Elsevier Editorial System(tm) for Fisheries Research

Manuscript Draft

Manuscript Number: FISH8518R1

Title: The importance of distinguishing pufferfish species (Lagocephalus spp.) in the Mediterranean Sea for ensuring public health: evaluation of the genetic databases reliability in supporting species identification

Article Type: Research Paper

Keywords: toxic pufferfish, Lagocephalus guentheri, Lagocephalus spadiceus, Mediterranean environment, genetic databases

Corresponding Author: Dr. Andrea Armani,

Corresponding Author's Institution: University of Pisa

First Author: Alice Giusti

Order of Authors: Alice Giusti ; Nir Stern; Marcella Guarducci; Nadav Davidovich; Daniel Golani; Andrea Armani

Manuscript Region of Origin: ITALY

Abstract: Taxonomic identification of marine organisms is sometimes hindered by morphological similarities and utilization of wrong criteria. Therefore, the morphological approach often requires the support of molecular tools which usually rely on a comparison of DNA sequences available in free publicly-accessible databases. However, the process can be affected by wrongly deposited sequences which lead to specimens' misidentification. This is the case of two toxic pufferfish species (Lagocephalus spadiceus and L. guentheri), both reported as Lessepsian invasive species, whose actual presence in the Mediterranean is debated within the scientific community. In this study, the reliability of the genetic databases GenBank and BOLD in supporting the discrimination of L. spadiceus and L. guentheri was assessed as it has been already debated in literature. Twenty Mediterranean specimens of L. guentheri were collected and morphologically identified. COI and cytb reference sequences were then produced and included in two separate analyses (one for each gene) together with corresponding online sequences of L. spadiceus and L. guentheri from all the available localities. A high percentage of sequences with non-valid taxonomic identification was observed, involving 32.5% of the COI and 43.7% of the cytb sequences from GenBank and 30% of the COI sequences from BOLD. The majority of sequences deposited under L. spadiceus, mostly of Mediterranean origin, were genetically confirmed to be misidentified L. guentheri. Outcomes highlighted two main shortcomings: i) Errors by some contributors to official sequence databases, due to misidentification of the species from which these sequences have been obtained, thus attributing sequences to a wrong species when submitting their sequences ; (ii) a significant underestimation of L. guentheri presence in the Mediterranean Sea. This study, therefore, underlines the necessity to improve the methods which need to involve careful scientific morphological identification of the voucher specimens prior to DNA sequence submissions. In this specific

case, accuracy is even more important, as different species may have different toxicity and effects to potential accidental consumers.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Revision Notes
Click here to view linked References

The importance of distinguishing pufferfish species (*Lagocephalus* spp.) in the Mediterranean Sea for ensuring public health: evaluation of the genetic databases reliability in supporting species identification

Giusti A. [1¶], Guarducci M. [2¶], Stern N.[3], Davidovich N.[4], Golani D.[5], Armani A.[1]*

[1]FishLab, Department of Veterinary Sciences, University of Pisa, Via delle Piagge 2, 56124, Pisa (Italy).

[2]Istituto Zooprofilattico Sperimentale del Lazio e della Toscana, S.S. dell'Abetone e del Brennero 4, 56123 Pisa, (Italy).

[3]National Institute of Oceanography, Israel Oceanographic and Limnological Research, Haifa 3108001, Israel.

[4]Israeli Veterinary Services, P.O. Box 12, Bet Dagan 5025001, Israel.

[5]Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

* Correspondig author

Email: andrea.armani@unipi.it

Abstract

~~Taxonomical~~ Taxonomic identification of marine ~~species~~ organisms is sometimes hindered by morphological similarities and utilization of wrong criteria. Therefore, the morphological approach often requires the support of molecular tools which usually rely on a comparison ~~to~~ of DNA sequences available in free publicly-accessible databases. However, the process can be affected by wrongly deposited sequences ~~and~~ which lead to specimens' misidentification~~, consequently involving in gaps on the marine environmental status~~. This is the case of two toxic pufferfish species (*Lagocephalus spadiceus* and *L. guentheri*), both reported as Lessepsian invasive species, whose actual presence in the Mediterranean is debated within the scientific community. In this study, the reliability of the genetic databases GenBank and BOLD in supporting the discrimination of *L. spadiceus* and *L. guentheri* was assessed as it has been already debated in literature. Twenty Mediterranean specimens of *L. guentheri* were collected and morphologically identified. *COI* and *cytb* reference sequences were then produced and included in two separate analyses (one for each gene) together with corresponding online sequences of *L. spadiceus* and *L. guentheri* from all the available localities. A high percentage of sequences with non-valid taxonomic identification was observed, involving 32.5% of the *COI* and 43.7% of the *cytb* sequences from GenBank and 30% of the *COI* sequences from BOLD. The majority of sequences deposited under *L. spadiceus*, mostly of Mediterranean origin, were genetically confirmed to be misidentified *L. guentheri*. Outcomes highlighted two main shortcomings: i) Errors by some contributors to official sequence databases, due to misidentification of the species from which these sequences have been obtained, thus attributing sequences to a wrong species when submitting their sequences ~~a low taxonomic accuracy of official databases~~; (ii) a significant underestimation of *L. guentheri* presence ~~and a possible absence of *L. spadiceus*~~ in the Mediterranean Sea. This study, therefore, underlines the necessity to improve the methods which need to involve careful scientific morphological identification of the voucher specimens prior to DNA sequence submissions~~the databases accuracy~~

~~in term of deposited sequences reliability~~. In this specific case, accuracy is even more important, as different species may have different toxicity and effects to potential accidental consumers~~considering the involved toxic species and the potential concern for public health associated with their accidental entering in the seafood chain~~.

Keywords: toxic pufferfish, *Lagocephalus guentheri*, *Lagocephalus spadiceus*, Mediterranean environment, genetic databases

1. Introduction

~~The marine ecosystem is characterized by an extremely high biodiversity (Trivedi et al. 2016)~~. According to the last update of the World Register of Marine Species database - WoRMS (http://www.marinespecies.org/index.php), more than 240~~.~~000 accepted species names~~species~~ are known to be ~~globally~~ distributed in the marine environment. Acquainting with biodiversity and identifying species is fundamental to sustain the marine ecosystem, ~~allowing to~~ guard against worldwide threats such as climate change, pollution, overfishing, illegal fishing, habitat destruction, and invasive species (Duffy et al. 2013; Coll et al. 2014; Bunholi et al. 2018). It even acquires fundamental importance in the prevention of phenomena of counterfeiting and adulteration that entail economic loss for consumers and potentially have a damaging effect on public health if toxic species accidentally enter in the seafood chain.

Species identification usually requires the expertise of taxonomists and has been conventionally based on morphological observations and measurements. However, in some cases these practices are inadequate for their purpose (Wheeler et al. 2004). For instance, the morphological approach may not be applicable to species identification in early life history stages (eggs and larvae), as well as to the detection of cryptic species (Pampoulie and Daníelsdóttir, 2008; Trivedi et al. 2016~~3~~).

In view of this~~With this respect~~, DNA-based methods are nowadays increasingly used to reveal~~for revealing~~ new non-indigenous species, as well as serving to accurately identify and detect species in seafood products. Sequencing techniques are currently the most applied and, as a rule, rely on comparisons of DNA sequences from publicly available reference libraries. GenBank (https://www.ncbi.nlm.nih.gov/genbank/) is a comprehensive database that contains nuclear and mitochondrial nucleotide sequences for more than 400,000 named organisms, obtained primarily through submissions from individual laboratories and batch submissions from large-scale sequencing projects (Benson et al. 2018). Moreover, the Barcode of Life Data system - BOLD (http://www.boldsystems.org/) has gained worldwide popularity with the development and the success of DNA Barcoding, based on the use of a partial sequence of the mitochondrial cytochrome

c oxidase subunit 1 (COI) gene as a target region for species identification and discrimination (Hebert et al. 2003), and includes nowadays sequences from more than 270,000 organisms. However, pitfalls in both online-depositories are related to the fact that many of the available sequences lack validation in the form of voucher information (Zanzi and Martinsohn, 2017). Therefore, DNA sequences obtained from specimens that were not properly vouchered may belong to misidentified species, leading to a worrisome decrease in database reliability and to a consequent unavoidable presence of gaps and imprecisions in the knowledge of the marine environmental status.

The spreading of the invasive alien pufferfish species in the Mediterranean Sea initiated an emerging concern for public health, considering their poisonous viscera and flesh (Rambla-Alegre et al. 2017). In fact, they naturally harbour a heat-stable neurotoxin, called tetrodotoxin (TTX), which is potentially lethal if ingested in sufficient quantities (Bane et al. 2014). Nevertheless, pufferfish are known to possess varied levels of ~~toxicology~~toxin, thus it is important to confirm their exact identification for assessing and managing the risk related to their accidental fishing and subsequent entering in the seafood chain. With this regard, previous studies have especially highlighted problems in discriminating two pufferfish species from the genus *Lagocephalus*: *L. spadiceus* (Richardson, 1845) and *L. guentheri* Miranda Ribeioro, 1915, both morphologically and genetically similar (~~Matsuura et al. 2011; Tuney, 2016;~~ Vella et al. 2017).

~~Until now, t~~The available scientific literature has reported them as invasive 'Lessepsian species' that is~~, i.e.,~~ originated from the Red Sea (El-Haweet et al. 2016; Farrag et al. 2016; Turan et al. 2017). ~~However, c~~Considering the morphological resemblance between the two species and their recent revision (Matsuura et al. 2011; Psomadakis, 2015), it has been hypothesized that many of the old Mediterranean records of *L. spadiceus* could have been erroneously identified (Vella et al. 2017). The debate has gone so far that the real occurrence of *L. spadiceus* is reputed to be questionable (Vella et al. 2017). Moreover, although both the species are reported as "harmful" on the official finfishes' database (www.fishbase.org), *L. spadiceus* is described in literature as

occasionally poisonous (Chulanetra et al. 2011) and therefore represents an actual health risk if accidentally by-caught during commercial fishing. In this light, solving the issues hindering the proper discrimination between the aforementioned species might be essential for both clarifying the actual status of Mediterranean Sea in term of toxic species presence and assessing the consumers' risk exposure.

In this study, the accuracy of GenBank and BOLD databases regarding the deposited sequences of *L. spadiceus* and *L. guentheri* was in-depth evaluated. The taxonomic classification of the specimens from which the public sequences belong was revised after performing a phylogenetic analysis using reference sequences of *L. guentheri* produced in this study and problems related to database accuracy in identification of toxic species were discussed. Moreover, the Mediterranean records of the two species were re-evaluated in the light of the molecular results.

2. Materials and Methods

2.1 Collection of *L. guentheri* specimens and reference sequences production

2.1.1 Specimens collection and morphological identification. Twenty specimens were collected from the Israeli Mediterranean waters, and have been weighted, measured, and identified morphologically as *L. guentheri* according to the key features proposed by Matsuura et al. (2011) and Psomadakis (2015). The samples were then labelled with an internal code (Table 1) and stored at -20°C until further analysis.

2.1.2 Total DNA extraction, amplification and sequencing. Total DNA extraction was performed starting from ~50 mg of muscle tissue following the protocol described by Armani et al. (2014). Each DNA sample was stored at -20 °C until further analysis.

The primer pair Glu-PUF (5'-AACCACCGTTGTGATTCHACTACAA-3') and THR-PUF (5'-CGGCATCCGGYTTACAAGAC-3'), designed by modifying the primer pair proposed by Sevilla et al. (2007), was used to amplify 1134 bp of the *cytb* gene. The following PCR protocol was applied: 20 µl reaction volume containing 2 µl of a 10X buffer (BiotechRabbit GmbH, Berlin, Germany), 100 mM of each dNTP (Euroclone Spa, Milano), 200 nM of forward primer, 200 nM of

reverse primer, 1.0 U PerfectTaq DNA Polymerase (BiotechRabbit GmbH, Berlin, Germany), 100 ng of DNA and DNase free water (Euroclone Spa, Milano) with the following cycling program: denaturation at 95 °C for 3 min; 40 cycles at 95 °C for 30 s, 57 °C for 30 s, and 72 °C for 50 s; final extension at 72 °C for 7 min. In case of amplification failures, DNA samples were reamplified by coupling the primers Glu-PUF and THR-PUF with the primers 4R and 7F (Sevilla et al. 2007), respectively, to finally obtain the required full *cytb* fragment. PCR protocol followed Sevilla et al. (2007). Moreover, the standard ~655 bp *COI* barcode (Hebert et al. 2003) was amplified for three specimens of *L. guentheri* with the universal primer pair proposed by Handy et al. (2011) using the same PCR protocol and with the following cycling program: denaturation at 94°C for 3 min; 45 cycles at 94°C for 30 s, 53°C for 30 s, 72°C for 35 s; final extension at 72°C for 7 min. Five µl of each PCR products were inspected by gel electrophoresis on a 2% agarose gel. The successful PCR products were then purified with EuroSAP PCR Enzymatic Clean-up kit (EuroClone Spa, Milano) and sequenced by the Ictiopathology and Aquaculture Laboratory of the Istituto Zooprofilattico Sperimentale del Lazio e della Toscana (Pisa, Italy). Only three *COI* reference sequences were produced given the satisfactory number of *L. guentheri COI* sequences (n=40) available on the official databases (Table 2). Contrariwise, as only one *L. guentheri cytb* sequence (JQ681891) was available (Table 3), the *cytb* gene was sequenced from all the twenty specimens collected in this study to perform a fullest possible phylogenetic analysis as well as to contribute to the updating of the official database. All the sequences were then deposited in GenBank.

2.2 Sequences retrieval from databases and phylogenetic analysis

2.2.1 Sequences retrieval from databases. All the *COI* and *cytb* sequences belonging to specimens identified as *L. spadiceus* and *L. guentheri*, along with their corresponding sampling information, were retrieved from Genbank (https://www.ncbi.nlm.nih.gov/genbank/) and BOLD (http://www.boldsystems.org/), in the case of *COI* gene. The sequences belonging to the species *L. wheeleri* were also collected and treated as *L. spadiceus* according to Matsuura (2010). Sequences were then grouped into Mediterranean~~/Red Sea~~ (M~~/R~~), ~~ or~~ extra-Mediterranean (EM) or not

reported (NR) origin according to the available information. Sequences that lacked information for their locality were also included in our analyses, in order to maximize the assessment of the databases reliabilities. The entire inventory used for *COI* and *cytb* are listed in Table 2 and Table 3, respectively.

2.2.2 Phylogenetic analysis. The sequences produced in this study (Table 1) were combined with the previously published *COI* or *cytb* sequences of *L. guentheri* and *L. spadiceus* (Table 2 and Table 3) and were aligned and trimmed in two separate datasets. Genetic analysis was chosen for each dataset based on the amount of sequences involved: the large aligned assemblage of *COI* sequences was screened for population structure through haplotype diversities and frequencies, and a graphic median-joining un-rooted network was constructed using NETWORK v.4.6.1.2 (http://www.fluxus-engineering.com/sharenet.htm). The smaller size *cytb* assemblage was used to produce a neighbour-joining dendrograms (Saitou and Nei, 1987) using ~~200~~ 500 iterations of the Kimura 2-parameter (K2P) model (Kimura, 1980) in MEGA version 7.0 (Kumar et al. 2016). Pairwise comparisons of genetic distances were computed for both genes using the K2P model with ~~200~~ 500 bootstrap re-samplings. Finally, the taxonomical classification of the specimens used in this study was revised in the light of our outcomes.

2.3 Collection of Mediterranean records of *L. spadiceus* and *L. guentheri*

Reports of the invasive *L. spadiceus* and *L. guentheri* in Mediterranean waters were obtained from the online databases FishBase (www.fishbase.org), the European Alien Species Information Network (EASIN)(https://easin.jrc.ec.europa.eu/Services/SpeciesSearch), and previously published scientific articles available on Web of Science, PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Google scholar (https://scholar.google.it/). The search has used different combinations of the following keywords: Mediterranean Sea, pufferfish, Tetraodontidae, *Lagocephalus*, *Lagocephalus spadiceus*, *Lagocephalus guentheri*, Lessepsian species, alien fish species. Records from 1990 to date were reported in S1 Table. The following informative parameters were considered: year of

report, collection area, identification method (morphological or molecular) and classification criteria. Records were then re-evaluated in regard to the obtained molecular results.

3. Results and Discussion

3.1 Specimens collection: distinguishing morphological characters of *L. guentheri* and *L. spadiceus*

Recently, a taxonomic revision study involving Red Sea and the type specimens of *L. guentheri* was conducted, describing its specific differences of the shape and colour of the caudal fin from the closely-related *L. spadiceus* (Matsuura et al. 2011). In detail, *L. guentheri* has a slight posterior medially extension that makes the caudal fin appear to be doubly emarginate (Fig 1) whereas slightly more lunate in *L. spadiceus*. Moreover, in *L. guentheri* the caudal fin is dark brown or almost black excepting the dorsal and ventral white tips, while in *L. spadiceus* the dorsal two-thirds of the caudal fin are dark-yellow and the ventral one-third is white. Thesesis characteristic was also recently reported in the "*FAO species identification guide for fishery purpose*" by Psomadakis (2015) and it is thereforecan be considered as a reliable distinguishing character. As such, special attention to the shape and colour of the caudal fin has been given to identify the twenty specimens collected in this study. Overall, twenty-three reference sequences (twenty of *cytb* and three of *COI* gene) of *L. guentheri* were then produced and deposited in GenBank (Table 1).

3.2 Sequences retrieval from databases and phylogenetic analysis

3.2.1 COI sequences retrieval from databases Eighty-eight sequences, 40 from *L. guentheri* and 48 from *L. spadiceus,* were retrieved from the official databases (Table 2). *L. spadiceus* comprised 34 sequences that were deposited under its valid nomenclature and 14 sequences deposited with its synonym *L. wheeleri*. In three cases, the same sequences carried the valid taxon on BOLD and its synonym on GenBank.

In 63 out of the total 88 sequences (71.6%) the specimen's locality was reported, while this information lacked in the other 25 sequences (28.4%); in detail, of the 31 out of 40 *L. guentheri* sequences reporting the locality, 12 (38.7%) belonged to Mediterranean or Red Sea specimens

while the other 19 (61.3%) belonged to specimens from other areas (Table 2). Of the 33 out of 48 *L. spadiceus* sequences reporting the locality, 21 (63.6%) belonged to the invasive Mediterranean populations while the other 13 (39.4%) belonged to specimens from other areas. Of the 15 *L. spadiceus* sequences not reporting the locality, 12 were considered as belonging to extra-Mediterranean specimens (EM) as they were all deposited under the synonym *L. wheeleri*, that has never been reported from the Mediterranean Sea. Summarizing the origins of the available *COI* data, the specimens were grouped as follows: for *L. guentheri*, 12 M~~/R~~ (30%), 19 EM (47.5%) and 9 NR (22.5%) sequences; for *L. spadiceus*, 21 M~~/R~~ (43.7%), 24 EM (50%) and 3 NR (6.2%) sequences (Table 2).

3.2.2 Cytb sequences retrieval from database. The available inventory of *cytb* included 15 sequences of *L. spadiceus* (four of which have been deposited under *L. wheeleri*), and a single sequence of *L. guentheri*. Localities were reported for 11 out of the 16 total retrieved sequences (68.7%) (Table 3). All the 11 sequences reporting the locality belonged to *L. spadiceus*: two (18.2%) were sampled from the Mediterranean and nine (81.8%) from extra-Mediterranean populations. Within the 4 *L. spadiceus* sequences not reporting the locality, 2 (AP009538 and JQ681896) were considered as belonging to extra-Mediterranean (EM) specimens, as they were deposited under *L. wheeleri* (according to section 3.2.1), while for the other 2 (KM667972 and NC_026194) it was not possible to extrapolate information on the collection area (NR), as they were associated to unpublished studies. Altogether, the available *cytb* data of *L. spadiceus* has been grouped as follows: 2 M~~/R~~ (13.3%), 11 EM (73.3%) and 2 NR (13.3%) sequences (Table 3). The single *L. guentheri* sequence (JQ681891) was produced in the work of Santini et al. (2013) from the same specimens used for the production of the *COI* gene (previous section) and thus considered as EM as well (Table 3).

3.2.3 Population structure of COI sequences. The *COI* neighbour-joining network illustration contained 43 *L. guentheri* and 47 *L. spadiceus* sequences, and a single *L. sceleratus* sequence as an outgroup. As shown in Fig 2, the sequences clustered into three haplogroups that we conventionally

termed α(COI), β(COI) and γ(COI) clusters (Table 2). The α(COI) cluster included 42 *L. guentheri* sequences and 23 *L. spadiceus* sequences that belonged to 16 haplotypes; the β(COI) cluster included the remaining 24 *L. spadiceus* sequences that belonged to five haplotypes; the γ(COI) cluster contained two closely-related singletons that were vouchered under *L. spadiceus* (KP266858) and *L. guentheri* (JQ681796). Overall, this finding highlights issues associated with the deposited sequences of *L. spadiceus* sequences, as their position in the α(COI) cluster seemed to be associated with misidentified *L. guentheri* specimens. In this respect, 20 out of the 23 *L. spadiceus* sequences (87%) from the α(COI) cluster were obtained from specimens collected in Mediterranean Sea (M~~/R~~), 1 from India (EM) and 2 with ~~unknown~~ not reported ~~origin~~ (NR) origin (Table 2). Contrariwise, 22 out of 24 sequences of the β(COI) cluster (91.7%) belong to extra-Mediterranean (EM) populations, while only 1 sequence (4%) has a confirmed Mediterranean origin (HQ167726). However, it belongs to an unpublished study. Last, an identification error has been observed in two sequences (JQ681796 and KP266858) belonging to the γ(COI) cluster, showing a 11.4% genetic divergence from the *L. guentheri* sequences of α(COI) cluster, but revealed an intra-specific genetic resemblance of 0.3% with *Lagocephalus cheesmanii* (Clarke, 1897) or its recently synonymized taxon *Lagocephalus gloveri* Abe & Tabeta, 1983 (Matsuura and Satoh, 2017).

3.2.4 Phylogenetic analysis of cytb gene sequences. The *cytb* neighbour-joining dendrogram was produced with 15 sequences of *L. spadiceus*, 20 new *L. guentheri* sequences and a single sequence of *L. sceleratus* as an outgroup. The unique available *cytb* sequences of *L. guentheri* retrieved from the database (JQ681891) was ~~instead~~ not included in the analysis as it was obtained from the same individual used for producing the misidentified *L. cheesmanii*. Two haplogroups have been formed: α(cytb) cluster with ten haplotypes that comprised the twenty newly produced sequences and six sequences deposited as *L. spadiceus*; and β(cytb) cluster that contained a single shared haplotype for the remaining nine *L. spadiceus* sequences (Table 3; Fig 3). In regard to the localities of the different haplogroups, α(cytb) contained both Mediterranean, Thailand and unknown localities; β(cytb) contained only extra-Mediterranean localities (Table 3). Therefore, also in this case, it was

possible to confirm that *L. spadiceus* sequences from the α(cytb) cluster are misidentified *L. guentheri* specimens.

3.3 Database taxonomic inconsistencies

Online databases offer scientists the opportunity to easily access a wide variety of biologically relevant data, including DNA sequences of an increasingly broad range of organisms (Baxevanis and Bateman, 2006). Studies employing classical *Forensically Informative Nucleotide Sequencing* – FINS (Bartlett and Davidson, 1992) or DNA Barcoding (Hebert et al. 2003) generally rely on publicly available reference libraries deposited on NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and, especially in the case of DNA barcoding, on BOLD database (http://www.boldsystems.org/). These online depositories require therefore, an accurate and updated taxonomic support, and more preferably vouchered specimens in official natural history museums. While NCBI GenBank often contains unvouchered sequences with missing information, the BOLD platform generally encourages vouchered sequences with a collateral data, such as date, location of capture or~~,~~ images of the sequenced specimen ~~and more~~ (Ratnasingham and Hebert, 2007). However, cases of poor accuracy of both databases still exist (Armani et al. 2015; Costa et al. 2017; Vella et al. 2017), and may possesses unknown impacts on practical biological investigations, in terms of redundancies and inconsistencies in taxonomic identifications. Over the years, efforts have been carried out to improve the public data sources. Especially regardings~~s~~ fish species, Zanzi and Martinsohn (2017) recently suggested the use of the database FishTrace (https://fishtrace.jrc.ec.europa.eu/), that collects genetic data from specimens that have been identified by taxonomists and stored in natural history museums. However, this database only contains sequences from 200 fish species of commercial interest. Regarding these limitations, morphologically-similar species may represent a continuous obstacle for correct taxonomical identification and, consequently, increase the probability to deposit sequences from wrongly identified species. In our findings, neither GenBank nor BOLD databases were proved as sufficiently accurate, with a high percentage of not valid taxonomic identity in both databases

(Table 2, Table 3 and S2 Table). Overall, 32.5% (25/77) of the *COI* sequences and 43.7% (7/16) of the *cytb* sequences deposited on GenBank presented inaccuracy. Similarly, 30% (18/30) of the *COI* sequences deposited in BOLD were inaccurate. Noteworthy, some of these sequences (KM538367; KM538378; KM538379; KM5383817~~9~~) were actually associated to pictures belonging to misidentified specimens, since the morphology of the caudal fin was absolutely compatible with that of *L. guentheri*.

3.4 *L. spadiceus* and *L. guentheri* records in the Mediterranean waters

3.4.1 L. spadiceus records. While *L. spadiceus* presence in Mediterranean waters was not reported on the European Alien Species Information Network - EASIN (https://easin.jrc.ec.europa.eu/Services/SpeciesSearch), the finfishes' database (www.fishbase.org) states *L. spadiceus* as Lessepsian invasive species throughout the waters of Israel, Lebanon, Cyprus and Greece. According to the oldest literature, it seemed that its first record was from Iskenderun Bay (Turkey), followed by Samos Island (Greece) and Israel coast through the years from 1930-1940 (Tuncer et al. 2008), including multiple observations from the Aegean and Mediterranean Turkish seas from the 1960s to the 1990s ~~'60s to the '90s~~ (Bilecenoglu et al. 2002). Throughout the reports and until 1990, invalid nomenclature taxa such as *Sphoeroides spadiceus* or *Tetrodon spadiceus* were often used in the literature, further hindering the reliability of the identifications. Prior to the works of Matsuura et al. (2011) and Psomadakis (2015), the morphology and the colour of the caudal fin, that is crucial to distinguish between *L. spadiceus* and *L. guentheri*, has not been used and thus indicates a major flaw in the taxonomic verifications (S1 Table). In this study, we further elucidate this flaw showing that in our collected 32 records, two (Bariche et al. 2015; Tuney, 2016) out of the only five specimens (40%) genetically identified as *L. spadiceus* were instead *L. guentheri* (S2 Table).

3.4.2 L. guentheri records. Opposite from *L. spadiceus*, the presence of *L. guentheri* in the Mediterranean Sea has never been reported on the finfishes' database (www.fishbase.org). Actually, no records of this species were highlighted in literature before 2004, except for one in Greece,

reported on the European Alien Species Information Network - EASIN (https://easin.jrc.ec.europa.eu/Services/SpeciesSearch) and dating back to 1930. The only records of *L. guentheri* in the Mediterranean waters were the eight ones listed in S1 Table, and almost all were observed between 2011 and 2015 (Erguden et al. 2017). Noteworthy, in four out of the five records detailing the morphological classification on which they relied, the key features proposed by Matsuura et al. (2011) and Psomadakis (2015) were considered (S1 Table). Therefore, these records could be overall stated as reliable, even though molecular tools have not always been used to confirm the morphological identification. Our outcomes, corroborate those by Vella et al. (2017) who indicated that a problem with the identification of *L. spadiceus* and *L. guentheri* in the Mediterranean exists.~~Moreover, considering the outcomes of this study,~~ In this case, especially regarding the evidence of misidentification issues involving the databases, we could certainly affirm that most of the *L. spadiceus* specimens recorded in Mediterranean prior to the current morphological identification criteria were misidentified individuals of *L. guentheri*. ~~Thus, the actual presence of this species could not be assessed through this analysis, since it has been undoubtedly underestimated.~~

3.4.3 Underestimation of L. guentheri and possible absence of L. spadiceus in the Mediterranean Sea. The actual presence of the pufferfish *L. spadiceus* and *L. guentheri* in Mediterranean Sea represent a recent debated topic (Akyol and Aydın, 2016; Vella et al. 2017). Through the investigation in this study, we basically recognized an actual difficulty in taking an absolute stand. It seems in fact that the turning point to solve this conundrum was solely represented by the awareness that the two species only differ each other from the shape and colour of the caudal fin. Therefore, it is clear that, on the one hand, all the current records and studies that had not considered this feature should not be assumed as reliable. On the other hand, we can neither unquestionably confirm they were wrong, as the key feature of the recorded whole specimens can no longer be observed. The veracity of the theory that *L. spadiceus* has never been present in Mediterranean waters and all its recorded were instead *L. guentheri* might be suggested by the fact

366 that the European Alien Species Information Network - EASIN
367 (https://easin.jrc.ec.europa.eu/Services/SpeciesSearch), curated by the European Commission, only
368 reports *L. guentheri* as alien species. This hypothesis could not be confirmed in this study since one
369 sequence analysed (HQ167726) was proved to actually belong to a Mediterranean *L. spadiceus*
370 specimen but ~~. However,~~ the reliability of this sequence is regarded questionable as it was not
371 associated to voucher specimen or published study. However, by the outcomes of our molecular
372 analysis and the data from the collected records, jointly with the outcomes provided by Vella et al.
373 (2017), we could certainly state that the presence of *L. guentheri* in the Mediterranean Sea was
374 significantly underestimated.

375 **3.5 Toxic species and public health implications**

376 The colonization of the Mediterranean by Indo-Pacific and Red Sea flora and fauna, known as a
377 Lessepsian migration, is an ongoing process that began in 1869, following the excavations of the
378 Suez Canal and it is one of the most potent mechanism and corridor for invasions by marine species
379 known worldwide (Galil et al. 2015). In the last years, special attention has been given to the
380 invasion success of the silver-cheekd toadfish *L.~~agocephalus~~ sceleratus* (Gmelin, 1789), a highly
381 toxic Lessepsian species that has spread throughout most of the Mediterranean Sea, representing an
382 actual hazard for public health and a significant impact on the surrounding ecosystem (Guardone et
383 al. 2018; Kaligirou et al. 2012). Although not as investigated as *L. sceleratus*, the Indo-Pacific
384 congeners *L. spadiceus* and *L. guentheri*, are also reported as "Lessepsian migrators", although
385 much more veteran in the Mediterranean. Considering their difficult morphological distinguishing
386 characters, the real occurrence of these species in the Mediterranean Sea is of paramount
387 importance to understand their specific impact on ecosystem and public health. This latter aspect is
388 primarily related to the possible toxicity that has been attributed to *L. spadiceus*: although
389 commonly known as harmless (www.fishbase.org), the study by Chulanetra et al. (2011) proved that
390 liver extracts may sometimes cause death in mice. Thus, even though no toxicological studies have
391 been conducted on Mediterranean specimens, it would be appropriate to consider it as a health

hazard. It should even be noted that, since pufferfish act as hosts of TTX-producing bacteria that live symbiotically in their bodies and are accumulated through the food chain, the toxicity degree of a species is largely environment dependent (Noguchi et al. 2006). This aspect, coupled with the actual lack of scientific literature, ~~not~~ does not allow to establish the real toxicity level of Mediterranean *L. spadiceus*, neither to exclude the presence of TTX in Mediterranean *L. guentheri*. For this reason, a proper discrimination between the two species might facilitate target toxicological studies, essential to further assess the risk.

4. Conclusions

Taxonomic accuracy within publicly-available genetic databases represents a key factor for the reliability of the molecular analysis aimed at identifying marine species. Here, a case study involving two pufferfish species has highlighted the pitfalls within the most commonly used databases. The absence of implementing a taxonomic revision within this invasive species complex led to underestimation of the occurrence of the invasive *L. guentheri* in the Mediterranean and to an uncertainty regarding the actual invasive presence of *L. spadiceus*. This issue involves a significative limit in marine environmental research field, as well as in the seafood inspection, as the actual citizens' risk exposure cannot be predicted. As a consequence, we recommend that the taxonomic validity of the online sequences should be occasionally re-visited to assure correct taxonomic and phylogenetic employments. Moreover, it is evident that maximal attention should be given prior to sequences depositing, especially in cases where the veracity of the key morphological features used for identification is specially needed. Last, we strongly recommend that scientists who employ DNA-sequencing inventories in their studies will conduct a targeted preliminary analysis concerning the reliability of the selected references.

Acknowledgments

This study was supported by the Ministry of Health (Current Research Grant IZS LT 08/14 RC).

We particularly thank the anonymous reviewer that provided us additional and useful information on the morphological criteria for distinguishing the pufferfish species considered in this study.

Supporting Information description

S1 Table. *L. spadiceus* and *L. guentheri* records in Mediterranean Sea. Studies that applied molecular techniques using the identified specimens were highlighted in grey boxes and the relative investigated gene was reported in the last right column; NR: not reported; N: specimens number; * specimens' number was not reported, so that a single specimen was assigned to the record; ** specimens' number was not reported, but the available data suggested at least the record number here reported; (a): Specimens were identified with criteria different from the classification by Psomadakis (2015) and Matsuura et al. (2011); (b): Specimens were identified according to Psomadakis (2015) and Matsuura et al. (2011); (-): The identification criteria were not reported.

S2 Table. *COI* and *cytb* sequences proved as wrongly deposited on GenBank (G) and BOLD (B) databases. AN: accession number; Un: Unpublished study.

References in S2 Table:

References

Akyol, O., Aydın, İ., 2016~~7~~. A new record of *Lagocephalus guentheri* (Tetraodontiformes: Tetraodontidae) from the north-eastern Aegean Sea. Zool. Middle East 62(3), 271-273.

Armani, A., Guardone, L., Castigliego, L., D'Amico, P., Messina, A., Malandra, R., Gianfaldoni, D., ~~...~~Guidi, A., 2015. DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. Food Control 50, 589-596.

Armani, A., Tinacci, L., Xiong, X., Titarenko, E., Guidi, A., Castigliego, L., 2014. Development of a simple and cost-effective bead-milling method for DNA extraction from fish muscles. Food Anal. Methods 7(4), 946-955.

Asgharian, H., Sahafi, H. H., Ardalan, A. A., Shekarriz, S., Elahi, E., 2011. Cytochrome c oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf and analysis using meta- data flag several cryptic species. Mol. Ecol. Resour. 11(3), 461-472.

Bane, V., Lehane, M., Dikshit, M., O'Riordan, A., Furey, A., 2014. Tetrodotoxin: Chemistry, toxicity, source, distribution and detection. Toxins 6(2), 693e755.

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. J. Biogeogr. 42(12), 2363-2373.

Bartlett, S.E., Davidson, W.S., 1992. FINS (Forensically Informative Nucleotide Sequencing): a procedure for identifying the animal origin of biological specimens. Biotechniques, 12(3), 408-411.

Baxevanis, A.D., Bateman, A., 2006. The importance of biological databases in biological discovery. Curr. Protoc. Bioinformatics, 1.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K.D., Sayers, E.W., 2018. GenBank. Nucleic Acid Res. 46, D41-D47.

Bilecenoglu, M., Taskavak, E., Mater, S., Kaya, M., 2002. Checklist of the marine fishes of Turkey. Zootaxa, 113(1),1-194.

Bunholi, I.V., da Silva Ferrette, B.L., De Biasi, J.B., de Oliveira Magalhães, C., Rotundo, M. M., Oliveira, C., Forest, F., ... Mendonça, F.F.. 2018. The fishing and illegal trade of the angelshark: DNA barcoding against misleading identifications. Fish. Res. 206, 193-197.

Chen, W.B., Lin, S.H., Shao, B.Y., Zhao, C., Jiang, S.X., Yan, C., Zheng, J., 2012. Partial DNA Sequence Analysis of Cyt b Gene in Puffer Fish. Food Science 20, 045.

Chulanetra, M., Sookrung, N., Srimanote, P., Indrawattana, N., Thanongsaksrikul, J., Sakolvaree, Y., Chongsa-Nguan, M., Kurazono, H., .. Chaicumpa, W., 2011. Toxic marine puffer fish in Thailand seas and tetrodotoxin they contained. Toxins 3(10), 1249e1262.

Coll, M., Carreras, M., Ciércoles, C., Cornax, M.J., Gorelli, G., Morote, E., Sáez, R., 2014. Assessing fishing and marine biodiversity changes using fishers' perceptions: the Spanish Mediterranean and Gulf of Cadiz case study. PLoS One, e85670.

Costa, F.O., Landi, M., Martins, R., Costa, M.H., Costa, M.E., Carneiro, M., Alves, M.J., Steinke, D., ... Carvalho, G.R., 2017. A ranking system for reference libraries of DNA barcodes: application to marine fish species from Portugal. PloS one 7(4), e35858.

Duffy, J.E., Amaral-Zettler, L.A., Fautin, D.G., Paulay, G., Rynearson, T.A., Sosik. H.M., Stachowicz, J.J., 2013. Envisioning a marine biodiversity observation network. Bioscience 63(5), 350-361.

El-Haweet, A.A., Farrag, M.M., Akel, E.S.A., Moustafa, M.A., 2016. Puffer Fish Catch in the Egyptian Mediterranean Coast" The Challenged Invaders. International Journal of Ecotoxicology and Ecobiology 1(1), 13-19.

Erguden, D., Kabaklı, F., Uyan, A., Doğdu, S.A., Karan, S., Gurlek, M., Turan,C. (2017). New record of diamondback puffer *Lagocephalus guentheri* Miranda Ribeiro, 1915 from the North-Eastern Mediterranean, Turkey. NESciences 2(3), 67-73.

Farrag, M., El-Haweet, A.A., Moustafa, M.A., 2016. Occurrence of puffer fishes (Tetraodontidae) in the eastern Mediterranean, Egyptian coast-filling in the gap. Bioinvasions Rec. 5(1), 47-54.

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. 2015. 'Double trouble': the expansion of the Suez Canal and marine bioinvasions in the Mediterranean Sea. Biol. Invasions 17(4), 973-976.

Guardone, L., Gasperetti, L., Maneschi, A., Ricci, E., Susini, F., Guidi, A., Armani, A., 2018. Toxic invasive pufferfish (Tetraodontidae family) in the Mediterranean Sea and along Italian coasts: data collection and analysis for the assessment of an emerging public health risk. Food Control 91, 330-338.

Handy, S.M., Deeds, J.R., Ivanova, N.V., Hebert, P.D., Hanner, R.H., Ormos, A., Weigt, L.A., Moore, M.M., Yancy, H.F., 2011. A single-laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. J. AOAC Int. 94(1), 201-210.

Hebert, P.D., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B Biol. Sci. 270(1), S96-S99.

Hsieh, Y.W., Hwang, P.A., Pan, H.H., Chen, J.B., Hwang, D.F., 2003. Identification of tetrodotoxin and fish species in an adulterated dried mullet roe implicated in food poisoning. J. Food Sci. 68(1),142-146.

Kaleshkumar, K., Rajaram, R., Vinothkumar, S., Ramalingam, V., Meetei, K.B., 2015. DNA barcoding of selected species of pufferfishes (Order: Tetraodontiformes) of Puducherry coastal waters along south-east coast of India. Indian J. Fish. 62(2), 98-103.

Kalogirou, S., Azzurro, E., Bariche, M., 2012. The ongoing shift of Mediterranean coastal fish assemblages and the spread of non-indigenous species. In Biodiversity enrichment in a diverse world. Chapter from "Enrichment in a Diverse World", InTech. available at https://www.intechopen.com/books/biodiversity-enrichment-in-a-diverse-world/the-ongoing-shift-of-mediterranean-coastal-fish-assemblages-and-the-spread-of-non-indigenous-species (accessed on 15 March 2018).

Kimura, M.A., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16(2), 111-120.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7), 1870-1874.

Lin, H.Y., Chiu, M.Y., Shih, Y.M., Chen, I.S., Lee, M.A., Shao, K.T., 2016. Species composition and assemblages of ichthyoplankton during summer in the East China Sea. Cont. Shelf Res. 126, 64-78.

Matsuura, K., 2010. *Lagocephalus wheeleri* Abe, Tabeta & Kitahama, 1984, a junior synonym of *Tetrodon spadiceus* Richardson, 1845 (Actinopterygii, Tetraodontiformes, Tetraodontidae). Mem. Natl. Mus. Nat. Sci. 46, 39-46.

Matsuura, K., Golani, D., Bogorodsky, S.V., 2011. The first record of *Lagocephalus guentheri* Miranda Ribeiro, 1915 from the Red Sea with notes on previous records of *L. lunaris* (Actinopterygii, Tetraodontiformes, Tetraodontidae). Bull. Natl. Mus. Nat. Sci. (Ser. A) 37, 163-169.

Matsuura, K., Satoh, T.P., 2017. Redescription of *Lagocephalus cheesemanii* (Clarke 1897), a senior synonym of *Lagocephalus gloveri* Abe and Tabeta 1983, based on morphological and genetic comparisons (Actinopterygii: Tetraodontiformes: Tetraodontidae). Ichthyol. Res. 64(1), 104-110.

Noguchi, T., Arakawa, O., Takatani, T., 2006. TTX accumulation in pufferfish. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 1(1), 145-152.

Pampoulie, C., Daníelsdóttir, A.K., 2008. Resolving species identification problems in the genus *Sebastes* using nuclear genetic markers. Fish. Res. 93(1-2), 54-63.

Psomadakis, P.N., 2015. Field identification guide to the living marine resources of Pakistan. *FAO species identification guide for fishery purposes*. Available at: http://agris.fao.org/agris-search/search.do?recordID=XF2016038532.

Rambla-Alegre, M., Reverté, L., del Río, V., de la Iglesia, P., Palacios, O., Flores, C., Caixach, J., Campbell, K., Elliott, C.T., Izquierdo-Muñoz, A., ... Campàs, M., Diogène, J. 2017. Evaluation of tetrodotoxins in puffer fish caught along the Mediterranean coast of Spain. Toxin profile of *Lagocephalus sceleratus*. Environ. Res. 158, 1-6.

Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol. Ecol. Resour. 7(3), 355-364.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4(4),406-425.

Sangthong, P., Ngernsiri, L., Sangthong, D., 2014. Identification of puffer fish of the genus *Lagocephalus*: *L. lunaris*, *L. spadiceus* and *L. inermis*, using multiplex PCR. Food Biotechnol. 28(3), 216-231.

Santini, F., Nguyen, M.T.T., Sorenson, L., Waltzek, T.B., Lynch Alfaro, J.W., Eastman, J.M., Alfaro, M., 2013. Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). J. Evol. Bio. 26(5), 1003-1018.

Sevilla, R. G., Diez, A., Norén, M., Mouchel, O., Jérôme, M., Verrez- Bagnis, V., Van Pelt, H., Favre-Krey, L., Krey, G., the FishTrace Consortium and ... Bautista,J. M., 2007. Primers and polymerase chain reaction

conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. Mol. Ecol. Resour. 7(5), 730-734.

Trivedi, S., Aloufi, A.A., Ansari, A.A., Ghosh, S.K., 2016. Role of DNA barcoding in marine biodiversity assessment and conservation: an update. Saudi J. Biol. Sci. 23(2), 161-171.

Tuncer, S., Aslan Cihangir, H., Bilecenoglu, M., 2008. First record of the Lessepsian migrant *Lagocephalus spadiceus* (Tetraodontidae) in the Sea of Marmara. Fish. Sci. 69, 1224-1230.

Tuney, I., 2016. Molecular identification of puffer fish *Lagocephalus sceleratus* (Gmelin, 1789) and *Lagocephalus spadiceus* (Richardson, 1845) from Eastern Mediterranean, Turkey. Fresen. Environ. Bull. 25(5),1428-1436.

Turan, C., Gürlek, M., Ergüden, D., Uyan, A., Karan, S., Doğdu, S.A., 2017. Assessing DNA Barcodes for Identification of Pufferfish Species (Tetraodontidae) in Turkish Marine Waters. Natural and Engineering Sciences 2(3), 55-66.

Vella, A., Vella, N., Karakulak, F.S., Oray, I., 2017. DNA Barcoding of Tetraodontidae Species from the Mediterranean Sea: Filling Knowledge Gaps for Improved Taxonomic Accuracy. GenAqua 1(2), 71-79.

Wheeler, Q.D., Raven, P.H., Wilson, E.O., 2004. Taxonomy: impediment or expedient? Science 303(5656), 285.

Yamanoue, Y., Miya, M., Matsuura, K., Miyazawa, S., Tsukamoto, N., Doi, H., Takahashi, H., Mabuchi, K., Nishida, M.,... Sakai, H., 2008. Explosive speciation of Takifugu: another use of fugu as a model system for evolutionary biology. Mol. Biol. Evol. 26(3), 623-629.

Zanzi, A., Martinsohn, J.T., 2017. FishTrace: a genetic catalogue of European fishes. Database, 1-11, doi: 10.1093/database/bax075.

Zhang, J.B., Hanner, R., 2012. Molecular approach to the identification of fish in the South China Sea. PLoS One 7(2), e30621.

Zhang, J.B., Hanner, R., 2011. DNA barcoding is a useful tool for the identification of marine fishes from Japan. Biochem. Syst. Ecol. 39(1), 31-42.

Figure captions

Fig 1. Specimen of *L. guentheri* collected in this study. The presence of the characteristic caudal fin's medially extension is well appreciable.

Fig 2. Population structure of COI sequences. Graphic median-joining un-rooted network constructed using 48 *L. spadiceus* and 43 *L. guentheri COI* sequences.

Fig 3. Neighbour-joining dendrograms obtained with 15 *L. spadiceus* and 20 *L. guentheri cytb* sequences. All the retrieved sequences' accession numbers were reported. Different species clusters were highlighted in different coloured circles. Sequences produced in this study were highlighted with empty circles. L. GUE: *L. guentheri;* L. SPA: *L. spadiceus;* (L. WHE): *L. wheeleri* (synonymous of *L. spadiceus*).

Tables

Locality	Date	Sample code	Sequenced gene(s)	Accession no.
Ashdod	10 May 2015	L. GUE-1	*cytb*	MG817087
			COI	MG793381
		L. GUE-2	*cytb*	MG817088
			COI	MG817098
		L. GUE-3	*cytb*	MG817089
			COI	MG817099
		L. GUE-4	*cytb*	MG817090
	23 Jun 2015	L. GUE-5	*cytb*	MG817091
		L. GUE-6	*cytb*	MG817092
		L. GUE-7	*cytb*	MG817093
		L. GUE-8	*cytb*	MG817094
		L. GUE-9	*cytb*	MG817095
		L. GUE-10	*cytb*	MG817096
		L. GUE-11	*cytb*	MG817097
	18 Nov 2014	L. GUE-20	*cytb*	MG987173
Jaffa	21 Jul 2017	L. GUE-12	*cytb*	MH047219
		L. GUE-13	*cytb*	MH047220
		L. GUE-14	*cytb*	MH047221
		L. GUE-15	*cytb*	MH047222
		L. GUE-16	*cytb*	MH047223
		L. GUE-17	*cytb*	MG987174
		L. GUE-18	*cytb*	MG987171
		L. GUE-19	*cytb*	MG987172

Table 1. Sampling information for Israeli specimens of *L. guentheri* collected for this study, with their correspondent accession numbers from GenBank depository.

Cluster	Deposition name	*n*	Locality	Origin	Accession no.	Reference
α	*L. guentheri*	1	Turkey: Antalya	M~~/R~~	KY176508[1]	Un
		6	Israel: Mediterranean	M~~/R~~	BIM026-13, BIM074-13, BIM075-13, BIM355-13 BIM430-15, BIM431-15[2]	Un
		5	Egypt	M~~/R~~	KU324611-KU324615[1]	Un
		1	Saudi Arabia	EM	KU170600[1]	Un
		2	Iran: Bushehr, Nayband	EM	HQ149858-HQ149859[1]	Asgharian et al. (2011)
		5	South Africa: Tugela Banks	EM	JF493720-JF493724[1]	Un
		5	India	EM	KX675919, KX758092, KC409369-KC409371[1]	Un
		1	India	EM	KF442241[1]	Kaleshkumar et al. (2015)
		2	India: Tamil Nadu	EM	MSRRK001-16, MSRRK006-16[2]	Un
		1	India: Gujarat	EM	ANGEN229-15[2]	Un
		1	India, Tamil Nadu	EM	BDUMS006-13[2]	Un
		1	NR	NR	LC155439[1]	Un
		8	NR	NR	KP641410-KP641417[1]	Un

	L. spadiceus	1	India	EM	FJ384711[1]	Un
		17	Israel	M/~~R~~	KM538365-KM538381[1]	Un
		1	Greece	M/~~R~~	KY130423[1]	Un
		2	Lebanon	M/~~R~~	KR861535-KR861536[1]	Bariche et al. (2015)
		1	NR	NR	NC_026194[1]	Un
		1	NR	NR	KM667972[1]	Un
β	*L. spadiceus*	1	Turkey	M/~~R~~	HQ167726[1]	Un
	L. spadiceus	1	Japan	EM	LC155438[1]	Matsuura and Satoh (2017)
	L. wheeleri	1	Japan: Nagasaki, Teguma	EM	JF952772[1]	Zhang and Hanner (2011)
	L. spadiceus /L. wheeleri	1	Taiwan	EM	FJ434551[1]	Un
	L. spadiceus	2	Taiwan	EM	KT718613, KT718614[1]	Lin et al. (2016)
	L. spadiceus	4	China: South China Sea	EM	EU595160-EU595163[1]	Zhang and Hanner (2012)
	L. spadiceus	1	China: Guangdong	EM	EF607419[1]	
	L. wheeleri	2	NR	EM	AP009538, NC_011637[1]	Yamanoue et al. (2008)
	L. wheeleri	1	NR	EM	KY514080[1]	Matsuura and Satoh (2017)
	L. wheeleri	8	NR	EM	KP641418-KP641425[1]	Un
	L. spadiceus	1	NR	NR	KM667972[1]	Un
	L. spadiceus /L. wheeleri	1	NR	EM	JQ681802[1]	Santini et al. (2013)
γ	*L. spadiceus*	1	China: South China Sea	EM	KP266858[1]	Un
	L. guentheri	1	NR	EM	JQ681796[1]	Santini et al. (2013)

582 **Table 2.** *COI* accessions retrievals of *Lagocephalus* used for this study, clustered according to the
583 phylogenetic analysis. The "wrongly clustering" sequences are shaded. M/~~R~~: Mediterranean ~~or Red~~
584 ~~Sea~~ specimen; EM: extra-Mediterranean specimen; [1] – GenBank retrievals; [2] – BOLD retrievals;
585 Un: unpublished study; NR: not reported. Cases of the same sequences deposited with different
586 nomenclature on GenBank and BOLD are specified with /

587

Cluster	Species	*n*	Locality	Origin	Accession no.	Reference
α	*L. spadiceus*	1	Turkey	M/~~R~~	KP683378	Tuney (2016)
		1	Greece	M/~~R~~	KY130422	Un
		2	NR	NR	KM667972, NC_026194	Un
		2	Thailand	EM	KF781158-9	Sangthong et al. (2014)
β	*L. spadiceus*	2	China	EM	KT696359, KT833747	Chen et al. (2012)
	L. spadiceus	3	Thailand	EM	KF781155-7	Sangthong et al. (2014)
	L. spadiceus *	1	Taiwan	EM	FJ823445	Un
	L. spadiceus *	1	Taiwan	EM	AY128531	Hsieh et al. (2003)
	L. spadiceus *	1	NR	EM	AP009538/ NC_011637 (a)	Yamanoue et al. (2008)
	L. spadiceus *	1	NR	EM	JQ681896	Santini et al. (2013)
γ	*L. guentheri*	1	NR	EM	JQ681891	

588 **Table 3.** *Cytb* sequences retrieved from GenBank and used for the phylogenetic analysis.
589 Sequences were grouped according to the clusters obtained through the phylogenetic analysis. *
590 Originally deposited as *L. wheeleri*. The "wrongly clustering" sequences were highlighted in grey

boxes. Un: unpublished study; M~~/R~~: Mediterranean ~~or Red Sea~~ specimen; EM: extra-Mediterranean specimen; NR: not reported.

Fig.1

Fig.2

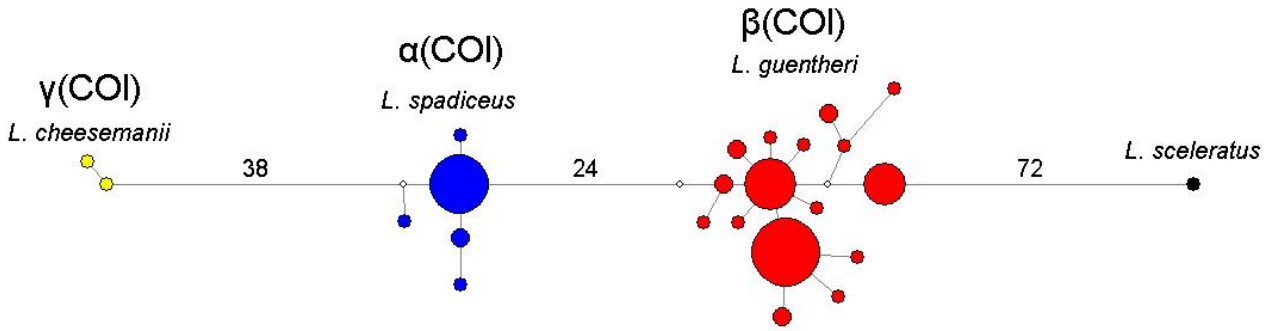
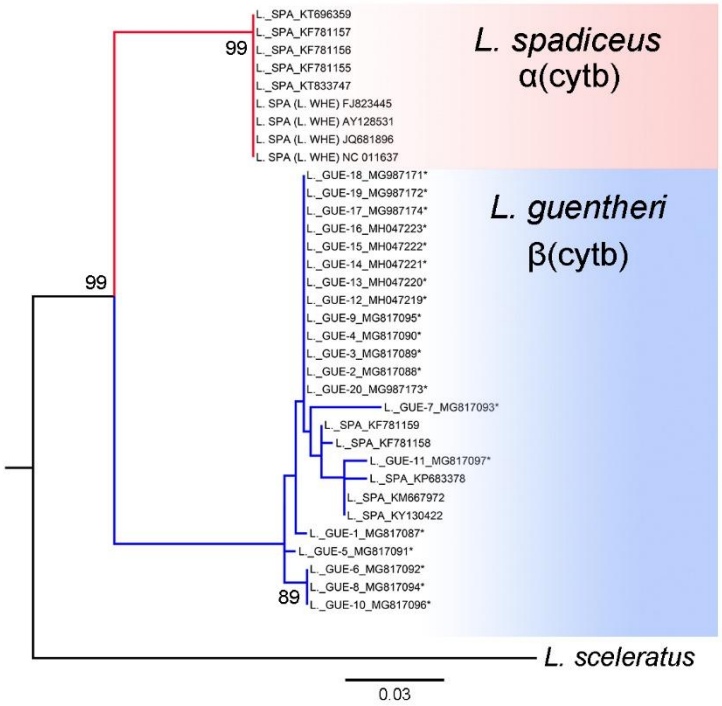


Fig.3



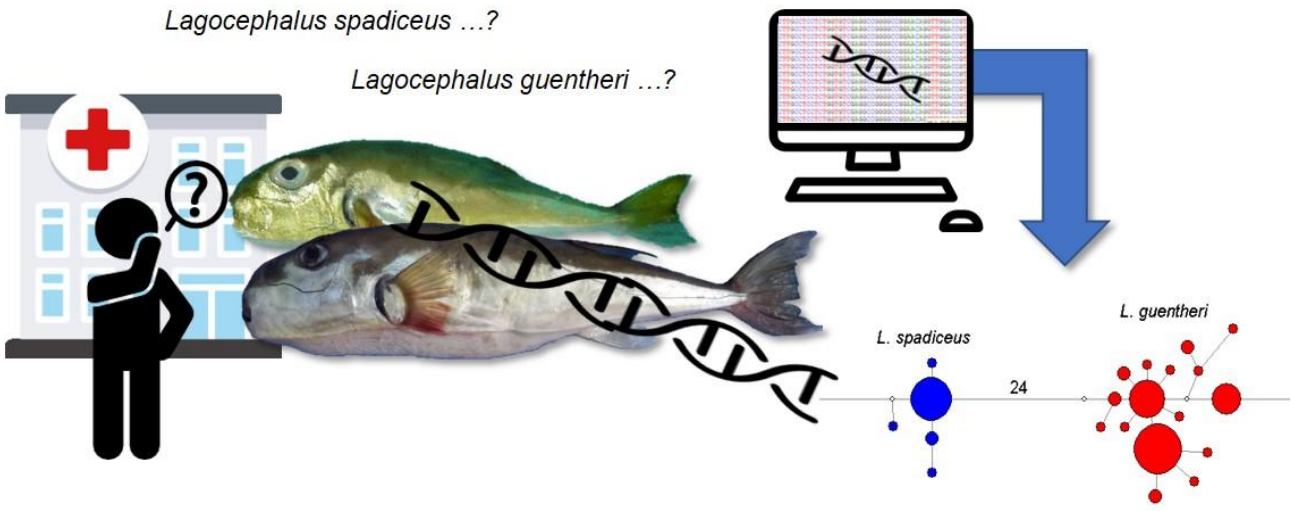
599

600

601 **Graphical abstract**

602 Molecular identification of two pufferfish species (*L. guentheri* and *L. spadiceus*), characterized by different
603 level of toxicity, recorded in the Mediterranean sea.

604



605

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: S1 Table (supplementary material).docx

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: S2 Table (supplementary material).docx