is that species belonging to small families were more prone to be introduced into the Mediterranean than species from large families. There are several possible reasons for this. Theoretical models of sympatric speciation (Doebeli & Dieckmann, 1999) and empirical analyses of herbivorous insects (Hardy & Otto, 2014) have shown that ecological specialization tends to be associated with higher diversification rates. Hence, species from highly diverse families might be more specialized and therefore less likely to colonize new habitats. Another possible explanation for the association between small family size and higher hazard rate could be that allopatric speciation creates a negative correlation between dispersal ability and family size. Members of families of good dispersers might have undergone allopatric speciation less often than members of families of poor dispersers, and as a result, families of good

dispersers are less diverse than families of poor dispersers. Lessepsian migrants might predominantly come from families of good dispersers and therefore from smaller families. The correlations between predictor variables indicate that species from small families tend to occur away from the bottom (surface or midwater) and in large schools (Table 3). Hence, two traits that are associated with higher hazard rates are overrepresented in small families, suggesting an association between good dispersal and low within-family diversity. A full analysis of this association would require a phylogeny-based estimation of the influence of these two traits on diversification rate (FitzJohn *et al.*, 2009), which is beyond the scope of this study. Regardless of the reason for the association between intra-family diversity and introduction risk, the fact that species from small families are more likely to pass into the Mediterranean provides a new explanation for

	Minimum distance to canal	Number of species per family	Demersal – pelagic	Schooling	Type of bottom	<i>SSTmax</i>
Minimum distance to Canal	–	0.02	0.07	–0.11	0	–0.19
Number of species per family	–	–	0.17	–0.31	0.27	–0.06
Demersal – pelagic	–	–	–	–0.57	–0.24	–0.11
Schooling	–	–	–	–	0	0.11
Type of bottom	–	–	–	–	–	–0.06
<i>SSTmax</i>	–	–	–	–	–	–

**Table 3** Results of the five best models fitted to times at which fish species introduced from the Red Sea are first detected in the Mediterranean Sea.

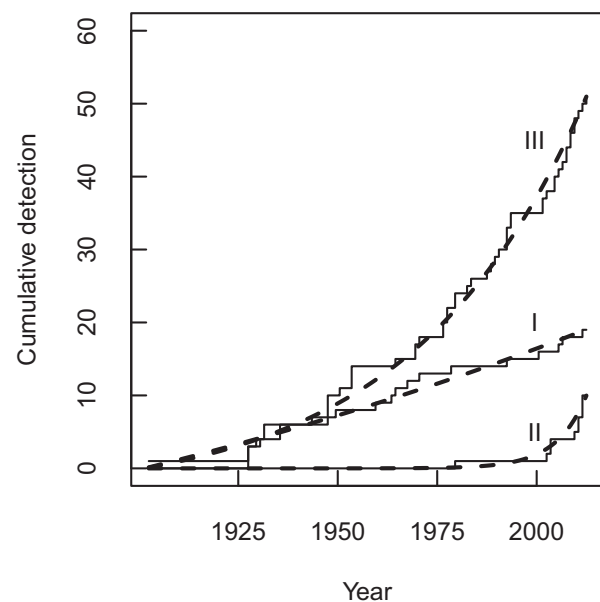
Parameters constrained to zero	Parameters shared between groups	Log-likelihood	Number parameters	AIC	$\Delta$ AIC
$\gamma_1$ and $\gamma_2$	$\gamma_0, \gamma_1$ and $\gamma_2$	–148.5	7	311	0
$\gamma_1$ and $\gamma_2$	None	–147.5	9	313	2
None	$\gamma_0, \gamma_1$ and $\gamma_2$	–148.5	9	315	4
None	None	–144.1	15	318.1	7.1
$\beta_1$	None	–148.7	12	321.3	10.3

Belmaker *et al.*'s (2013) observation that Lessepsian migrants show a higher trait evenness and divergence than randomly selected Red Sea species. The high trait evenness and divergence of Lessepsian migrants has previously been interpreted as evidence that migrants are filtered by competition (Belmaker *et al.*, 2013; Azzurro *et al.*, 2014). We provide the alternative explanation that Lessepsian migrants represent more families than a random sample from the source pool.

Our analysis confirms the previously discovered associations between the hazard rate and the variables *SSTrange*, *SSTmax*, type of bottom, schooling level and a species' occurrence in the water column (Golani, 1993, 2010; Belmaker *et al.*, 2013). The high hazard rate among soft-bottom fishes detected in our study matches the previous observation that Lessepsian species make up a large percentage among species that are common in sandy habitats in the northern Red Sea (Golani, 1993).

Our results regarding the associations between the hazard rate and the variables *SSTrange* and *SSTmax* match the results of a previous study (Belmaker *et al.*, 2013), but our results also provide new interpretations for these associations. The association between hazard rate and *SSTrange* was not strong enough to overcome our false discovery rate

**Table 2** Correlation estimates among the six traits (predictor variables) associated with the rate at which fish species are introduced from the Red Sea to the Mediterranean Sea.



**Figure 4** Observed (solid step lines) and fitted (dashed lines) cumulative number of detected fish species introduced from the Red Sea to the Mediterranean for the three groups that have a different baseline hazard rate according to the proportionality test. The labels next to the lines indicate the groups (I–III). See text for description of groups. The fitted line corresponds to the highest ranked model (constant detection shared by all groups and increase in introduction rates).

threshold. This could be because our set of sampling locations might differ from the ones used by Belmaker *et al.* (2013) or because we used a large number of variables and controlled for a false discovery rate. The association between high *SSTmax* and high hazard rate was strong enough to pass the screening, but the variable *SSTmax* was not part of the best model when other correlated variables were included. The variable *SSTmax* showed the strongest correlation with the variable 'distance to the canal' (Table 2), which suggests a novel explanation of the effect of *SSTmax*. Since

*SSTmax* is higher in the Red Sea than in most parts of the Indo-Pacific (see Fig. S2.3), species occurring at sampling locations in the Red Sea tend to have a higher *SSTmax*. Therefore, *SSTmax* might simply be a proxy for a species' occurrence in the Red Sea sampling locations but itself not a predictor of the introduction probability. This interpretation of the effect of *SSTmax* is supported by two additional pieces of evidence. *SSTmax* is generally lower in the eastern Mediterranean than in the Red Sea (see Fig. S2.3), and its value in the native range does not predict the establishment success in the Mediterranean (Arndt & Schembri, 2015). In contrast, Belmaker *et al.* (2013) interpreted the association between *SSTmax* and introduction probability as evidence that high maximum temperatures in parts of the eastern Mediterranean restrict the introduction of species that occur at lower temperatures in the Indo-Pacific. According to their interpretation, global warming will slow down the invasion process from the Red Sea into the Mediterranean. Our interpretation of the effect of *SSTmax* does not suggest such an effect of global warming. More detailed data on the geographical distribution and thermal tolerance of Red Sea species are needed to disentangle the effects of *SSTmax* and geographical distance to the canal and to determine whether global warming will slow down Lessepsian migration as suggested by Belmaker *et al.* (2013) or facilitate it as suggested earlier by other researchers (Lasram & Mouillot, 2009; Raitos *et al.*, 2010).

Our results show that the average hazard rate is higher among species living close to the surface than among species living midwater or close to the bottom (Fig. 3b), which is in accordance with Belmaker *et al.* (2013). However, our analysis of introduction rate trends also shows that demersal species, while having a lower average introduction rate than pelagic species (Fig. 3b), made up the majority of the alien species that were recorded in recent time (groups II and III in Fig. 4). This is partly because there are more demersal than pelagic species in the source pool and partly because the introduction rate remained roughly constant over the last century for pelagic species (group I) but increased over time among demersal species (groups II and III) (Figs 2f & 3). As a result, the current introduction rate is highest among demersal species. The current increase in introduction rate is highest among solitary hard-bottom species (group II) as 9 out of the 10 introduced solitary hard-bottom species have been recorded in the Mediterranean since 2002 (Fig. 4).

The increase in true introduction rate among species of groups II and III is at odds with Belmaker *et al.* (2009), who analysed temporal trends in true introduction rates of Lessepsian species and found no evidence for a temporal increase in the true introduction rate. Belmaker *et al.* (2009) used a different analysis method and a different dataset than our study and did not separate species into groups with different baseline hazard rates. To determine whether their result deviated from ours due to a difference in data or methodology, we repeated Belmaker *et al.*'s analysis on their dataset but introduced a distinction between

demersal and pelagic species. When applied separately to demersal and pelagic species, the method of Belmaker *et al.* provided statistical evidence for an increase in the true introduction rate in demersal species but not for pelagic species (results not shown). This indicates that Belmaker *et al.* (2009) did not detect increasing true introduction rates because they did not separate species according to different baseline hazard rates and not because they used different data or methodology.

According to the highest ranked model in our study, the increase in the hazard rate is due to an increase in the true introduction rate rather than due to an improved detection. However, it is unclear to what degree our result is a consequence of the particular functions we used. The temporal change in records of introduced species is well approximated by the exponential function we used for the true introduction rate, and as a result, the highest ranked model assumes an immediate detection of introduced species – an assumption that is certainly not true. The highest ranked model is, therefore, not necessarily an accurate description of the underlying processes but rather a benchmark to measure alternative models against. The highest ranked model performed substantially better than any model we fitted with constant true introduction rates. Hence, the models we fitted provide strong evidence for an increase in the true introduction rate among demersal species.

Our estimation of temporal trends in introduction rates did not consider the effects of source pool depletion. The true introduction rate in our analysis is the total number of species introduced per time, i.e. the number of species in the source pool multiplied by the per-species introduction probability. Since we considered first records, a species ceases to be part of the source pool once it has been recorded in the Mediterranean, and hence, each introduction event reduces the introduction rate. Therefore, the per-species introduction probabilities must have increased even more than the increase in introduction rate shown in our analysis.

The ecological reason for the increase in the true introduction rate of demersal species (group III) over time is unclear. This increase could have been due to repeated expansions of the Suez Canal, the change in its salinity over time, or other unknown environmental changes within the canal. It could also be due to an increase in the susceptibility of the Mediterranean to the establishment of new species, as the invasibility of an environment can be influenced by an assortment of biotic and abiotic events and processes (Davis, 2011).

The detected sharp increase in the last decade of solitary species living mainly on hard bottom (group II, Fig. 4) might be results of aquarium release events. Species that are potentially used in aquariums constitute 80% of the species in group II (i.e. 8/10) and only around 18% of the species in groups I and III (i.e. 12/68). It is, therefore, conceivable that the recent surge in the detection of group II species is due to the increase in the trade of tropical marine species in the Middle East, rather than to environmental changes. Genetic

studies that can better locate the geographical source of newly introduced alien species are likely to shed more light on this hypothesis (Bariche *et al.*, 2015).

Regardless of the underlying reasons, the recent increase in the true introduction rate of demersal species should be a cause of concern. Introduced demersal species have the potential to drastically affect sensitive Mediterranean coastal ecosystems (Sala *et al.*, 2011). Our analysis treated the first Mediterranean record as event time and therefore focused on the early stage of the introduction process. The date of collection may lag years behind the date of introduction, and it could take years for an introduced species to become established (Galil, 2008a). Therefore, we assume that the full ecological cost of the recent surge in introductions of demersal species is not yet apparent. Furthermore, the concentration of recent records of newly introduced species with historically low introduction rates indicates a shift in the nature of species introductions into the Mediterranean. It is important to understand the reason for this shift, particularly in light of the further expansion of the Suez Canal. If we do not understand and mitigate the ecological risks associated with the expansion of the Suez Canal, the integrity of a large part of the Mediterranean ecosystem could be in jeopardy.

## ACKNOWLEDGEMENT

We thank Ms Cristina Garilao for supplying data from the Fishbase database in a table format.

## REFERENCES

- Arndt, E. & Schembri, P.J. (2015) Common traits associated with establishment and spread of Lessepsian fishes in the Mediterranean Sea. *Marine Biology*, **116**, 2141–2153.
- Azzurro, E., Tuset, V.M., Lombarte, A., Maynou, F., Simberloff, D., Rodríguez-Pérez, A. & Solé, R.V. (2014) External morphology explains the success of biological invasions. *Ecology Letters*, **17**, 1455–1463.
- Bariche, M., Letourneur, Y. & Harmelin-Vivien, M. (2004) Temporal fluctuations and settlement patterns of native and Lessepsian herbivorous fishes on the Lebanese coast (eastern Mediterranean). *Environmental Biology of Fishes*, **70**, 81–90.
- Bariche, M., Torres, M., Smith, C., Sayar, N., Azzurro, E., Baker, R. & Bernardi, G. (2015) Red Sea fishes in the Mediterranean Sea: a preliminary investigation of a biological invasion using DNA barcoding. *Journal of Biogeography*, **42**, 2363–2373.
- Belmaker, J., Brokovich, E., China, V., Golani, D. & Kiflawi, M. (2009) Estimating the rate of biological introductions: Lessepsian fishes in the Mediterranean. *Ecology*, **90**, 1134–1141.
- 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.
- Ben Souissi, J., Rifi, M., Ghanem, R., Boughedir, W., Capapé, C. & Azzurro, E. (2014) First record of the two bar sea bream *Acanthopagrus bifasciatus* (Teleostei: Sparidae) in the Mediterranean Sea. *Mediterranean Marine Science*, **15**, 437–439.
- Benjamini, Y. & Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, **25**, 60–83.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference*, 2nd edn. Springer-Verlag, New York.
- Coll, M., Piroddi, C., Steenbeek, J. *et al.* (2010) The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PLoS ONE*, **5**, e11842.
- Costello, C.J. & Solow, A.R. (2003) On the pattern of discovery of introduced species. *Proceedings of the National Academy of Sciences USA*, **100**, 3321–3323.
- Costello, M.J., Coll, M., Danovaro, R., Halpin, P. & Miloslavich, P. (2010) A census of marine biodiversity knowledge, resources and future challenges. *PLoS ONE*, **5**, e12110. doi:10.1371/journal.pone.0012110.
- Cox, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- Davis, M.A. (2011) Invasion biology. *Encyclopedia of biological invasions*, 3rd edn (ed. by D. Simberloff and M. Rejmánek), pp. 364–369. University of California Press, Los Angeles, CA.
- Doebeli, M. & Dieckmann, U. (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354–354.
- Edelist, D., Rilov, G., Golani, D., Carlton, J.T. & Spanier, E. (2013) Restructuring the sea: profound shifts in the world's most invaded marine ecosystem. *Diversity and Distributions*, **19**, 69–77.
- FitzJohn, R., Maddison, W. & Otto, S. (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, **58**, 595–611.
- Froese, R. & Pauly, D. (2013) *FishBase*. Available at: <http://www.fishbase.org> (accessed 30 September 2013).
- Galil, B.S. (2008a) Alien species in the Mediterranean sea – which, when, where, why? *Hydrobiologia*, **606**, 105–116.
- Galil, B.S. (2008b) Taking stock: inventory of alien species in the Mediterranean sea. *Biological Invasions*, **11**, 359–372.
- Galil, B.S., Boero, F., Campbell, M.L., Carlton, J.T., Cook, E., Fraschetti, S., Gollasch, S., Hewitt, C.L., Jelmert, A., Macpherson, E., Marchini, A., McKenzie, C., Minchin, D., Occhipinti-Ambrogi, A., Ojaveer, H., Olenin, S., Piraino, S. & Ruiz, G.M. (2014) 'Double trouble': the expansion of the Suez Canal and marine bioinvasions in the Mediterranean Sea. *Biological Invasions*, **17**, 973–976.
- Goeman, J.J. (2010) L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, **52**, 70–84.
- Golani, D. (1993) The sandy shore of the Red Sea-launching pad for Lessepsian (Suez Canal) migrant fish colonizers of the eastern Mediterranean. *Journal of Biogeography*, **20**, 559–585.

- Golani, D. (2010) Colonization of the Mediterranean by Red Sea fishes via the Suez Canal – Lessepsian migration. *Fish invasions of the Mediterranean Sea: change and renewal* (ed. by D. Golani and B. Appelbaum-Golani), pp. 145–188. Pensoft Publishers, Sofia.
- Golani, D. & Bogorodsky, S.V. (2010) The fishes of the Red Sea – reappraisal and updated checklist. *Zootaxa*, **2463**, 1–135.
- Golani, D., Massuti, E., Orsi-Relini, L., Quignard, J.P., Dulcic, J. & Azzurro, E. (2013) *CIESM atlas of exotic fish species*. Available at: <http://www.ciesm.org/atlas/appendix1.html> (accessed 5 May 2014).
- Grambsch, P.M. & Therneau, T.M. (1994) Proportional hazards test and diagnostics based on weighted residuals. *Biometrika*, **81**, 515–526.
- Hardy, N.B. & Otto, S.P. (2014) Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132960. doi:10.1098/rspb.2013.2960.
- Ives, A.R. & Garland, T. (2010) Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, **59**, 9–26.
- Katsanevakis, S., Zenetos, A., Belchior, C. & Cardoso, A.C. (2013) Invading European Seas: assessing pathways of introduction of marine aliens. *Ocean & Coastal Management*, **76**, 64–74.
- Lasram, F.B.R. & Mouillot, D. (2009) Increasing southern invasion enhances congruence between endemic and exotic Mediterranean fish fauna. *Biological Invasions*, **11**, 697–711.
- Lipej, L., Mavrič, B., Žiža, V. & Dulčić, J. (2008) The largescaled terapon *Terapon theraps*: a new Indo-Pacific fish in the Mediterranean Sea. *Journal of Fish Biology*, **73**, 1819–1822.
- Parzen, M. & Lipsitz, S.R. (1999) A global goodness-of-fit statistic for Cox regression models. *Biometrics*, **55**, 580–584.
- R Core Team (2013) *R: A language and environment for statistical computing, version 3.2.1*. R Foundation for Statistical Computing, Vienna, Austria.
- Raitsos, D.E., Beaugrand, G., Georgopoulos, D., Zenetos, A., Pancucci-Papadopoulou, A., Theocharis, A. & Papanthanasios, E. (2010) Global climate change amplifies the entry of tropical species into the eastern Mediterranean sea. *Limnology and Oceanography*, **55**, 1478–1484.
- Rilov, G. & Galil, B.S. (2009) Marine bioinvasions in the Mediterranean sea – history, distribution and ecology. *Biological invasions in marine ecosystems: ecological, management and geographic perspectives* (ed. by G. Rilov and J.A. Crooks), pp. 549–575. Springer-Verlag, Berlin.
- Sala, E., Kizilkaya, Z., Yildirim, D. & Ballesteros, E. (2011) Alien marine fishes deplete algal biomass in the Eastern Mediterranean. *PLoS ONE*, **6**, 1–5.
- Solow, A.R. & Costello, C.J. (2004) Estimating the rate of species introductions from the discovery record. *Ecology*, **85**, 1822–1825.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, **21**, 272–281.
- Vergés, A., Tomas, F., Cebrian, E., Ballesteros, E., Kizilkaya, Z., Dendrinos, P., Karamanlidis, A.A., Spiegel, D. & Sala, E. (2014) Tropical rabbitfish and the deforestation of a warming temperate sea. *Journal of Ecology*, **102**, 1518–1527.

## SUPPORTING INFORMATION

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

**Appendix S1** Tables of species and variables included in the analysis.

**Appendix S2** Supplementary figures.

## BIOSKETCHES

**Carol Samaha** completed her Master's at the American University of Beirut (AUB). Her research interests focus on the ecology and biogeography of marine fish species.

**Heinrich zu Dohna** is assistant professor at AUB with research interests in quantitative methods in ecology and population genetics.

**Michel Bariche** is associate professor at AUB. His research focuses on the biogeography, ecology and taxonomy of fish species.

Author contributions: M.B. and H.z.D. conceived the ideas; C.S. and M.B. collected the data; C. S. and H.z.D. analysed the data; all authors wrote the manuscript.

---

Editor: Luiz Rocha