Invasive genetics of the mummichog (*Fundulus heteroclitus*): recent anthropogenic introduction in Iberia

Teófilo Morim, Grant R Bigg, Pedro M Madeira, Jorge Palma, David D Duvernell, Enric Gisbert, Regina L Cunha, Rita Castilho

1 Center for Marine Sciences, University of Algarve, Faro, Portugal
2 Department of Geography, University of Sheffield, Sheffield, United Kingdom
3 Department of Biological Sciences, Missouri University of Science and Technology, Rolla, Missouri, United States of America
4 RTA, Sant Carles de la Ràpita, Spain

Corresponding Author: Rita Castilho
Email address: rita.castil@gmail.com

Human activities such as trade and transport have increased considerably in the last decades, greatly facilitating the introduction and spread of non-native species at a global level. In the Iberian Peninsula, *Fundulus heteroclitus*, a small euryhaline coastal fish with short dispersal, was found for the first time in the mid-1970s. Since then, *F. heteroclitus* underwent range expansions, colonizing the southern region of Portugal, southwestern coast of Spain and the Ebro Delta in the Mediterranean Sea. Cytochrome *b* sequences were used to elucidate the species invasion pathway in Iberia. Three Iberian locations (Faro, Cádiz and Ebro Delta) and 13 other locations along the native range of *F. heteroclitus* in North America were sampled. Results revealed a single haplotype, common to all invasive populations, which can be traced to the northern region of the species’ native range. We posit that the origin of the founder individuals is between New York and Nova Scotia. Additionally, the lack of genetic structure within Iberia is consistent with a recent invasion scenario and a strong founder effect. We suggest the most probable introduction vector is associated with the aquarium trade and we further discuss the hypothesis of a second human-mediated introduction responsible for the establishment of individuals in the Ebro Delta supported by the absence of adequate muddy habitats linking Cádiz and the Ebro Delta. Although the species has a high tolerance to salinity and temperature, benthic habitat constraints prevent along-shore colonisation suggesting that such expansions would need to be aided by human release.
Invasive genetics of the mummichog (*Fundulus heteroclitus*): recent anthropogenic introduction in Iberia

Teófilo Morim¹, Grant R. Bigg², Pedro M. Madeira¹, Jorge Palma¹, David D. Duvernell³, Enric Gisbert⁴, Regina L. Cunha¹, and Rita Castilho*¹

¹ CCMAR - Centre for Marine Sciences, University of Algarve, Campus de Gambelas, Faro 8005-139, Portugal.
² Department of Geography, University of Sheffield, Winter Street, Sheffield S10 2TN, UK.
³ Department of Biological Sciences, Missouri University of Science and Technology, Rolla, Missouri 65409, USA
⁴ RTA - Sant Carles de la Ràpita, Crta Poble Nou km 5.5, P.O. Box 200, 43540 Sant Carles de la Ràpita, SPAIN

*Corresponding author:
Email: rcastil@ualg.pt
Telephone: +351 918397282

Running head: invasive Iberian mummichog

Keywords: invasive species, introduction vector, cytochrome *b*, mitochondrial DNA
Abstract

Human activities such as trade and transport have increased considerably in the last decades, greatly facilitating the introduction and spread of non-native species at a global level. In the Iberian Peninsula, *Fundulus heteroclitus*, a small euryhaline coastal fish with short dispersal, was found for the first time in the mid-1970s. Since then, *F. heteroclitus* underwent range expansions, colonizing the southern region of Portugal, southwestern coast of Spain and the Ebro Delta in the Mediterranean Sea. Cytochrome *b* sequences were used to elucidate the species invasion pathway in Iberia. Three Iberian locations (Faro, Cádiz and Ebro Delta) and 13 other locations along the native range of *F. heteroclitus* in North America were sampled. Results revealed a single haplotype, common to all invasive populations, which can be traced to the northern region of the species’ native range. We posit that the origin of the founder individuals is between New York and Nova Scotia. Additionally, the lack of genetic structure within Iberia is consistent with a recent invasion scenario and a strong founder effect. We suggest the most probable introduction vector is associated with the aquarium trade and we further discuss the hypothesis of a second human-mediated introduction responsible for the establishment of individuals in the Ebro Delta supported by the absence of adequate muddy habitats linking Cádiz and the Ebro Delta. Although the species has a high tolerance to salinity and temperature, benthic habitat constraints prevent along-shore colonisation suggesting that such expansions would need to be aided by human release.
Introduction

As a consequence of human activities involving large distance marine transportation and trade, worldwide marine biological invasion rates have increased dramatically in the last 30 years (Bax et al. 2003; Perrings et al. 2005; Meyerson and Mooney 2007; Hulme 2009). Along the European coasts, there are over 850 invasive species, of which 237 occur along the western European margins, 680 in the Mediterranean Sea and the remaining in the Baltic (Galil et al. 2014). Marine invasive species pose a significant environmental threat as they are one of the major drivers of biodiversity loss (Millennium Ecosystem Assessment 2005), altering ecosystems and their dynamics, shifting the community structure and displacing endemic species (Bax et al. 2003; Wallentinus and Nyberg 2007; Molnar et al. 2008). Negative impacts may also be registered at the economic and social levels, affecting fisheries, aquacultures, tourism or human health (Bax et al. 2003; Molnar et al. 2008). Invasive species spread and occupy new marine and coastal ecosystems through several maritime introduction vectors such as ballast water, biofouling of vessels, aquaculture escape or ornamental species trade (see Williams et al. 2013 and references therein).

It is almost impossible to completely eradicate invasive non-native species after their establishment (Katsanevakis and Crocetta 2014). Preventing the establishment of new invasive species is, therefore, of extreme importance and probably the best way to avoid the negative impacts associated with the spread of non-native species (Otero et al. 2013; Katsanevakis and Crocetta 2014). Furthermore, a deep understanding of the invasion pathways and population sources is relevant not only to prevent new invasions but also to aid the development of effective conservation measures (Lawson Handley et al. 2011; Ghabooli et al. 2013; Galil et al. 2014). For instance, information on the ecology of the invasive species within its native range may help in choosing adequate biological control agents for containment measures (Roderick and Navajas 2003). Knowledge of the introduction vectors, population sources and their genetic characteristics can be used to create risk maps, which provide useful insights into the potential ecological impacts (Kulhanek et al. 2011), predicting the impacts of different routes and the probability of new invasions occurring (Molnar et al. 2008; Hulme 2009).

Reconstructing the invasion pathways and identifying the putative source populations with historical and contemporary vector records is a difficult task (Estoup and Guillemaud 2010;
In this context, molecular genetic data is a powerful tool to reconstruct the invasive history by identifying putative source populations and genetic bottlenecks (Estoup and Guillemaud 2010; Lawson Handley et al. 2011; Bock et al. 2015; Cristescu 2015). Although genetic data does not always allow for successful identification of these processes, there are three conditions which increase the probability of accurate reconstruction of the invasive pathway, all fulfilled for Fundulus heteroclitus: (1) extensive sampling along the native range to ensure all the putative source populations are known, (2) the presence of genetic structure in the native range to narrow down the putative source regions and (3) a short amount of time passed since the invasion so that processes such as genetic drift do not increase the genetic differentiation between native and invasive populations (Geller et al. 2010).

The mummichog, F. heteroclitus (Linnaeus, 1766), is a small teleost naturally occurring almost continuously in saltmarshes of the North American east coast, from Newfoundland to Florida (Hardy Jr 1978). This species is extremely resistant to a wide range of salinities and temperatures, and can be found in freshwater, brackish or saltwater, inhabiting sheltered coastal areas such as saltmarshes, tidal creeks, estuaries or bays all year-round (Bigelow and Schroeder 1953; Hardy Jr 1978). F. heteroclitus is one of the most stationary marine species (Bigelow and Schroeder 1953), with short dispersal distances (1-2 km, Fritz et al. 1975; Lotrich 1975), high site fidelity closely related to the saltmarshes (Kneib 1984) and short home ranges (36-38 m, Lotrich 1975).

Due to its biology and ecological patterns, F. heteroclitus is an important model for the study of adaptive genetic clines, exhibiting distinct and extensive latitudinal clinal variation in relation to morphological, physiological and genetic traits along the North American Atlantic coast, which have been thoroughly studied for a long time (e.g., Powers et al. 1986). Geographical analysis of morphological traits (Morin and Able 1983; Able and Felley 1986), enzyme-coding loci (Powers and Place 1978; Powers et al. 1986; Ropson et al. 1990), mitochondrial DNA (González-Vilaseñor and Powers 1990; Bernardi et al. 1993; Smith et al. 1998), putative neutral nuclear microsatellite loci (Adams et al. 2006; Duvernell et al. 2008), and a variety of nuclear single-nucleotide polymorphisms (Williams et al. 2010; Strand et al. 2012; McKenzie et al. 2015; McKenzie et al. 2016) demonstrated concordance of clinal patterns, where a break is placed between the meridians 40º and 41º N, along the coast of New Jersey. These studies concur with Morin and Able (1983), supporting the division into two subspecies: F.
heteroclitus heteroclitus (Linnaeus, 1766) to the north of the cline and F. heteroclitus macrolepidotus (Walbaum, 1792) to the south of the cline.

In the Iberian Peninsula, F. heteroclitus was first detected in the 1970s, in the Guadalquivir and Guadiana saltmarshes (southwestern coast of Spain) (Hernando 1975) and in the Guadiana Delta (Coelho et al. 1976). More recently, its presence was also reported in the Ebro Delta, northeastern coast of Spain (Gisbert and López 2007) and in the Ria Formosa, south of Portugal (e.g., Catry et al. 2006; Paiva et al. 2006b). The species was probably introduced between 1970 and 1973 in the Spanish saltmarshes (Fernández-Delgado 1989) either involuntarily via aquarium trade (Hernando and Soriguer 1992; Bernardi et al. 1995; Elvira and Almodóvar 2001), ballast water (García-Revillo and Fernández-Delgado 2009; Fernández-Delgado 2010), or intentionally for purposes of biological control (Gozlan 2010). The Ebro Delta individuals were probably caught in the Spanish saltmarshes to be used in aquaculture and the aquarium trade (Gisbert and López 2007) and were later released from captivity. Another possibility for the Ebro Delta introduction is an accidental escape from a research centre nearby, where this species had already been used as a scientific model (Gisbert and López 2007). Two previous studies have analysed the origin of the southwestern Spanish populations using mitochondrial DNA (mtDNA), concluding the founder individuals were original from the northern native region between Maine and Nova Scotia where the northern subspecies F. h. macrolepidotus is distributed (Bernardi et al. 1995; Fernández-Pedrosa et al. 1996).

In the present study, we aim to build on previously published studies on the invasive range of F. heteroclitus (Bernardi et al. 1995; Fernández-Pedrosa et al. 1996) by using (1) more sampling locations (one in the Mediterranean Sea and two locations in the eastern Atlantic, and thirteen native locations), (2) a significantly larger number of individuals (248 in total), and (3) a three times larger fragment of the mitochondrial DNA cytochrome b gene; and to evaluate the genetic diversity and invasion pathways. Given that the species was recently reported in the Iberian Peninsula and has a limited adult dispersal capability, we tested the hypothesis of a human-mediated single Iberian introduction followed by dispersal promoted along the main oceanographic currents. This hypothesis leads to the expectation of an Iberian invasion based on a few founder individuals, with consequent lower genetic diversity than the putative identified source population. Also, the Iberian populations are expected to show no evident genetic structure given the short time since invasion.
Materials and methods

Sampling. A total of 248 Fundulus heteroclitus individuals from 16 locations: 13 sites in the western Atlantic, one in the Mediterranean Sea and two locations in the eastern Atlantic (Table 1 and Fig. 1) were obtained and stored in 96% ethanol and kept at -20 ºC. The populations from the western Atlantic sampled above 40º N are hereafter referred to as northern locations (ID 1-8), while samples collected below that latitude are referred to as southern locations (ID 9-13). Samples collected from the Mediterranean and eastern Atlantic are referred to as Iberian (ID 14-16).

DNA extraction, PCR amplification and sequencing. Total genomic DNA was extracted from caudal fin tissue following a standard Chelex 100 protocol (Walsh et al. 1991). Extraction results were checked by electrophoresis in 0.8% agarose gel stained with GelRed. Polymerase Chain Reactions (PCR) were conducted in a total volume of 25 µL, with 1X buffer, 10 mM dNTPs, 10 mM of each primer, 1U Taq Advantage 2 Polymerase mix DNA polymerase (CLONTECH-TaKaRa), 2 µL of DNA and Milli-Q water to the final volume. A fragment of the cytochrome b (cyt b) gene (1000 base pairs) was amplified with the forward primer GludG-L14724 (5’-TGACTTGAARAACCAYCGTTG-3’) (Palumbi et al. 1991) and the reverse primer cb6b.h (5’-GGAATTCACCTCTCCGGTTTACAAGAC-3’) (Martin and Bermingham 1998). PCR amplification consists of an initial 4 min denaturation step at 95 ºC, followed by 40 cycles of 1 min at 94 ºC (denaturation), 1 min at 50 ºC (annealing) and 1.5 min at 72 ºC (extension), and a 5 min final extension step. When amplification was not successful the following profile was used: initial 3 min denaturation step at 95 ºC, followed by 32 cycles of denaturation for 30 s at 95 ºC, annealing for 30 s at 54 ºC and extension for 1 min at 68 ºC, and a final extension step for 4 min at 68 ºC. PCR products were checked afterwards by electrophoresis in a 1% agarose gel stained with GelRed. Mitochondrial DNA was purified by ethanol/sodium acetate precipitation (Sambrook and Russel 2001). Its purity and quantity were analysed using a NanoDrop1000 spectrophotometer (Thermo Fisher Scientific, USA). Sequencing was performed on an ABI 3130xl capillary sequencer (Applied Biosystems – CCMAR, Portugal) using the forward primer from the PCR amplification (GludG-L14724).
**Genetic analysis.** Cyt b sequences were aligned and manually checked using the software Geneious v4.8.2 (Biomatters, Ltd., Auckland, New Zealand). The number of haplotypes ($n$), number of private haplotypes ($n_p$), and the haplotype ($h$) (Nei and Tajima 1981) and nucleotide diversities ($\pi$) (Nei 1987) were calculated for each location using the DNAsP v5.10.1 (Librado and Rozas 2009). Sequences were uploaded to GenBank (MH809691 - MH809938). To represent the phylogeographic relationships among haplotypes, a haplotype network was constructed using the Median Joining algorithm implemented in NETWORK v5.0 (Bandelt et al. 1999, fluxus-engineering.com).

We used two approaches to infer the most probable source area of the introduced Iberian populations within the native range of *F. heteroclitus*: (1) the geographical distribution of haplotypes in native populations, and (2) the phylogeographical relationships among haplotypes.

**Ecological niche modelling.** To explore the potential spread of *F. heteroclitus* within European waters beyond its currently known locations we used an ecological niche model. We selected the interactive AquaMaps system in FISHBASE (www.fishbase.org). Using observed locations of adults to construct a range of acceptable environmental conditions within which a species can exist, this model uses a c-squares distribution modelling approach (Rees 2003) to predict the probability of occurrence of the adult fish for a particular location, possessing specific environmental parameters. While this is normally computed just within the native range of a species it can be extended within the interactive system to consider all possible locations, and the associated probability of occurrence linked to the location’s mean environmental state if the species was able to reach that region. The basic statistical approach is given in Kaschner et al. (2006) and the AquaMaps version is described in Kesner-Reyes et al. (2012). Another ecological niche model, based on maximum entropy principles (MAXENT, Phillips and Dudík 2008), was also tested using environmental parameters of sea surface temperature (SST) and sea surface salinity (SSS), but the specialist ecological needs of *F. heteroclitus* led to poor solutions for range prediction. This latter approach is therefore not considered further here.

A set of favourable and extreme environmental conditions compatible with the native occurrence of *F. heteroclitus* is automatically specified by AquaMaps, relying heavily on Page and Burr (2011). It includes parameter ranges for water depth, SST, SSS, primary production rates and sea-ice cover, leading to probabilities of occurrence exceeding 0.6 along the whole
eastern seaboard of North America from South Carolina to the Canadian Maritime Provinces south to the Gulf of St. Lawrence. This matches well to the known range (Page and Burr 2011). However, the long-term survival of *F. heteroclitus* in two Iberian estuaries has implications for extending the extreme salinity level that this species can tolerate beyond that automatically specified. The extreme salinity tolerance has therefore been raised to 39.6 psu, compared to the automatic 36.47 psu, consistent with the values found for the Ria Formosa (Cristina et al. 2016). In addition, the sea-ice variable has been excluded from constraining the AquaMaps solution, as this is not relevant to the area being considered in this study, and also the primary production variable, as coastal environments satisfy the automatic requirement almost uniformly. The environmental variables used for the AquaMaps simulation shown here are given in Table 2. AquaMaps was re-run with these environmental constraints and using the combined occurrence data from the native range, along the North American eastern seaboard, and the two Iberian sites of Ria Formosa and the Ebro River Delta. There is no change to the North American predictions, so here we concentrate only on those for European waters. The occurrence of *F. heteroclitus* was constrained to coastal zones with the muddy benthic habitats in which *F. heteroclitus* is found. These are defined as the coastal fine muds to muddy sands benthic environments of the EMODnet Seabed Habitats project (www.emodnet-seabedhabitats.eu). These environments are formally defined by Long (2006), but basically include sediments with more than a 10% mud component.

**Results**

Data analysis on the sampled specimens resulted in cyt *b* sequences with a final length of 700 base pairs (bp), which comprised 77 (11%) polymorphic sites and 32 (41.6%) parsimony-informative sites. These polymorphisms defined 70 haplotypes, of which 62 (88.6%) are private haplotypes (present in one location only) and 55 (78.6%) are singletons (present in one individual only). Overall haplotype diversity was high (0.74 ± SD 0.03), ranging from null in Iberia (ID 14-16) to 0.93 in Tuckerton (ID 9) and Roanoke Island (ID 12), whereas mean nucleotide diversity was low (0.26% ± SD 0.06%) ranging from null diversity in Iberia to 0.91% in Woods Hole (ID 4) (Table 1 and Fig. 2).
The most abundant haplotype in North American locations is shared by 50.4% (N = 125) of the individuals and is present in all northern group locations, in one southern location and in all Iberian locations (Fig. 1c). This is the only haplotype detected in the invasive range of the species (Faro, Cádiz and Ebro Delta). The second most frequent haplotype in North American locations is shared by 9.7% (N = 24) individuals in six locations, although in higher frequency in the southern group locations. Two other haplotypes were found in 3.2% (N = 8) and in 2.4% (N = 6) of the individuals from two locations. All other haplotypes were present in five or less individuals and in less than three locations (Fig.1d).

The cyt b-based haplotype network (Fig. 3) displays two haplogroups separated by 9 mutational steps. Haplogroup A is constituted by all eight northern group locations (Bridgewater to Red Bank, ID 1-8), two individuals from the southern group (Suffolk, ID 11) and by the Iberian locations (Faro to Ebro Delta, ID 14-16). Haplogroup B is formed by all the other individuals from all southern locations (Tuckerton to Skidaway Island, ID 9-13) and includes eight individuals from northern locations: one from Bridgewater (ID 1) and seven from Woods Hole (ID 4). Overall, both haplogroups display star-like configurations with different levels of complexity. The northern haplogroup (A) is simpler with 90% of the haplotypes separated by a single mutation, while the southern haplogroup (B) is more complex, with 3 mini-stars interconnected by one mutation each, and haplotypes separated by up to five mutations. The haplotype found in the Iberian Peninsula belongs to the northern haplogroup.

Conditions for the spread of *F. heteroclitus* are fundamentally favourable (probability>0.5), or at least acceptable (probability>0.25), along much of the Atlantic coastline of Europe and around the Alboran Sea in the western Mediterranean (Fig. 4).

**Discussion**

Results revealed the presence of a single haplotype common to all individuals in the Iberian Peninsula. This haplotype is the most abundant in the northern group of the native distribution, indicating the most probable origin of the invasion. We posit as the most likely introduction vector the aquarium trade and propose that the Ebro Delta colonization results from an independent human-mediated secondary introduction. Before addressing the main interpretations and conclusions of these results, one main caveat must be addressed. Successful understanding
of the invasion pathway relies on (1) comparable genetic data retrieved from an adequate number
of sampled individuals throughout the entire native range, (2) presence of genetic clines within
the native populations and (3) the use of adequate molecular markers (Geller et al. 2010). While
the first two points are fulfilled for *F. heteroclitus*, the use of a single mitochondrial DNA
marker constitutes the main caveat of this study. Mitochondrial DNA has been a widely used
molecular marker in population genetics studies (Ballard and Whitlock 2004). Nevertheless, the
use of high variable nuclear markers, such as microsatellites, provides an opportunity to perform
assignment tests based on their multiple-locus genotypes, to test for recent reductions in
population sizes and to estimate effective population sizes. However, previously published
studies show relatively low microsatellite genetic variation in the northern group (Duvernell et
al. 2008). Single nucleotide polymorphisms (SNPs, Morin et al. 2004), extend the previously
referred analytical possibilities improving their statistical power because of the sheer number of
existing loci genome wide.

Genetic diversity

The presence of a single haplotype common to all *F. heteroclitus* sampled in the Iberian
Peninsula lends support to the hypothesis of an extremely recent introduction of the species
which has not allowed the accumulation of mutations at the mtDNA level, and with a single
introduction event composed by a very small number of individuals (Roman and Darling 2007).
Theoretically, invasive species are expected to suffer loss of genetic variation since the new
established populations are often based on a few individuals, which by definition, have lower
genetic diversity than the native source populations (Nei et al. 1975; Allendorf and Lundquist
2003; Dlugosch and Parker 2008). The single-haplotype characteristic can be found in other
invasive species, such as *Equulites elongatus*, the slender pony fish (Sakinan et al. 2017);
*Cercopagis pengoi*, the fishhook waterflea, a planktonic cladoceran crustacean (Cristescu et al.
2001); *Corbicula fluminea*, the Asian clam (Gomes et al. 2016) and *Didemnum perlucidum*, a sea
squirt (Dias et al. 2016). However, many successful invasive species do not display significant
erosion of genetic diversity (Roman and Darling 2007; Dlugosch and Parker 2008; Rius et al.
2014). For example, in a recent review of the literature on European seas invasion genetics, in
54% of studies that compared the genetic diversity between introduced species and their native
range, 74% reported comparable levels of diversity between them, while only 23% displayed a
reduction in the genetic diversity of introduced species, and the remaining 2% showed an increase in diversity (Rius et al. 2014).

**Population sources**

We identified a single Iberian haplotype present in all northern populations in high frequency (between 47 and 93%) and in two individuals in Suffolk, one of the southern locations. According to our results, the northern group is the most probable source of the founder individuals, which corresponds to the natural range of the subspecies *F. h. macrolepidotus*. However, we cannot definitely exclude Suffolk as a presumptive population source. The absence of genetic diversity in the Iberian Peninsula populations prevents the precise determination of the putative source population. The low spatial resolution of our data arises from the lack of genetic variability in the invasive range, rather than insufficient sampling of *F. heteroclitus* individuals in the Iberian Peninsula or in its native range (Muirhead et al. 2008).

Nevertheless, our findings are consistent with two previous studies on the origin of invasive individuals found in the Guadalquivir. First, based on mtDNA restriction fragment length polymorphisms (RFLP) (Fernández-Pedrosa et al. 1996) reported the presence of two haplotypes: the most abundant corresponding to the northern haplotype 1, dominant between Maine and Nova Scotia in North America; the other haplotype did not match any of the sampled native haplotypes and we found no evidence of its presence in the present study, using a larger number of individuals. It was previously suggested to be either a native unsampled haplotype or an endemic haplotype from Iberia, which is rather unlikely due to its recent invasion (Fernández-Pedrosa et al. 1996) and total absence of records in the area. Although there are no reports of hybridization between *F. heteroclitus* and any of the Spanish endemic species, the presence of a new haplotype could nonetheless be due to hybridization (Rius et al. 2014).

Secondly, a study based on cyt b sequences (Bernardi et al. 1995) concluded the individuals from the Guadalquivir originated in the region between Maine and Nova Scotia. Although our results are congruent with this proposed origin, we show that source populations could come from lower latitudes (40° N), even if only the northern group is considered.
Introduction vector

Since it was first recorded in Iberian saltmarshes, several studies linked *F. heteroclitus* introduction to different vectors. The aquarium trade has been suggested as the most important vector responsible for the introduction of this species (Hernando and Soriguer 1992; Elvira and Almodóvar 2001; García-Llorente et al. 2008; Ribeiro et al. 2008a; Gozlan 2010), followed by ballast water (García-Revillo and Fernández-Delgado 2009; Fernández-Delgado 2010), biological control (Gozlan 2010) and unknown origins (Fernández-Delgado 1989; Gutiérrez-Estrada et al. 1998). We posit that the introduction via the aquarium trade is the most probable scenario responsible for the establishment of the first individuals in Iberia, given the genetic results obtained.

It is well known that vessels can transport large numbers of organisms from several species at the same time in their ballast water (Carlton 2001; Gollasch 2007). Since there are countless vessels active around the world (e.g., Kaluza et al. 2010), ballast water-mediated transport allows the possibility of multiple introduction events, each with large groups of individuals (Hulme 2009). Considering such large groups usually comprise higher genetic diversity than fewer individuals alone, the assemblages transported are likely to display levels of genetic diversity similar to the levels found within their native range (Roman 2006; Dlugosch and Parker 2008; Wilson et al. 2009). Our findings, however, suggest that this was not the introduction vector responsible for the spread of the species; in contrast to the expectation of similar levels of genetic diversity between the native and invasive range, the Iberian locations display a strong founder effect, with all samples sharing a single haplotype.

Similarly to ballast water, the aquarium and ornamentals trade transport many species at a global scale (Padilla and Williams 2004). For instance, at least 19% of the invasive fishes found in the Iberian Peninsula were introduced via the aquarium trade (Maceda-Veiga et al. 2013). However, there are two main differences between introductions that follow ballast water or the aquarium trade. First, each introduction event after aquarium release is likely to comprise a small number of individuals (Duggan et al. 2006; Roman and Darling 2007). Thus, even though this vector may be responsible for the establishment of several non-native species (Padilla and Williams 2004), multiple introductions would be necessary for an invasive species to display high genetic diversity (Facon et al. 2003; Roman and Darling 2007). Secondly, the individuals released by aquarists are usually adults of higher fitness, which makes them better adapted to
survive in a natural environment (Padilla and Williams 2004). Thus, not only are the $F. \text{heteroclitus}$ invasive genetic diversity and structure consistent with an introduction of a low number of individuals via the aquarium trade, but it is also plausible that a few resistant individuals would manage to survive, reproduce and colonize the environment in which they were released.

**Human mediated introduction in the Ebro Delta**

Although the absence of genetic structure within Iberia limits possible insights into the invasion pathway, our data support the hypothesis of a human-mediated introduction episode responsible for the establishment of $F. \text{heteroclitus}$ in the Ebro Delta, as previously suggested by Gisbert and López (2007) based on taxonomic identification. The hypothesis of a long-distance colonization via natural dispersal is quite unlikely.

According to a review of the geographical distribution of Cyprinodontiformes along the northeastern coast of Spain by (García-Berthou and Moreno-Amich 1991), no $F. \text{heteroclitus}$ individuals were found at the Ebro Delta in 1989; the first record of the species was only registered 16 years later by Gisbert and López (2007). Thus, we estimate the date of establishment in the Ebro Delta ranges between 12 and 28 years ago. Assuming this estimate is correct, a natural colonization hypothesis implies individuals would have taken roughly two decades to travel more than 1000 km from their southernmost limit located in the Guadalquivir saltmarshes (Gutiérrez-Estrada et al. 1998) to the Ebro Delta. However, when compared with the natural colonization of the Ria Formosa, this hypothesis seems quite improbable. While $F. \text{heteroclitus}$ was never collected during sampling events that happened in the Ria Formosa between 1980 and 2006 (Ribeiro et al. 2006; Ribeiro et al. 2008b; França et al. 2009), analysis of prey remains left by Little Terns ($Sterna \text{albifrons}$) in the salt-pans and barrier islands revealed this prey species was present in the salt-pans and adjacent channels at least since 2002 (Catry et al. 2006; Paiva et al. 2006a). Although this may sound contradictory, $F. \text{heteroclitus}$ could in fact have been present in the Ria Formosa in specific unsampled locations or at extremely low densities, avoiding capture. Nonetheless, assuming that colonization happened around 2002, it seems that $F. \text{heteroclitus}$ took no more than ca. 20 years to travel around 50 km from the Guadiana Delta, where it was first detected in 1976 (Coelho et al. 1976). This estimate indicates that if the Ebro colonization happened via natural dispersal, it must have happened 20 times
faster than the natural colonization of the Ria Formosa. Given that *F. heteroclitus* has very low dispersal abilities (e.g., Fritz et al. 1975; Lotrich 1975), the natural colonization scenario for the Ebro seems unlikely. Furthermore, if establishment in the Ebro Delta followed a natural range expansion one would expect to find several established populations between the Barbate marshes in the Guadalquivir, its southernmost location (Gutiérrez-Estrada et al. 1998), and the Ebro Delta. However, we know that no invasive individuals can be found at the Mar Menor (personal communication), a coastal lagoon with suitable conditions for successful establishment (Gutiérrez-Estrada et al. 1998; Pérez-Ruzafa et al. 2006), located midway between the Guadalquivir and the Ebro Delta.

Finally, the Strait of Gibraltar would represent a barrier to dispersal and gene flow, limiting *F. heteroclitus* natural range expansion towards the Mediterranean Sea (Fernández-Delgado et al. 1986; Moreno-Amich et al. 1999; Doadrio et al. 2002) because of the strong currents prevailing in the area and the absence of suitable habitats. Although uncertain for *F. heteroclitus*, this has already been documented for two other Iberian toothcarps, where restricted gene flow in this region led to speciation of the *Aphanius iberus* in the Mediterranean Iberian coast and *Aphanius baeticus* in the southwestern Atlantic Spanish coast (Perdices et al. 2001; Doadrio et al. 2002).

**Ecological niche modelling**

Conditions for the spread of *F. heteroclitus* are limited by the existence of benthic muddy saltmarsh environments compatible with the species’ ecological requirements. These habitats are only found near major estuaries or lagoons areas along the Atlantic and western Mediterranean coastlines. The exception to this is along the North African coast, where such environments are more continuous, although other environmental aspects are less favourable there, as suggested by the rapid decline in occurrence probability eastwards. The consequence of this benthic habitat constraint is to make along-shore colonisation unlikely in most areas, suggesting that any such sudden expansion would need to be aided by human release.

**Conclusions**
Fundulus heteroclitus invasive Iberian populations revealed the presence of a single cyt b haplotype common to all individuals. This haplotype is predominant in the northern group of the native distribution, and although we cannot determine which exact location was at the origin of the introduced individuals, one can identify the northern group as the source of the introduction. The lack of genetic diversity is consistent with a strong founder effect at the origin of F. heteroclitus in Iberia. Although there is no direct evidence, we infer that the most likely vector was the aquarium trade, and that the Ebro Delta colonization results from a human-mediated secondary introduction isolated from the rest of Iberia. Considering the tolerance of the species to high salinity, its temperature range, and the significant amount of colonized area in the Iberian southern region, we predict that F. heteroclitus will most likely keep on expanding its invasive range until it faces unfavourable environmental conditions.

Acknowledgements

Occurrence data for the AquaMaps modelling was provided through the AquaMaps website (www.aquamaps.org), and the model simulation was run through the same site. Information on seabed habitats contained in this study has been derived from data that is made available under the European Marine Observation Data Network (EMODnet) Seabed Habitats project (http://www.emodnet-seabedhabitats.eu/), funded by the European Commission’s Directorate-General for Maritime Affairs and Fisheries (DG MARE). RLC was supported by a post-doctoral fellowship (SFRH/BPD/109685/2015) from FCT (Fundação para a Ciência e Tecnologia, Portugal) and FSE (Fundo Social Europeu). The work was funded by RC funds from FCT strategic plan UID/Multi/04326/2013 granted to CCMAR. North American sample collection was supported by NSF grant OCE-0221879 to DDD. Authors are thankful to N. Franch and J.M. Queral (Parc Natural del Delta de l’Ebre, PNDE) for providing animals from the Ebro River Delta.

References


Fernández-Delgado C (1989) Life-history patterns of the salt-marsh killifish *Fundulus heteroclitus* (L.) introduced in the estuary of the Guadalquivir River (South West Spain). Estuarine, Coastal and Shelf Science 29: 573-582


Long D (2006) BGS detailed explanation of seabed sediment modified folk classification, Keyworth, UK


Millennium Ecosystem Assessment (2005) Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington, DC


distinguished from history in geographic clines? A population genomic study of killifish
(Fundulus heteroclitus). PLOS ONE 7: e45138

Pollution Bulletin 55: 323-332

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of
DNA for PCR-based typing from forensic material. BioTechniques 10: 506-513

Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF (2010) SNP identification,
verification, and utility for population genetics in a non-model genus. BMC Genetics 11: 1-14

Williams SL, Davidson IC, Pasari JR, Ashton GV, Carlton JT, Crafton RE, Fontana RE,
in an increasingly connected world. BioScience 63: 952-966

you move: dispersal pathways affect invasion success. Trends in Ecology & Evolution 24:
136-144
**Figure 1** (on next page)

Distribution map of localities and haplotypes

Distribution map of localities along the coast of (a) North America, and (b) Iberian Peninsula of *Fundulus heteroclitus*. Locations and sample details can be found in Table 1. (c) Coloured circles display the distribution of the relative proportions of the cytochrome *b* haplotypes from each location: the four most common haplotypes, the private haplotypes, and all the other shared, but less frequent haplotypes.
Haplotype and nucleotide diversity of *Fundulus heteroclitus* from all sampled locations. Abbreviations in the legend are defined in Table 1, and colours are the same as in Fig. 1.
Median-joining cytochrome b haplotype network for *Fundulus heteroclitus*

Median-joining cytochrome b haplotype network for *Fundulus heteroclitus*. Each circle represents a different haplotype with size proportional to the frequency of the haplotype within the sample. Line length is proportional to the number of mutations between haplotypes. Each colour corresponds to a different location. Colours are the same as in Figure 1. Locations details can be found in Table 1. The two black dots represent putative unsampled haplotypes, and the number in black square shows the number of mutations separating haplogroup A from the haplogroup B.
Figure 4 (on next page)

Western European and Mediterranean coastal environments showing AquaMap probabilities of occurrence of *F. heteroclitus*.

Western European and Mediterranean coastal environments showing AquaMap probabilities of occurrence of *F. heteroclitus*. Those areas in bold show coastal seabed habitats with a mud content > 10%.
Relative probabilities of occurrence

- 0.75 - 1.00
- 0.50 - 0.74
- 0.25 - 0.49
- 0.00 - 0.24
Table 1 (on next page)

Sample details

Sample location, sample abbreviations and summary statistics for a cytochrome b sequence fragment from *Fundulus heteroclitus*. ID refers to numbers in Fig 1
Table 1: Sample location, sample abbreviations and summary statistics for a cytochrome \( b \) sequence fragment from *Fundulus heteroclitus*. ID refers to numbers in Fig 1.

<table>
<thead>
<tr>
<th>Location</th>
<th>ID</th>
<th>Code</th>
<th>Latitude / longitude</th>
<th>( n )</th>
<th>( n_h )</th>
<th>( n_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridgewater</td>
<td>1</td>
<td>HV</td>
<td>44°22.0′N / 64°31.0′W</td>
<td>15</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Chewonki</td>
<td>2</td>
<td>CM</td>
<td>43°57.3′N / 69°43.2′W</td>
<td>15</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Wells</td>
<td>3</td>
<td>WM</td>
<td>43°19.2′N / 70°34.2′W</td>
<td>15</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Woods Hole</td>
<td>4</td>
<td>WH</td>
<td>41°31.5′N / 70°40.4′W</td>
<td>16</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Jerusalem</td>
<td>5</td>
<td>JR</td>
<td>41°23.1′N / 71°31.5′W</td>
<td>15</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Clinton</td>
<td>6</td>
<td>CC</td>
<td>41°15.3′N / 72°32.8′W</td>
<td>16</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Newark Bay</td>
<td>7</td>
<td>NB</td>
<td>40°41.2′N / 74°06.7′W</td>
<td>15</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Red Bank</td>
<td>8</td>
<td>RE</td>
<td>40°20.9′N / 74°05.0′W</td>
<td>15</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Tuckerton</td>
<td>9</td>
<td>TN</td>
<td>39°32.2′N / 74°19.4′W</td>
<td>15</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Speace</td>
<td>10</td>
<td>SP</td>
<td>38°09.1′N / 75°17.2′W</td>
<td>15</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Suffolk</td>
<td>11</td>
<td>CH</td>
<td>36°51.8′N / 76°28.7′W</td>
<td>16</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Roanoke Island</td>
<td>12</td>
<td>RI</td>
<td>35°53.8′N / 75°36.9′W</td>
<td>15</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Skidaway Island</td>
<td>13</td>
<td>SI</td>
<td>31°56.8′N / 81°04.2′W</td>
<td>16</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Faro</td>
<td>14</td>
<td>RF</td>
<td>37°00.3′N / 07°58.0′W</td>
<td>16</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cádiz</td>
<td>15</td>
<td>CD</td>
<td>36°31.4′N / 06°11.4′W</td>
<td>17</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ebro Delta</td>
<td>16</td>
<td>ED</td>
<td>40°37.38′N / 0°39.44′E</td>
<td>16</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\( n \), number of individuals; \( n_h \), number of haplotypes; \( n_p \), number of private haplotypes
**Table 2 (on next page)**

Aquamaps environmental envelope

Aquamaps default environmental envelope for *F. heteroclitus*
## Table 2 Aquamaps default environmental envelope for *F. heteroclitus*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Absolute</th>
<th>Preferred</th>
<th>Preferred</th>
<th>Absolute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth [m]</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Sea Surface Temperature</td>
<td>4.76</td>
<td>6.41</td>
<td>23.53</td>
<td>26.1</td>
</tr>
<tr>
<td>Sea Surface Salinity (SSS)</td>
<td>28.96</td>
<td>29.94</td>
<td>35.74</td>
<td>39.6</td>
</tr>
</tbody>
</table>