

## ***Invasive traits of freshwater fish database (ITOFF)***

### ***Abstract***

**AIM:** Species invasions are a major driver of global biodiversity loss, but only a minority of invasions are successful. Evidence suggests that invasive success is linked to life-history traits. Yet, data on invasive success and species' traits remain fragmented across multiple sources. Here we present the Invasive Traits of Freshwater Fish (ITOFF) database, an interdisciplinary framework that integrates multiple datasets to elucidate the role of life-history traits in shaping invasive success. ITOFF allows seamless access to invasive species data and fosters collaborative actions through knowledge sharing. ITOFF is supported by an innovative web-application that makes complex relationships between invasive and native species accessible to a broad audience. The scientific contribution of ITOFF is illustrated by examining the role of life-history traits and phylogeny in invasion success.

**LOCATION:** Global.

**METHODS:** Generalized linear models were used to test the contribution of generation time, trophic level, longevity, and temperature range to invasive success. Through divisive cluster analysis we investigate the role of multiple traits in determining invasive success. Finally, we construct phylogenetic trees to investigate the role of evolutionary history in the invasion process.

**RESULTS:** ITOFF unifies data for 1917 freshwater fish species representative of invasive species, those species they endanger, and species impacted by invasives but not considered endangered. Invasive species are generally characterized by greater temperature ranges, but are indistinguishable from impacted, endangered, and critically endangered species for the remaining life-history traits. Further, we show that invasive species are generally not distinct from impacted or endangered species when considering multiple traits or phylogeny.

**MAIN CONCLUSIONS:** ITOFF provides an accessible platform for the improved forecasting of species invasions. ITOFF data shows that classical predictions of life-history traits determining invasive success do not hold amongst freshwater fish species. Forecasting of invasive species must therefore shift towards a wholistic approach encompassing the species and the environment.

# 1 **INTRODUCTION**

2 Invasive species are a leading threat to global biodiversity (Reid *et al.*, 2019; WWF, 2020).  
3 Further, the economic and social costs of invasive species are severe (Andersen *et al.*,  
4 2004; Pimentel, Zuniga and Morrison, 2004; Hoffmann and Broadhurst, 2016). Nevertheless,  
5 measures can be implemented to mitigate the negative impacts of invasion (Olson, 2006).  
6 For example, recreational and commercial fisheries have been used to control invasive  
7 lionfish (*Pterois miles*) populations (Ulman *et al.*, 2021). However, the control of invasive  
8 species can be expensive (Heikkila and Peltola, 2004) and the impacts of invasion are often  
9 irreversible (Andersen *et al.*, 2004; Heikkila and Peltola, 2004; Keith and Spring, 2013).  
10 Prevention is, therefore, the optimal management strategy (Andersen *et al.*, 2004). As  
11 invasive propensity differs amongst species (Manchester and Bullock, 2000; Sakai *et al.*,  
12 2001), species with greater invasive propensity must be identified for the improved efficacy  
13 of preventative management (Sakai *et al.*, 2001). Life-history traits, such as fecundity and  
14 age of sexual maturation, are generally acknowledged to be critical determinants of invasive  
15 success (Holway and Suarez, 1999; Deacon, Ramnarine and Magurran, 2011; Chapple,  
16 Simmonds and Wong, 2012). However, few comparative studies have identified traits that  
17 promote invasive success in freshwater fish species at the global or major geographical  
18 region scale (Bernery, 2022). To address this gap, we present the Invasive Traits of  
19 Freshwater Fish (ITOFF) database. The ITOFF database provides easy access to research-  
20 ready data concerning the life-history traits and impacts of invasive freshwater fish species  
21 at the global scale.

22 Life-history traits that promote invasive success ('invasive traits') are predicted to be shared  
23 by a significant proportion of successful invasive species (Van Kleunen, Weber and Fischer,  
24 2010). The comparison of trait values between invasive and 'non-invasive' species allows  
25 the identification of invasive traits (Capellini *et al.*, 2015). However, identifying 'non-invasive'

26 species has proved challenging because it is extremely difficult to predict if a currently non-  
27 invasive species will become invasive in the future (Garcia-Berthou, 2007). For mammalian  
28 species, comparing successful and unsuccessful invaders enables the identification of traits  
29 critical to invasive success (Capellini *et al.*, 2015). This is possible because of the large  
30 number of documented, failed mammalian invasions (Capellini *et al.*, 2015). However,  
31 similar analyses are not appropriate for freshwater fish as introductions are not commonly  
32 reported in sufficient detail (Garcia-Berthou, 2007). To circumvent this issue, data can be  
33 collated on traits of species that are under threat of extinction from invasive species,  
34 because endangered species are unlikely to become invasive themselves. Further, we can  
35 incorporate trait data for species known to be impacted by invasives but not considered to be  
36 at risk of extinction (here onwards referred to as impacted species). This would be an  
37 appropriate control group. Data on the life-history traits of invasive species, those species  
38 they endanger, and impacted species have already been recorded (ISSG, 2000; Froese and  
39 Pauly, 2022; IUCN, 2022). Nevertheless, these data are stored in fragmented datasets  
40 managed by different bodies. Moreover, in their current format, data are not easily  
41 accessible for research purposes. The unified and public provision of life-history trait data for  
42 invasive species in an immediately accessible and re-usable format for research is urgently  
43 needed for the improved, predictive understanding of species invasions.

44 The ITOFF database creates the unique opportunity to address critical questions related to  
45 invasion biology by collating publicly available data on the life-history traits of the most  
46 common invasive, critically endangered and impacted freshwater fish species (freely  
47 available at <https://itoff-dataset.wp.st-andrews.ac.uk>). The accessibility of ITOFF is further  
48 supported by a web application (available at: [https://itoff-dataset.wp.st-](https://itoff-dataset.wp.st-andrews.ac.uk/webapplication/)  
49 [andrews.ac.uk/webapplication/](https://itoff-dataset.wp.st-andrews.ac.uk/webapplication/)) that facilitates the visualization of links between invasive,  
50 impacted, endangered and critically endangered species. Networks can be constructed  
51 based on the user's focal species and climate region of choice. To our knowledge, this is the  
52 first visualisation of the network of impacts associated with invasive species. These networks

53 reveal the complexity of impacts associated with invasive species. The harmonious  
54 incorporation of invasive, impacted, endangered, and critically endangered freshwater fish  
55 species in one dataset provides a novel approach to the study of invasion biology, while  
56 creating a powerful yet simple tool for future research.

57 To illustrate the scientific significance and accessibility of ITOFF, we investigate differences  
58 in life-history traits often predicted to facilitate species invasions among invasive, critically  
59 endangered, endangered, and impacted species for temperate, subtropical, and tropical  
60 regions. We focus on four traits shown to be critical to invasive success (Pimm, 1991; Crowder  
61 and Snyder, 2010; Zerebecki and Sorte, 2011; Deacon and Magurran, 2016; Rosenthal *et al.*,  
62 2021). Namely, generation time, trophic level, longevity, and temperature range. We expected  
63 invasive species to have shorter generation times and reduced longevity to facilitate fast  
64 population growth (Pimm, 1991) and rapid adaptation (Rosenthal *et al.*, 2021). Further, we  
65 expected invasive species to have wider temperature ranges, as this increases the likelihood  
66 of survival across a range of environments (Zerebecki and Sorte, 2011). Finally, we expected  
67 invasive species to be characterised by intermediate trophic levels (i.e. generalists).  
68 Generalist species have broader resource use which increases competitive advantage  
69 (Crowder and Snyder, 2010) and facilitates behavioural flexibility that in turn promotes survival  
70 under novel conditions (Deacon and Magurran, 2016).

71 There is the possibility that invasive success is not shaped by individual traits but rather by a  
72 combination of several. For example, invasive gammarid amphipods cannot be identified from  
73 individual traits but can be predicted using ecological profiles that combine multiple  
74 characteristics (Grabowski, Bacela and Konopacka, 2007). We therefore test the hypothesis  
75 that invasive propensity is not tied to individual traits but arises from a combination of multiple  
76 traits. As phylogenetic history influences trait similarity, we additionally utilise the detailed  
77 taxonomic information provided by the ITOFF database, to untangle the role of adaptation,  
78 chance, and evolutionary history in the invasion process.

## 79 ***METHODS***

### 80 ***The ITOFF Database***

81 A dedicated worksheet was created for the collection of all data. Data fields include group  
82 (i.e., invasive, endangered, or impacted), taxonomy, climate region (divided broadly into  
83 temperate, tropical, subtropical, polar, and boreal), IUCN status (a detailed measure of  
84 extinction risk), impacts on other species, and life-history traits predicted to be critical to  
85 invasive success (Table A1 Appendix).

86 A list of non-native freshwater fish species was generated using data from the Global  
87 Invasive Species Database (GISD; ISSG, 2000), the U.S. Geological Survey's  
88 Nonindigenous Aquatic Species (NAS) database (USGS, 2022), CABI's Invasive Species  
89 Compendium (ISC; CABI, 2022), the R interface to FishBase (Froese and Pauly, 2022), and  
90 data provided by the Global Register of Introduced and Invasive Species (ISSG, 2021). The  
91 IUCN Red List of Threatened Species (2022) was used to generate three lists: (1) critically  
92 endangered and (2) endangered species for which invasive species are considered a threat,  
93 and (3) least concern species that are known to be impacted by invasives. Finally, through  
94 the screening of invasive species' profiles, provided in text format by online databases  
95 (ISSG, 2000; CABI, 2022; USGS, 2022; Froese and Pauly, 2022), a list of species known to  
96 be impacted by invasive species was generated. Cross-referencing this list with the IUCN  
97 Red List allowed for the identification of impacted species considered to be of 'least concern'  
98 in terms of extinction risk; these species were included in the impacted group of the dataset.  
99 Scientific names for all species were updated to be in accordance with FishBase (Froese  
100 and Pauly, 2022). Further, the database was curated to identify and omit cryptic duplicate  
101 entries.

102 All species in the ITOFF database are categorised as non-native (n = 979), endangered (n =  
103 299), critically endangered (n = 213), or impacted (n = 381). Non-native species are further  
104 categorised as invasive, established, extirpated, failed, or reported. Categorisation of non-  
105 native species within ITOFF reflects the policy of governments, NGOs, and international  
106 agencies by defining species as invasive only if there is an associated negative impact  
107 (Rejmanek *et al.*, 2002). Species are considered reported if known to have been introduced  
108 to a novel ecosystem, but survival beyond introduction is unknown. Reported is the default  
109 categorisation for all non-native species. Non-native species that have successfully founded  
110 one or more self-sustaining populations are considered 'established'. Non-native species that  
111 have never survived to establishment are considered 'failed'. Established species that were  
112 later eradicated across their non-native range by human management are considered  
113 'extirpated'. Finally, species are categorised as 'invasive' if there are known (or a reported risk  
114 of) negative economic, ecological or social impacts associated with the species within its non-  
115 native range. It should be noted that of the 535 extant invasive and established species in the  
116 ITOFF database, for which IUCN status is reported, only 7 % are endangered or critically  
117 endangered.

118 The ITOFF dataset reports IUCN status and native/non-native status as separate variables.  
119 Accordingly, critically endangered, endangered, and impacted species with known non-  
120 native populations have been identified and recategorized as non-native within the ITOFF  
121 database. Further, impacted species reported by the IUCN Red List of Threatened Species  
122 to have undergone large-scale population declines or localised extinctions due to invasive  
123 species were recategorized as 'locally threatened'. This was necessary to ensure that  
124 impacted species in the dataset are representative of native species that are tolerant of  
125 invaders.

126 All data were collected by a small team of researchers (N = 11). All researchers followed the  
127 same systematic methodology for data collection as outlined below. For all species, online

128 databases, namely FishBase (Froese and Pauly, 2022) and the IUCN Red List of Threatened  
129 Species (IUCN, 2022), were screened for data on relevant traits. FishBase was further used  
130 for its unique estimate of trophic level, and the life-history tool was used to produce appropriate  
131 estimates for key traits (age of sexual maturity, generation time, and longevity) when data  
132 were otherwise unavailable. Note that estimates for trait values were only included if data were  
133 available for the species or the species' family. Estimated and observed values for the same  
134 trait are recorded in separate columns within the ITOFF database for clarity. For all non-native  
135 species, data on traits were additionally collected from species profiles provided by the GISD  
136 (ISSG, 2000), ISC (CABI, 2022), and the NAS database (USGS, 2022). Means of introduction  
137 was extracted from non-native species' profiles and the R interface to FishBase (USGS, 2022;  
138 CABI, 2022; Froese and Pauly, 2022). Links between species (e.g., predation of an invader  
139 on one or more native species) were recorded from species profiles provided by the IUCN  
140 Red List of Threatened Species, and above-mentioned databases for non-native species.  
141 Links between species were only included if reported to the species level (i.e., reports of an  
142 impact by an invasive species on a taxonomic rank higher than the species level are not  
143 included in the ITOFF database).

144 For all traits, if different data sources provided contrasting values, the mean value was used  
145 to populate the dataset. Detailed taxonomic information for all species was extracted from the  
146 GISD (ISSG, 2000), NAS database (USGS, 2022), and the IUCN Red List of Threatened  
147 Species (IUCN, 2021).

148 To validate the data, hybrids and species of taxonomic uncertainty (i.e., disputed subspecies  
149 or inconsistent and unclear nomenclature used in literature) were identified and omitted from  
150 the species pool. Conditional formatting rules, created via the Excel Visual Basic for  
151 Applications (VBA) tool, were used to control for data entry errors. Namely, entries not  
152 matching specified entry formats were visually highlighted, easily recognised, and corrected.



153 Checks were then made in R (R Core Team, 2021) using base functions to ensure consistent  
154 formatting was used throughout and to control for duplicates and typos.

### 155 ***Linking life-history traits and invasive success***

156 We used the ITOFF database to examine differences in life-history traits amongst invasive,  
157 critically endangered, endangered, and impacted species in temperate, subtropical, and  
158 tropical regions. Polar and boreal regions were excluded from the investigation because of  
159 limited data availability. The four traits of interest, generation time, longevity, temperature  
160 range, and trophic level, have been systematically reported as key traits for invasive success  
161 (Pimm, 1991; Crowder and Snyder, 2010; Zerebecki and Sorte, 2011; Deacon and Magurran,  
162 2016; Rosenthal *et al.*, 2021).

163 Values of life-history traits were compared between the four treatment groups using four  
164 generalized linear models (one per trait). Both observed and estimated trait values as reported  
165 within ITOFF were included in the analyses. Each full model included trait value as the  
166 response variable and species' status (i.e. invasive, impacted, endangered) as a fixed factor.  
167 Separate analyses were conducted for each climate region. Species that occupy multiple  
168 climate regions were represented in multiple counts. Locally threatened species were  
169 excluded from the analyses as they are representative of native species that have been  
170 severely impacted by invasive species within part or all of their sympatric range. Error  
171 distributions were chosen based on data type and distribution of the response variable.

### 172 ***Multiple traits in determining invasive success***

173 To investigate the role of multiple traits in determining invasive success a hierarchical cluster  
174 analysis, divisive analysis using the *diana* function of the *Cluster* package in R v 4.2.0.  
175 (Maechler *et al.*, 2022), based on age of sexual maturity, trophic level, and preferred  
176 temperature range was performed. We utilised a random subset of 75 tropical species  
177 included in the ITOFF database (25 invasive, 25 endangered, and 25 impacted). A subset was

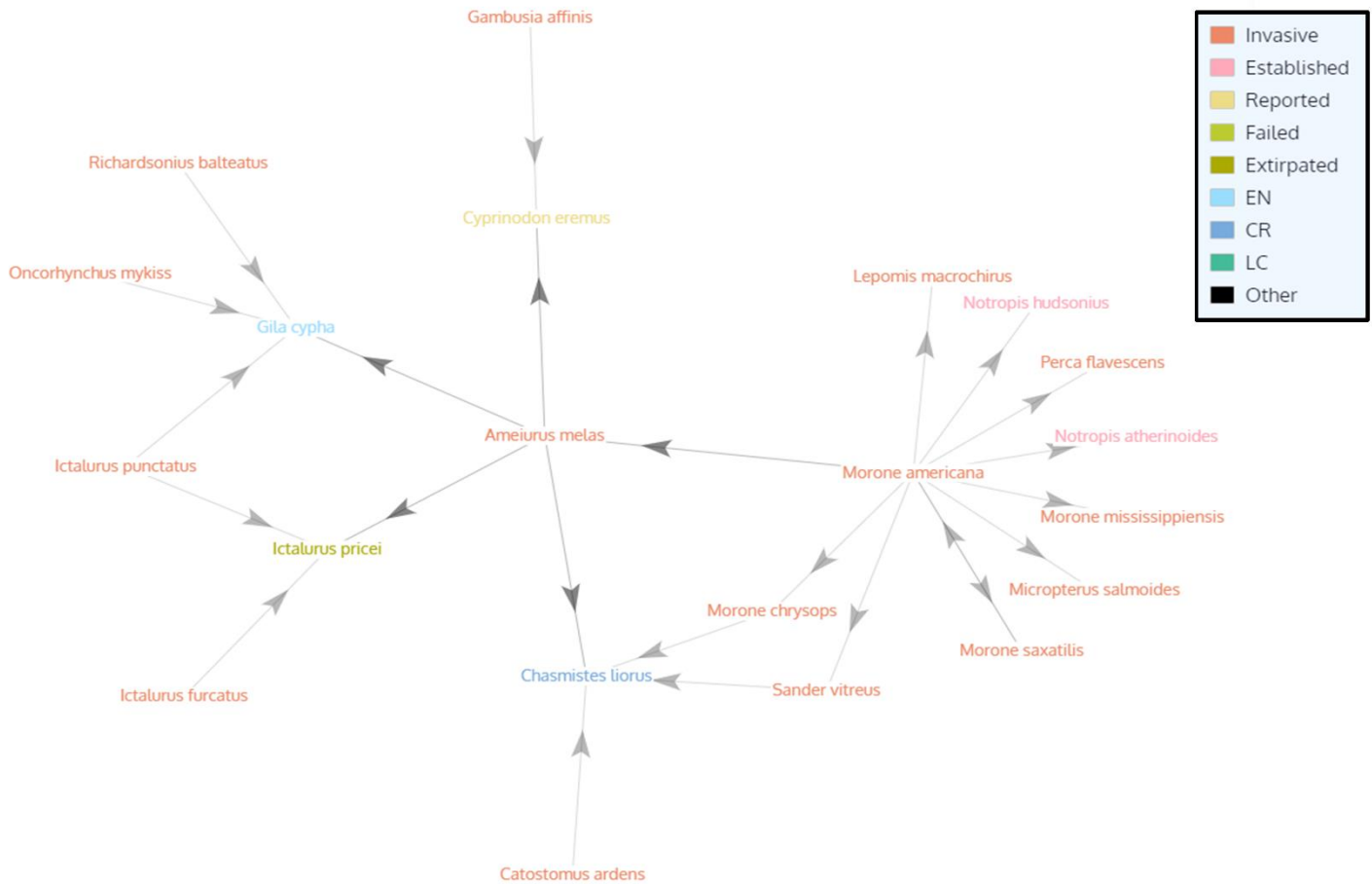
178 used because data on the three traits of interest were often not reported collectively for all  
179 species, and data availability differed between treatment groups. To investigate the  
180 relationship between adaptation and evolutionary history in the context of invasive species,  
181 we constructed a taxonomic tree utilising the species pool used for the cluster analysis.

## 182 **RESULTS**

### 183 ***ITOFF Database***

184 Data availability and sample size differ amongst species categories within ITOFF. Limited data  
185 availability reflects knowledge gaps which can be mitigated with the ITOFF database. For  
186 example, of all non-native species in the ITOFF database, only 31 are recorded in polar  
187 climatic zones. Further, relatively few species are reported as failed invaders, and data  
188 availability within this group is limited.

189 The ITOFF web application, clearly shows that endangered and impacted species are affected  
190 by multiple invasive species (Figure 1). Further, invasive species are shown to have impacts  
191 on other invasive species. Information on the type of impact (i.e., competition, predation), and  
192 profiles for the species selected are provided when using the web application directly.



193 **Figure 1.** Network showing the direction of impacts of invasive species, as taken from the ITOFF web  
194 application. The network is illustrated with *Ameiurus melas* selected as the focal species. The degree  
195 of separation between species is set to two, distribution of the species across climate regions is not  
196 specified. The web application is illustrated using a colourblind-friendly palette.

## 197 **Linking life-history traits and invasive success**

198 Invasive species had greater preferred temperature ranges than impacted species across  
199 temperate ( $13.23 \pm 1.07$ ,  $p = 0.005$ ), subtropical ( $12.2 \pm 1.06$ ,  $p = 0.001$ ), and tropical ( $8.38 \pm$   
200  $1.05$ ,  $p = 0.001$ ) climate regions. Endangered and critically endangered species did not differ  
201 significantly from impacted species for preferred temperature range across any of the three  
202 climate regions. For all other traits, observed differences between invasive, endangered,  
203 critically endangered, and impacted species did not match predictions (Table 1, Figure 2).

204 Further, there was no consistent pattern in trait values between climate zones. The full model  
205 outputs are provided in Table A2 (Appendix).

206 In temperate zones, invasive, endangered, and critically endangered species did not differ  
207 from impacted species in terms of mean trophic level (Table 1, Figure 2). Contrary to our  
208 predictions, invasive, and critically endangered species had significantly greater generation  
209 times than impacted species. Further, invasive, endangered, and critically endangered  
210 species all had greater longevity than impacted species (Table 1, Figure 2).

211 Similar to temperate zones, invasive, endangered, and critically endangered species had  
212 greater generation times and longevity than impacted species in subtropical zones (Table 1,  
213 Figure 2). Additionally, neither invasive nor critically endangered species differed significantly  
214 from impacted species for trophic level (Table 1, Figure 2). Conversely, endangered species  
215 had significantly lower trophic level values than impacted species.

216 In tropical zones, invasive species had significantly greater generation times, longevity, and  
217 preferred temperature ranges than impacted species but did not differ significantly for trophic  
218 level (Table 1, Figure 2). In contrast, endangered species did not differ significantly from  
219 impacted for any of the four traits. Contrary to our predictions, critically endangered species  
220 had greater trophic level scores and reduced longevity relative to impacted species (Table 1,  
221 Figure 2). Critically endangered species did not differ from impacted species for generation  
222 time or temperature range in tropical zones.

223 **Table 1.** General direction of predicted and observed differences between critically endangered,  
 224 endangered, and invasive fish species relative to species impacted by invasives but considered least  
 225 concern in terms of extinction risk (impacted; control group). Green arrows indicate significantly  
 226 greater trait values for that group, red arrows indicate significantly lower trait values, while a dash  
 227 indicates no significant differences. Data are split between three major climate regions (Temperate,  
 228 Subtropical, and Tropical). Asterisks denote statistical significance (\* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq$   
 229  $0.001$ ). N represents the total number of species included from the corresponding climate region. n  
 230 indicates the number of species for which data are available for the corresponding trait.

| Climate Region: Temperate<br>(N = 744)   |                       |                  |            |                 |           |                  |
|--|-----------------------|------------------|------------|-----------------|-----------|------------------|
| Trait                                    | Critically endangered |                  | Endangered |                 | Invasive  |                  |
|  | Predicted             | Observed         | Predicted  | Observed        | Predicted | Observed         |
| Generation time<br>(n = 567)             | ↑                     | ↑ <sup>***</sup> | ↑          | —               | ↓         | ↑ <sup>**</sup>  |
| Trophic level<br>(n = 725)               | ↓                     | —                | ↓          | —               | —         | —                |
| Longevity<br>(n = 596)                   | ↑                     | ↑ <sup>***</sup> | ↑          | ↑ <sup>**</sup> | ↓         | ↑ <sup>***</sup> |
| Preferred temperature range<br>(n = 309) | ↓                     | —                | ↓          | —               | ↑         | ↑ <sup>**</sup>  |

| Climate Region: Subtropical<br>(N = 800) |                       |                  |            |                  |           |                  |
|--|-----------------------|------------------|------------|------------------|-----------|------------------|
| Trait                                    | Critically endangered |                  | Endangered |                  | Invasive  |                  |
|  | Predicted             | Observed         | Predicted  | Observed         | Predicted | Observed         |
| Generation time<br>(n = 518)             | ↑                     | ↑ <sup>***</sup> | ↑          | ↑ <sup>*</sup>   | ↓         | ↑ <sup>***</sup> |
| Trophic level<br>(n = 779)               | ↓                     | —                | ↓          | ↓ <sup>*</sup>   | —         | —                |
| Longevity<br>(n = 545)                   | ↑                     | ↑ <sup>***</sup> | ↑          | ↑ <sup>***</sup> | ↓         | ↑ <sup>***</sup> |
| Preferred temperature range<br>(n = 342) | ↓                     | —                | ↓          | —                | ↑         | ↑ <sup>***</sup> |

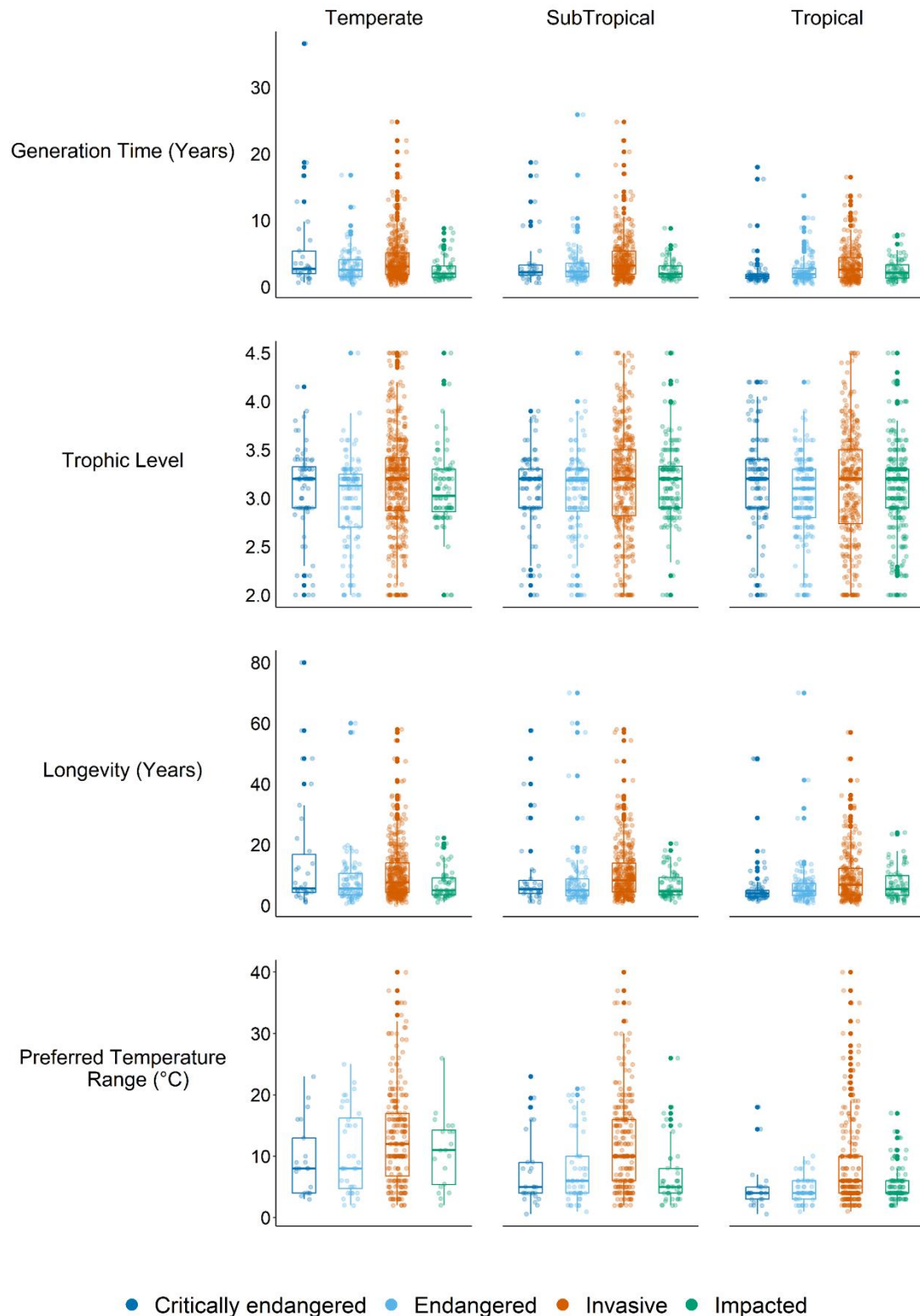
| Climate Region: Tropical<br>(N = 974)    |                       |                 |            |          |           |                  |
|--|-----------------------|-----------------|------------|----------|-----------|------------------|
| Trait                                    | Critically endangered |                 | Endangered |          | Invasive  |                  |
|  | Predicted             | Observed        | Predicted  | Observed | Predicted | Observed         |
| Generation time<br>(n = 519)             | ↑                     | —               | ↑          | —        | ↓         | ↑ <sup>***</sup> |
| Trophic level<br>(n = 935)               | ↓                     | ↑ <sup>*</sup>  | ↓          | —        | —         | —                |
| Longevity<br>(n = 531)                   | ↑                     | ↓ <sup>**</sup> | ↑          | —        | ↓         | ↑ <sup>***</sup> |
| Preferred temperature range<br>(n = 423) | ↓                     | —               | ↓          | —        | ↑         | ↑ <sup>***</sup> |

231

232

233

234



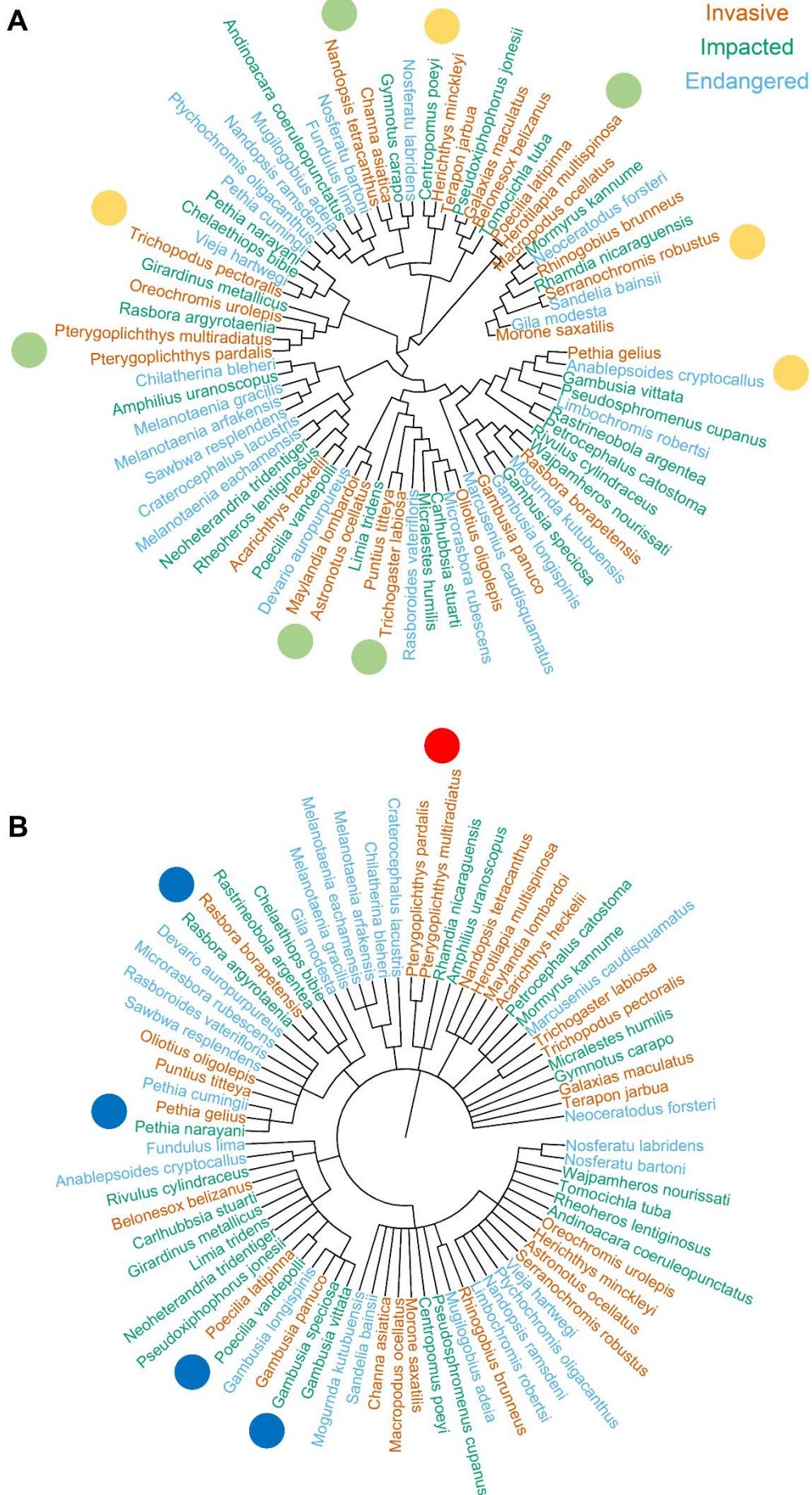
235

236 **Figure 2.** Box and scatter plots showing observed values for generation time (years) (n = 1604), trophic  
237 level (n = 2439), longevity (years) (n = 1672), and preferred temperature range (°C) (n = 1074) for  
238 invasive species, those species they endanger, critically endangered, and species impacted by invasives  
239 but not considered to be at risk of extinction (impacted; control group). Species that occupy multiple  
240 climate regions are represented in multiple counts.

241 ***Multiple traits in determining invasive success***

242 Following a divisive cluster analysis based on three traits, some clusters were found to be  
243 composed exclusively by invasive species suggesting that combined traits may be linked to  
244 invasive success (Figure 3A, green circles). However, other clusters contained invasive,  
245 impacted and/or endangered species (Figure 3A, yellow circles). Accordingly, multiple  
246 invasive species have greater similarity to impacted and endangered species as opposed to  
247 other invasives. Following phylogenetic analyses of the same species, we identified genera  
248 that contain a mix of invasive, impacted and/or endangered species (Figure 3B, blue circles).  
249 However, a genera (*Pterygoplichthys*) containing only invasive species was also identified  
250 (Figure 3B, red circle).

251





252 **Figure 3.** (A) Hierarchical cluster (using the Divisive ANALysis [DIANA] clustering algorithm) based on  
253 trait values for age of sexual maturity, trophic level, and temperature range for a random subset of 75  
254 tropical species in the ITOFF database (25 invasive species, 25 species endangered due to the spread  
255 of invasive species, and 25 species impacted by invasives but not endangered [impacted]). Green  
256 circles indicate clusters composed of exclusively invasive species. Yellow circles indicate clusters  
257 containing invasive, impacted and/or endangered species. (B) Taxonomic tree constructed for the same  
258 species pool used in (A). Red circles indicate genera containing only invasive species. Blue circles  
259 indicate genera containing invasive, impacted and/or endangered species.

260

## 261 **DISCUSSION**

262 ITOFF offers a unique and accessible platform to comprehensively address key questions in  
263 invasion biology. ITOFF data provides strong evidence that classical predictions for life-  
264 history traits facilitating invasive success do not hold for freshwater fish on the major climate  
265 zone scale (Capellini *et al.*, 2015; Chapple, Simmonds, and Wong, 2012; Crowder and  
266 Snyder, 2010). While invasive freshwater fish species are characterized by greater preferred  
267 temperature ranges, they cannot be reliably identified by generation time, trophic level,  
268 longevity, or multiple trait analyses. It is therefore essential that we shift the focus of invasive  
269 species management towards a holistic approach that considers the species (Chapple,  
270 Simmonds, and Wong, 2012), environment (Pyšek *et al.*, 2020), and evolutionary history  
271 (Mazzamuto *et al.*, 2016).

## 272 **ITOFF Database**

273 The ITOFF database reveals key gaps in data availability amongst invasive species and  
274 those species they impact. For example, there is data inequality amongst non-native  
275 species. Specifically, trait data are unavailable for many failed invaders. Nonetheless,  
276 comparisons between non-native groups (i.e. invasive, and failed invaders) will allow the  
277 identification of traits that facilitate invasive success during one or more stages of the  
278 invasion process (Capellini *et al.*, 2015). Future research should target both successful and  
279 unsuccessful invasive freshwater fish species. Additionally, by using ITOFF it was possible  
280 to identify key data gaps concerning the distribution of invasive species across climate

281 regions. One of the major factors affecting the spread of non-native species is tropicalization  
282 (Osland *et al.*, 2021). High latitude regions are therefore hot spots for the spread of non-  
283 native species (Goldsmith *et al.*, 2020; Hughes *et al.*, 2020). However, data on freshwater  
284 invasions in these regions are limited relative to tropical, subtropical and temperate regions.  
285 It is therefore critical that invasive species research is distributed evenly across climate  
286 zones.

287 Another unique addition provided by ITOFF to invasion biology research is the visualization  
288 of impacts associated with invasive species. This is an accessible tool with applications for  
289 academic research, citizen science, and education. Networks highlight the complexity of  
290 impacts associated with invasive species and can be used to guide invasive species  
291 management and identify target study species.

### 292 ***Linking life-history traits and invasive success***

293 Collectively, the results of our trait analyses provide weak evidence to support the classical  
294 predictions that life-history traits promote invasive success (Capellini *et al.*, 2015; Chapple,  
295 Simmonds, and Wong, 2012; Crowder and Snyder, 2010). The influence of environmental  
296 variability and propagule pressure on invasive success may explain the failure to detect life-  
297 history traits that ubiquitously promote invasive success (Moyle and Light, 1996; Tabak, Webb,  
298 and Miller, 2018). However, a key finding of this study is the greater preferred temperature  
299 ranges for invasive species across temperate, tropical, and subtropical climate zones. This  
300 result confirms predictions that broader abiotic tolerances increase the likelihood of invasive  
301 success by facilitating survival in a greater range of novel environments (Zerebecki and Sorte,  
302 2011; Bates *et al.*, 2013). Nonetheless, it should be noted that there is large overlap in trait  
303 values amongst all treatment groups for all traits, including preferred temperature range  
304 (Figure 2).

## 305 ***Multiple traits in determining invasive success***

306 The confounding results of the hierarchical cluster and phylogenetic analyses stress the  
307 complexity of the invasion process and suggest that both life-history traits and context are key  
308 in determining invasive success. These results prompt discussion into the role of ancestral  
309 traits in determining invasive success and highlight the need for targeted research of closely  
310 related species that differ in invasive status.

311 Overlap of trait values between invasive species and those species they impact limits the  
312 successful identification of potential invasive species on the climatic zone or global scale. This  
313 key result is in line with previous conceptual research which predicted that ubiquitous invasive  
314 traits are unlikely to be identified (Moyle and Light, 1996). Universal invasive species  
315 management is likely to be less efficient than localized management strategies and we advise  
316 that this be reflected in policy. Successful prediction of invasion on reduced spatial scales is  
317 possible with detailed knowledge of the biotic and abiotic conditions of an ecosystem (Kolar  
318 and Lodge, 2002; Marchetti, Moyle and Levine, 2004; Vila-gispert, Alcaraz and Garcia-  
319 Berthou, 2005). However, local scale approaches to invasive species management are work  
320 intensive. Nonetheless, as current ability to forecast aquatic species invasions on the large  
321 spatial scale remains limited, localised management remains the most viable approach (Moyle  
322 and Light, 1996).

## 323 ***Future directions***

324 The accurate prediction of future invasive species is a central component of governmental  
325 non-native species management. In the United Kingdom, 'The Great Britain Invasive Non-  
326 native Species Strategy' (Department for Environment Food and Rural Affairs [DEFRA], 2015)  
327 is used as a strategic framework for current management (e.g. 'River basin management  
328 plans'; Environment Agency, 2022). Horizon scanning is the only predictive management  
329 measure named in the DEFRA framework, and is largely reliant on the use of 'biological and  
330 ecological criteria' to identify species with a high-risk of becoming invasive in the UK. The

331 results of this study demonstrate that predictive frameworks must consider ecosystem  
332 characteristics, evolutionary history and contemporary evolution in combination with life-  
333 history traits for improved accuracy of invasive species forecasts. We argue that urgent  
334 research concerning the relationship of life-history traits, evolutionary history, rapid adaptation  
335 and invasive success is needed. ITOFF opens the opportunity to test the above avenues of  
336 research, which ultimately will further our understanding of the factors contributing to  
337 successful invasion in freshwater fish, a model taxon for the study of invasion biology.

### 338 ***Data Availability Statement***

339 The ITOFF core dataset is available at <https://doi.org/10.5281/zenodo.10135093>  
340

### 341 ***Acknowledgements***

342 The authors wish to acknowledge the U.S. Geological Survey, the Invasive Species  
343 Specialist Group, the IUCN Red List of Threatened Species, FishBase, and CABI for the  
344 provision of data that have made the ITOFF project possible. M.B. is supported by  
345 a FCT/MCTES (Fundação para a Ciência e Tecnologia/Ministério da Ciência, Tecnologia e  
346 Ensino Superior) postdoctoral fellowship ([SFRH/BPD/82259/2011](https://doi.org/10.5281/zenodo.10135093)). M.B. also acknowledges  
347 the financial support provided by CESAM ([UIDP/50017/2020+UIDB/50017/2020](https://doi.org/10.5281/zenodo.10135093)).

## REFERENCES

- Amiel, J. J., Tingley, R. and Shine, R. (2011) 'Smart moves: Effects of relative Brain size on Establishment success of invasive amphibians and reptiles', *PLoS ONE*, 6(4), pp. 4–7. doi: 10.1371/journal.pone.0018277.
- Andersen, M. C. *et al.* (2004) 'Risk assessment for invasive species', *Risk Analysis*, 24(4).
- Bates, A. E. *et al.* (2013) 'Geographical range, heat tolerance and invasion success in aquatic species', *Proceedings of the Royal Society B: Biological Sciences*, 280(1772). doi: 10.1098/rspb.2013.1958.
- Bernery, C., Bellard, C. A., Courchamp, F., Brosse, S., Gozlan, R., Jaric, I., Teletchea, F., and Leroy, B. (2022) 'Freshwater fish invasions: A comprehensive review', *Annual Review of Ecology, Evolution, and Systematics*, 53, pp. 427-456 doi: 10.1146/annurev-ecolsys-032522-015551
- CABI (2022) *Invasive Species Compendium*. Wallingford, UK: CAB International. [www.cabi.org/isc](http://www.cabi.org/isc).
- Capellini, I. *et al.* (2015) 'The role of life-history traits in mammalian invasion success', *Ecology Letters*, 18(10), pp. 1099–1107. doi: 10.1111/ele.12493.
- Chapple, D. G., Simmonds, S. M. and Wong, B. B. M. (2012) 'Can behavioral and personality traits influence the success of unintentional species introductions?', *Trends in Ecology & Evolution*, 27(1), pp. 57–64. doi: 10.1016/j.tree.2011.09.010.
- Crowder, D. W. and Snyder, W. E. (2010) 'Eating their way to the top? Mechanisms underlying the success of invasive insect generalist predators', *Biological Invasions*, 12(9), pp. 2857–2876. doi: 10.1007/s10530-010-9733-8.
- Deacon, A. E., Ramnarine, I. W. and Magurran, A. E. (2011) 'How Reproductive Ecology Contributes to the Spread of a Globally Invasive Fish', *PLoS ONE*, 6(9), pp. 1–7. doi: 10.1371/journal.pone.0024416.

Drake, J. M. (2007) 'Parental investment and fecundity, but not brain size, are associated with establishment success in introduced fishes', *Functional Ecology*, 21(5), pp. 963–968. doi: 10.1111/j.1365-2435.2007.01318.x.

Garcia-Berthou, E. (2007) 'The characteristics of invasive fishes: what has been learned so far?', *Journal of Fish Biology*, 71, pp. 33–55. doi: 10.1111/j.1095-8649.2007.01668.x.

Froese, R. and Pauly, D. Editors. (2022) FishBase, World Wide Web electronic publication. [www.fishbase.org](http://www.fishbase.org)

Goldsmith, J., McKindsey, C. W., Schlegel, R. W., Stewart, D. B., Archambault, P., and Howland, K. L. (2020) 'What and where? Predicting invasion hotspots in the Arctic marine realm', *Global Change Biology*, 26, pp. 4752–4771. doi: <https://doi.org/10.1111/gcb.15159>

Grabowski, M., Bacela, K. and Konopacka, A. (2007) 'How to be an invasive gammarid (Amphipoda: Gammaroidea) – comparison of life-history traits', *Hydrobiologia*, 590, pp. 75–84. doi: 10.1007/s10750-007-0759-6.

Heikkila, J. and Peltola, J. (2004) 'Analysis of the Colorado potato beetle protection system in Finland', *Agricultural Economics*, 31, pp. 343–352. doi: 10.1016/j.agecon.2004.09.019.

Hoffmann, B. D. and Broadhurst, L. M. (2016) 'The economic cost of managing invasive species in Australia', *NeoBiota*, 31, pp. 1–18. doi: 10.3897/neobiota.31.6960.

Holway, D. A. and Suarez, A. V (1999) 'Animal behavior: an essential component of invasion biology', *Tree*, 14(8), pp. 328–330. doi: [https://doi.org/10.1016/S0169-5347\(99\)01636-5](https://doi.org/10.1016/S0169-5347(99)01636-5).

Hughes, KA, Pescott, OL, Peyton, J, Adriaens, T., Cottier-Cook, E. J., Key, G., Rabitsch, W., Tricarico, E., Barnes, D. K. A., Baxter, N., Belchier, M., Blake, D. Convey, P., Dawson, W., Frohlich, D., Gardiner, L. M., González-Moreno, P., James, R., Malumphy, C., Martin, S., Martinou, A. F., Minchin, D., Monaco, A., Moore, N., Morley, S. A., Ross, K., Shanklin, J., Turvey, K., Vaughan, D., Vaux, A. G. C., Werenkraut, V., Winfield, I. J., Roy, H. E. (2020) 'Invasive non-native species likely to threaten biodiversity

and ecosystems in the Antarctic Peninsula region', *Global Change Biology*, 26, pp. 2702–2716. doi: <https://doi.org/10.1111/gcb.14938>

Jessop, A., Michalopoulou, A., Coonan, C., Mazzei, L., Sutherland O'Brien, E., Brady, G., Davison, C., Gourlay, W., Henderson, E., Lornie, A., McCloskey, E., Ramsay, H., Wilson, S., Shimadzu, H., and Barbosa, M. (2023). Invasive traits of freshwater fish database (ITOFF) (Version 5) [Data set] Zenodo. <https://doi.org/10.5281/zenodo.10135093>

Keith, J. M. and Spring, D. (2013) 'Agent-based Bayesian approach to monitoring the progress of invasive species eradication programs', *PNAS*, 110(33), pp. 1–6. doi: [10.1073/pnas.1216146110](https://doi.org/10.1073/pnas.1216146110) /DCSupplemental.[www.pnas.org/cgi/doi/10.1073/pnas.1216146110](http://www.pnas.org/cgi/doi/10.1073/pnas.1216146110).

Invasive Species Specialist Group (ISSG) (2000) *Global Invasive Species Database (GISD)*, <http://www.iucngisd.org/gisd/>

Invasive Species Specialist Group (ISSG) (2021) *Aquatic Species Extraction*, dataset, unpublished.

IUCN (2022) *The IUCN Red List of Threatened Species, Version 2022-2*. <https://www.iucnredlist.org>.

Manchester, S. J. and Bullock, J. M. (2000) 'The impacts of non-native species on UK biodiversity and the effectiveness of control', *Journal of Applied Ecology*, 37(5), pp. 845–864.

Marchetti, M. P., Moyle, P. B. and Levine, R. (2004) 'Invasive species profiling? Exploring the characteristics of non-native fishes across invasion stages in California', *Freshwater Biology*, 49(5), pp. 646–661. doi: [10.1111/j.1365-2427.2004.01202.x](https://doi.org/10.1111/j.1365-2427.2004.01202.x).

Mazzamuto, M.V., Galimberti, A., Cremonesi, G., Pisanu, B., Chapuis, J.-L., Stuyck, J., Amori, G., Su, H., Aloise, G., Preatoni, D.G., Wauters, L.A., Casiraghi, M. and Martinoli, A. (2016), 'Preventing species invasion: A role for integrative taxonomy?', *Integrative Zoology*, 11, pp. 214-228. doi: <https://doi.org/10.1111/1749-4877.12185>

Olson, L. J. (2006) 'The economics of terrestrial invasive species: a review of the literature', *Agricultural and Resource Economics Review*, 35(1), pp. 178–194.

Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., Langtimm, C. A., Williams, C. M., Keim, B. D., Terando, A. J., Reyier, E. A., Marshall, K. E., Loik, M. E., Boucek, R. E., Lewis, A. B. and Seminoff, J. A. (2021), 'Tropicalization of temperate ecosystems in North America: The northward range expansion of tropical organisms in response to warming winter temperatures', *Global Change Biology*, 27, pp. 3009-3034. doi: <https://doi.org/10.1111/gcb.15563>

Pimentel, D., Zuniga, R. and Morrison, D. (2004) 'Update on the environmental and economic costs associated with alien-invasive species in the United States', *Ecological Economics*, 52, pp. 273–288. doi: [10.1016/j.ecolecon.2004.10.002](https://doi.org/10.1016/j.ecolecon.2004.10.002).

Pimm, S. (1991) *The balance of nature?* Chicago: University of Chicago Press.

Pyšek, P., Hulme, P.E., Simberloff, D., Bacher, S., Blackburn, T.M., Carlton, J.T., Dawson, W., Essl, F., Foxcroft, L.C., Genovesi, P., Jeschke, J.M., Kühn, I., Liebhold, A.M., Mandrak, N.E., Meyerson, L.A., Pauchard, A., Pergl, J., Roy, H.E., Seebens, H., van Kleunen, M., Vilà, M., Wingfield, M.J. and Richardson, D.M. (2020), 'Scientists' warning on invasive alien species'. *Biol Rev*, 95, pp. 1511-1534. doi: <https://doi.org/10.1111/brv.12627>

Reid, A. J. *et al.* (2019) 'Emerging threats and persistent conservation challenges for freshwater biodiversity', *Biological Reviews*, 94(3), pp. 849–873. doi: [10.1111/brv.12480](https://doi.org/10.1111/brv.12480).

Rejmánek, M. *et al.* (2002) 'Biological invasions: politics and the discontinuity of ecological terminology', *Bulletin of the Ecological Society of America*, 83(2), pp. 131–133. doi: [10.2307/20168698](https://doi.org/10.2307/20168698).

Rosenthal, W. C. *et al.* (2021) 'Invasion and rapid adaptation of guppies (*Poecilia reticulata*) across the Hawaiian Archipelago', *Evolutionary Applications*, 14(7), pp. 1747–1761. doi: [10.1111/eva.13236](https://doi.org/10.1111/eva.13236).

Sakai, A. K. *et al.* (2001) 'The population biology of invasive species', *Annual Review of*



*Ecology and Systematics*, 32, pp. 305–332.

Sol, D. *et al.* (2008) 'Brain size predicts the success of mammal species introduced into novel environments', *American Naturalist*, 172(S1), pp. 63–71. doi: 10.1086/588304.

Sol, D. *et al.* (2012) 'Unraveling the Life history of successful invaders', *Science*, 337(6094), pp. 580–583.

Suarez, A. V. and Tsutsui, N. D. (2008) 'The evolutionary consequences of biological invasions', *Molecular Ecology*, 17(1), pp. 351–360. doi: 10.1111/j.1365-294X.2007.03456.x.

Tabak, M. A., Webb, C. T. and Miller, R. S. (2018) 'Propagule size and structure, life history, and environmental conditions affect establishment success of an invasive species' *Scientific Reports*, 8. doi: <https://doi.org/10.1038/s41598-018-28654-w>

Ulman, A. *et al.* (2021) 'Low pufferfish and lionfish predation in their native and invaded ranges suggests human control mechanisms may be necessary to control their Mediterranean abundances', *Frontiers in Marine Science*, 8, pp. 1–13. doi: 10.3389/fmars.2021.670413.

U.S. Geological Survey (USGS) (2022) *Nonindigenous Aquatic Species Database*, Gainesville, FL. <http://nas.er.usgs.gov>

Vall-Llosera, M. and Sol, D. (2009) 'A global risk assessment for the success of bird introductions', *Journal of Applied Ecology*, 46(4), pp. 787–795. doi: 10.1111/j.1365-2664.2009.01674.x.

Van Kleunen, M., Weber, E. and Fischer, M. (2010) 'A meta-analysis of trait differences between invasive and non-invasive plant species', *Ecology Letters*, 13(2), pp. 235–245. doi: 10.1111/j.1461-0248.2009.01418.x.

Vila-gispert, A., Alcaraz, C. and Garcia-Berthou, E. (2005) 'Life-history traits of invasive fish in small Mediterranean streams', *Biological Invasions*, 7(1), pp. 107–116. doi: 10.1007/s10530-004-9640-y.

Winker, H. *et al.* (2011) 'Life history and population dynamics of invasive common carp, *Cyprinus carpio*, within a large turbid African impoundment.', *Marine and Freshwater Research*, 62, pp. 1270–1280. doi: <https://doi.org/10.1071/MF11054>.

Wong, J. W. Y., Meunier, J. and Kölliker, M. (2013) 'The evolution of parental care in insects: The roles of ecology, life history and the social environment', *Ecological Entomology*, 38(2), pp. 123–137. doi: 10.1111/een.12000.

Wourms, J. P. and Lombardi, J. (1992) 'Reflections on the evolution of piscine viviparity', *American Zoologist*, 32(2), pp. 276–293. doi: 10.1093/icb/32.2.276.

WWF (2020) *Living Planet Report 2020 - Bending the curve of biodiversity loss*, *Living Planet Report 2020*.

Zerebecki, R. A. and Sorte, C. J. B. (2011) 'Temperature tolerance and stress proteins as mechanisms of invasive species success', *PLoS ONE*, 6(4). doi: 10.1371/journal.pone.0014806.

## **APPENDIX**

### ***ITOFF usage notes***

The core ITOFF dataset can be downloaded from the Zenodo public repository (<https://doi.org/10.5281/zenodo.10135093>) and should be referenced by citing the present paper. Exemplary R code and a document describing the variables included in the core ITOFF dataset are also available via the Zenodo public repository. Additionally, a dynamic, continuously updated, version of the ITOFF database can be downloaded from the ITOFF project website (<https://itoff-dataset.wp.st-andrews.ac.uk>). Data are organised such that the dataset is easy to use and serves as a readily available tool for invasion biology research. The web application can be accessed via the ITOFF project website (<https://itoff-dataset.wp.st-andrews.ac.uk/webapplication/>). Guidance on how to use the web application is also provided on the project website.

The primary intended use of the dataset is the identification of shared invasive traits amongst freshwater fish. Traits can be identified through broad-scale comparisons made between invasive, critically endangered, endangered, and impacted species. Once identified, these traits will improve current ability to forecast species invasions and highlight ‘high-risk’ species with great invasive potential. The same data offer insights into traits that may increase a native species’ susceptibility to invasion and can therefore be used to inform policy, conservation, and invasive species management. Future studies can address differences in invasive traits between climate regions. This research is of great importance given the current knowledge gap concerning how climate change will impact the outcome of future species invasions. Additionally, using the provided taxonomic information, future studies can untangle the role of adaptation, chance, and evolutionary history in the invasion process (Figure 1A, Appendix). Further, cluster analyses can investigate similarity across multiple traits for invasive, endangered, and impacted species (Figure 1B, Appendix). Cluster analyses will be important to invasive trait identification as invasive propensity is not always tied to individual traits but arises from a combination of multiple traits (Grabowski, Bacela and Konopacka, 2007).

**Table A1.** Descriptions of the life-history traits included in the ITOFF database and evidence for their link to invasive success.

| Trait  | Description   | Link to invasive success  | References   |
|--|---|---|--|
| <b>Reproductive guild (included as mode of reproduction, mode of fertilization, and extent of parental care)</b> | Species are categorised as non-guarders, guardians, and bearers. Bearers include livebearers, mouthbrooders, or any species that utilises a brood pouch. Fish that actively defend externally laid eggs are guardians. Non-guarders include egg scatterers and any species for which neither parent actively defends eggs. Additional information is provided in a supplementary column (i.e. mouth | Greater parental care promotes colonisation success, especially in harsh environments. Further, establishment success was correlated with greater parental investment in Californian fishes and a global study using 14 fish species. Further, live birth is linked to increased offspring survival, colonisation success, and is advantageous when mating opportunities are limited (i.e. when | (Wourms and Lombardi, 1992; Marchetti, Moyle and Levine, 2004; Drake, 2007; Deacon, Ramnarine, and Magurran, 2011; Wong, Meunier and Kölliker, 2013) |

|                                       |   |   |   |
|---------------------------------------|---|---|---|
|                                       | brooding, nesting, internal live-bearing).  | population sizes are limited following introduction).   |   |
| <b>Absolute fecundity</b>             | Total number of eggs in an average female (for egg-layers), or the average, total number of offspring produced over the entire lifetime of livebearers.   | Greater reproductive output is positively associated with mammalian invasion success on the global scale  | (Capellini <i>et al.</i> , 2015)  |
| <b>Brood size</b>                     | The number of offspring produced, or the number of eggs laid at one discrete time   | Production of larger broods reduces the risk of extinction for small populations and is linked to non-native mammalian establishment success  | (Pimm, 1991; Capellini <i>et al.</i> , 2015)                                  |
| <b>Age of sexual maturity</b>         | The average age (years) at which an individual is able to successfully reproduce. Provided as a mean value across sex in the ITOFF database   | Early maturity is linked to rapid population growth, establishment success, and improved resistance to demographic stochasticity and anthropogenic management strategies in the invasive common carp, <i>Cyprinus carpio</i> . Conversely, non-native fish species of Catalan streams are characterised by late maturity relative to native species | (Vila-gispert, Alcaraz and Garcia-Berthou, 2005; Winker <i>et al.</i> , 2011) |
| <b>Generation time</b>                | The average age (years) of parents at the time their first offspring are born   | Short generation times promote rapid population growth, colonisation success, and facilitates rapid adaptation  | (Rosenthal <i>et al.</i> , 2021)  |
| <b>Trophic level and feeding type</b> | Trophic level is a value estimated by FishBase to show the 'rank of a species in a food web'. Defined as 1 + the mean trophic level of the focal species' food items (Froese and Pauly, 2022). Under feeding type, fish are categorised by mode of feeding (e.g. grazer or predator). | Generalist diets allow for behavioural flexibility in novel environments and increase the competitive ability of non-native species   | (Crowder and Snyder, 2010)  |

|  |  |   |  |
|--|--|---|--|
| <p><b>Relative brain weight</b></p>  | <p>Given as brain weight (mg) divided by body weight (g).</p>  | <p>Larger brains may facilitate greater behavioural flexibility and promote survival in novel environments. Brain size is correlated with invasive success in birds, mammals, amphibians, and reptiles</p>  | <p>(Sol <i>et al.</i>, 2008; Vall-Llosera and Sol, 2009; Amiel, Tingley and Shine, 2011)</p> |
| <p><b>Longevity (included as maximum age in years, maximum weight, and maximum length)</b></p> | <p>The maximum age, weight, or length that an individual of each species can reach. Reported as the mean value of the approximate age, weight, or length of the oldest known individuals amongst populations.</p>  | <p>Though life-history strategies are generally acknowledged to be an important factor in determining invasive success, there is debate as to whether or not invasive species are more often characterised by fast or slow life-histories. Fast life-histories are predicted to promote rapid population growth and establishment success. However, slow life-histories can also increase establishment success in a wider range of environmental conditions by prioritising survival and delaying reproduction</p> | <p>(Sol <i>et al.</i>, 2012)</p>   |
| <p><b>Temperature range (included as both preferred and tolerable)</b></p>                     | <p>Preferred temperature range: the range of temperatures over which the species can survive indefinitely. Reported in ITOFF as the range of water temperatures experienced throughout the species range.</p> <p>Tolerable temperature range: the range of temperatures over which individuals of the species can survive brief exposure</p> | <p>Wider abiotic tolerances increase the likelihood of survival within a wider range of novel environments</p>  | <p>(Zerebecki and Sorte, 2011; Bates <i>et al.</i>, 2013)</p>                                |

**Table A2.** Summary of the results for generalized linear models investigating differences in life-history trait values between invasive species, species impacted by invasive species but not facing an elevated risk of extinction (impacted), endangered species that are threatened by invasive species, and critically endangered species that are threatened by invasive species. Results are shown for four life-history traits shown to be critical to invasive success. Data are split between major climate regions. Species present in multiple climate regions are represented in multiple counts. N represents the total number of species included for the corresponding climate region. n indicates the number of species for which data are available for the corresponding trait. Overall model adjusted  $R^2$  is reported for models with Gaussian error distributions, McFadden's pseudo- $R^2$  is reported for models with Poisson error distributions.

| <b>Climate Region: Temperate</b><br>(N = 744)                                  |  |  |                                       |                                       |
|--|--|--|---------------------------------------|---------------------------------------|
| <b>Trait</b>   | <b>Impacted</b><br>(Reference group)   | <b>Critically endangered</b>           | <b>Endangered</b>                     | <b>Invasive</b>                       |
| <b>Generation time</b><br>(n = 567, McFadden's R <sup>2</sup> = 0.02)          | 2.71 ± 1.08<br>(z = 12.92, p < 0.001)  | 5.73 ± 1.11<br>(z = 7.06, p < 0.001)   | 3.24 ± 1.1<br>(z = 1.84, p = 0.07)    | 4.06 ± 1.08<br>(z = 4.97, p < 0.001)  |
| <b>Trophic level</b><br>(n = 725, R <sup>2</sup> = 0.02)                       | 3.08 ± 0.05<br>(t = 60.18, p < 0.001)  | 3.09 ± 0.08<br>(t = 0.05, p = 0.95)    | 2.97 ± 0.07<br>(t = -1.65, p = 0.1)   | 3.16 ± 0.06<br>(t = 1.42, p = 0.16)   |
| <b>Longevity</b><br>(n = 596, McFadden's R <sup>2</sup> = 0.03)                | 7.16 ± 1.05<br>(z = 41.8, p < 0.001)   | 15.18 ± 1.07<br>(z = 11.66, p < 0.001) | 8.38 ± 1.06<br>(z = 2.63, p = 0.009)  | 10.56 ± 1.05<br>(z = 7.87, p < 0.001) |
| <b>Preferred Temperature range</b> (n = 309, McFadden's R <sup>2</sup> = 0.02) | 10.85 ± 1.07<br>(z = 35.12, p < 0.001) | 9.62 ± 1.1<br>(z = -1.23, p = 0.22)    | 10.61 ± 1.09<br>(z = -0.26, p = 0.79) | 13.23 ± 1.07<br>(z = 2.83, p = 0.005) |

| <b>Climate Region: Subtropical</b><br>(N = 800)                                |                                       |                                       |                                      |                                       |
|--|---------------------------------------|---------------------------------------|--------------------------------------|---------------------------------------|
| <b>Trait</b>   | <b>Impacted</b><br>(Reference group)  | <b>Critically endangered</b>          | <b>Endangered</b>                    | <b>Invasive</b>                       |
| <b>Generation time</b><br>(n = 518, McFadden's R <sup>2</sup> = 0.02)          | 2.59 ± 1.08<br>(z = 12.62, p < 0.001) | 3.85 ± 1.12<br>(z = 3.45, p < 0.001)  | 3.14 ± 1.1<br>(z = 2.05, p = 0.04)   | 4.2 ± 1.08<br>(z = 6.03, p < 0.001)   |
| <b>Trophic level</b><br>(n = 779, R <sup>2</sup> = 0.01)                       | 3.16 ± 0.04<br>(t = 82.26, p < 0.001) | 3.09 ± 0.06<br>(t = -1.23, p = 0.22)  | 3.03 ± 0.06<br>(t = -2.39, p = 0.02) | 3.16 ± 0.05<br>(t = 0.18, p = 0.86)   |
| <b>Longevity</b><br>(n = 545, McFadden's R <sup>2</sup> = 0.02)                | 6.84 ± 1.05<br>(z = 41.46, p < 0.001) | 10.63 ± 1.07<br>(z = 6.34, p < 0.001) | 8.38 ± 1.06<br>(z = 3.53, p < 0.001) | 10.71 ± 1.05<br>(z = 9.11, p < 0.001) |
| <b>Preferred Temperature range</b> (n = 342, McFadden's R <sup>2</sup> = 0.06) | 6.98 ± 1.06<br>(z = 33.65, p < 0.001) | 7.57 ± 1.09<br>(z = 0.91, p = 0.36)   | 7.9 ± 1.08<br>(z = 1.63, p = 0.1)    | 12.2 ± 1.06<br>(z = 9.18, p < 0.001)  |

| <b>Climate Region: Tropical</b><br>(N = 974)                                   |                                       |                                       |                                      |                                      |
|--|---------------------------------------|---------------------------------------|--------------------------------------|--------------------------------------|
| <b>Trait</b>   | <b>Impacted</b><br>(Reference group)  | <b>Critically endangered</b>          | <b>Endangered</b>                    | <b>Invasive</b>                      |
| <b>Generation time</b><br>(n = 519, McFadden's R <sup>2</sup> = 0.02)          | 2.49 ± 1.08<br>(z = 12.49, p < 0.001) | 2.21 ± 1.11<br>(z = -1.15, p = 0.25)  | 2.54 ± 1.1<br>(z = 0.19, p = 0.85)   | 3.35 ± 1.08<br>(z = 3.65, p < 0.001) |
| <b>Trophic level</b><br>(n = 935, R <sup>2</sup> = 0.01)                       | 3.09 ± 0.03<br>(t = 99.95, p < 0.001) | 3.21 ± 0.05<br>(t = 2.29, p = 0.02)   | 3.02 ± 0.05<br>(t = -1.6, p = 0.11)  | 3.11 ± 0.04<br>(t = 0.38, p = 0.71)  |
| <b>Longevity</b><br>(n = 531, McFadden's R <sup>2</sup> = 0.03)                | 7.16 ± 1.04<br>(z = 45.57, p < 0.001) | 5.93 ± 1.07<br>(z = -2.96, p = 0.003) | 6.71 ± 1.06<br>(z = -1.18, p = 0.24) | 9.52 ± 1.05<br>(z = 5.98, p < 0.001) |
| <b>Preferred Temperature range</b> (n = 423, McFadden's R <sup>2</sup> = 0.06) | 5.26 ± 1.04<br>(z = 38.66, p < 0.001) | 4.79 ± 1.11<br>(z = -0.91, p = 0.36)  | 4.55 ± 1.08<br>(z = -1.86, p = 0.06) | 8.38 ± 1.05<br>(z = 9.63, p < 0.001) |

**ITOFF Supplementary R code.** R code for basic analysis of the ITOFF database. This script will be provided in unison with the ITOFF database following publication of the ITOFF data paper.

### This code is for exploration, preliminary analyses, and visualisation of the ITOFF core database

### This code should be used as a guide and adapted to your needs

### First you will need to download the ITOFF core dataset. Data will be available at:

### <https://itoff-dataset.wp.st-andrews.ac.uk/> and via the Zenodo public repository (in .csv format)

### Load packages:

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(ggpubr)
```

### You may need to install these packages using `install.packages("...")`

### if you have not done so previously

### Load the database using the file pathway as saved on your device:

```
data <- read.csv("...", header=T, sep="," )
```

### Ensuring the dataset has loaded correctly

```
head(data)
```

```
str(data)
```

### Specifying factor and continuous variables

```
data$Polar <-as.factor(data$Polar)
```

```
data$Boreal <-as.factor(data$Boreal)
```

```
data$Temperate <-as.factor(data$Temperate)
```

```
data$SubTropical <-as.factor(data$SubTropical)
```

```
data$Tropical <-as.factor(data$Tropical)
```

```
data$ITOFF.Group <-as.factor(data$ITOFF.Group)
```

```
data$System <-as.factor(data$System)
```

```
data$Status <- as.factor(data$Status)
```



```
data$IUCN_Status <-as.factor(data$IUCN_Status)
data$ReproMode<-as.factor(data$ReproMode)
data$ReproductiveGuild <-as.factor(data$ReproductiveGuild)
data$RepGuild1<-as.factor(data$RepGuild1)
data$RepGuild2<-as.factor(data$RepGuild2)
data$MeanFecundity <- as.numeric(data$MeanFecundity)
data$MinFecundity <- as.numeric(data$MinFecundity)
data$MaxFecundity <- as.numeric(data$MaxFecundity)
data$BroodSize <- as.numeric(data$BroodSize)
data$AgeMaturity_Years_Estimate<- as.numeric(data$AgeMaturity_Years_Estimate)
data$MeanAgeMaturity_Years<- as.numeric(data$MeanAgeMaturity_Years)
data$GenerationTime_Years_Estimate<- as.numeric(data$GenerationTime_Years_Estimate)
data$TrophicLevel_Estimate<- as.numeric(data$TrophicLevel_Estimate)
data$TrophicLevel<- as.numeric(data$TrophicLevel)
data$FeedingType <-as.factor(data$FeedingType)
data$Relative_BrainWeight<- as.numeric(data$Relative_BrainWeight)
data$Longevity_Years_Estimate<- as.numeric(data$Longevity_Years_Estimate)
data$Tmax<- as.numeric(data$Tmax)
data$Wmax<- as.numeric(data$Wmax)
data$Lmax<- as.numeric(data$Lmax)
data$TempPref_min<- as.numeric(data$TempPref_min)
data$TempPref_max<- as.numeric(data$TempPref_max)
data$TempRange<- as.numeric(data$TempRange)
data$TolerableTemp_min<- as.numeric(data$TolerableTemp_min)
data$TolerableTemp_max<- as.numeric(data$TolerableTemp_max)
data$Temp_Range_tolerable<- as.numeric(data$Temp_Range_tolerable)
```

```
###-----
```

```
### To see all variables currently included in the database:
```

```
names(data)
```

```
### To see the pool of species:
```

```
species <- unique(data$Species)
```

```
species
```

```
###-----
```

```
### Extracting data - To create a new, derivative dataframe you can subset
```

```
### the data by categorical variables. For example, to create a dataframe comprised of only temperate
```

```
### species:
```

```
temperate <- subset(data, Temperate == "1")
```

```
head(temperate)
```

```
### To export this dataframe as a new csv file [ADD YOUR OWN CUSTOM PATH]
```

```
write.csv(temperate, "ITOFF_Temperate_species.csv")
```

```
###-----
```

```
### Data visualisation. Example 1.
```

```
### Plot of log(Absolute Fecundity) for groups of invasive, endangered, and impacted species
```

```
summary(data$Status)
```

```
### Creating a new variable for log(Absolute Fecundity)
```

```
data$logMeanFecundity<- log(data$MeanFecundity)
```

```
str(data)
```

```
### Creating a dataframe to plot from
```

```
invasive <- subset(data, Status=="Invasive")
```

```
EN <- subset(data, Status=="EN")
```

```
CR <- subset(data, Status=="CR")
```

```
LC <- subset(data, Status=="LC")
```

```
plottingdata<-rbind(invasive,EN,CR,LC)
```

```
### For simplicity we will focus only on species found in SubTropical, Temperate
```

```
### and Tropical climate zones:
```

```
plottingdata_temp <- subset(plottingdata , Temperate == "1")
```

```
plottingdata_subtrop <- subset(plottingdata , SubTropical == "1")  
plottingdata_trop <- subset(plottingdata , Tropical == "1")  
plottingdata <- rbind(plottingdata_temp, plottingdata_subtrop, plottingdata_trop)
```

### Creating the plot

```
FecundityPlot <- ggboxplot(plottingdata, x = "Status", y = "logMeanFecundity",  
  color = "Status", palette = "viridis",  
  ylab = "log(Mean Fecundity)", xlab = "",  
  add = "jitter",  
  order = c("LC", "EN", "CR", "Invasive")) +  
  scale_x_discrete(labels=c('Impacted', 'Endangered',  
    'Critically Endangered', 'Invasive'))
```

FecundityPlot

### Example 2. Plot of Generation Time (years) for groups of non-native, endangered,  
### and impacted species

### Creating the plot

```
GenerationTimePlot <- ggboxplot(plottingdata, x = "Status", y = "GenerationTime_Years_Estimate",  
  color = "Status", palette = "viridis",  
  ylab = "Generation \n Time \n (Years)", xlab = "",  
  add = "jitter",  
  order = c("LC", "EN", "CR", "Invasive")) +  
  scale_x_discrete(labels=c('Impacted', 'Endangered',  
    'Critically Endangered', 'Invasive'))
```

GenerationTimePlot

### To show these plots simultaneously:

```
ggarrange(FecundityPlot, GenerationTimePlot,  
  ncol = 1, nrow = 2, common.legend=TRUE)
```

```
#### To also show differences by climate region: Subsetting by climate region
#### First need to create a new dataset with Climate Region in long format:
temper <- subset(plottingdata, Temperate=="1")
subtrop <- subset(plottingdata, SubTropical=="1")
trop <- subset(plottingdata, Tropical=="1")

temper$ClimateRegion<- "Temperate"
subtrop$ClimateRegion<- "SubTropical"
trop$ClimateRegion<- "Tropical"

#### Merge
newplottingdata <- rbind(temper,subtrop,trop)
str(newplottingdata)

#### Plot for absolute fecundity as shown for each climate region
FecundityPlot2<- ggboxplot(newplottingdata, x = "Status", y = "logMeanFecundity",
  color = "Status", palette = "viridis",
  ylab = "log(Mean Fecundity)", xlab = "",
  add = "jitter",
  order = c("LC", "EN", "CR", "Invasive")) +
  scale_x_discrete(labels=c('Impacted', 'Endangered',
    'Critically Endangered', 'Invasive'))

FecundityPlot2
FecundityPlot2 <- facet(FecundityPlot2, facet.by="ClimateRegion", ncol=3)
FecundityPlot2

####-----
#### CLUSTER AND TAXONOMIC ANALYSES

#### Load required packages [NOTE THESE MAY HAVE TO BE INSTALLED FIRST]:
```

```
library(ape)
library(Rcpp)
library(ggtree)
library(cluster)
```

```
### Ensure variables are specified as factors:
```

```
data$Class<-as.factor(data$Class)
data$Order<-as.factor(data$Order)
data$Family<-as.factor(data$Family)
data$Species<-as.factor(data$Species)
data$Genus<-as.factor(data$Genus)
data$Phylum<-as.factor(data$Phylum)
data$Kingdom<-as.factor(data$Kingdom)
```

```
### HEIRARCHICAL CLUSTER ANALYSIS
```

```
### We will base our cluster analysis on three traits: age of sexual maturity,
```

```
### trophic level, and preferred temperature range
```

```
### Creating a dataframe for age of sexual maturity, trophic level, and temperature range
```

```
data2 <- data.frame(data$Species,data$Kingdom,data$Phylum,data$Class,
                   data$Order,data$Family,data$Genus,data$System,
                   data$ITOFF.Group,data$Status,data$AgeMaturity_Years,
                   data$TrophicLevel,data$TempRange)
```

```
str(data2)
```

```
data2 <- na.omit(data2)
```

```
str(data2)
```

```
names(data2)[names(data2) == "data.Species"] <- "Species"
```

```
names(data2)[names(data2) == "data.AgeMaturity_Years"] <- "AgeMaturity_Years"
```

```
names(data2)[names(data2) == "data.TrophicLevel"] <- "TrophicLevel"
```

```
names(data2)[names(data2) == "data.TempRange"] <- "TempRange"
```

```
names(data2)[names(data2) == "data.Kingdom"] <- "Kingdom"  
names(data2)[names(data2) == "data.Phylum"] <- "Phylum"  
names(data2)[names(data2) == "data.Class"] <- "Class"  
names(data2)[names(data2) == "data.Order"] <- "Order"  
names(data2)[names(data2) == "data.Family"] <- "Family"  
names(data2)[names(data2) == "data.Genus"] <- "Genus"  
names(data2)[names(data2) == "data.System"] <- "System"  
names(data2)[names(data2) == "data.ITOFF.Group"] <- "ITOFF.Group"  
names(data2)[names(data2) == "data.Status"] <- "Status"
```

```
### From this dataframe we will select a random subset of 90 species  
### that we will perform our heirarchical cluster analysis on
```

```
### Create list of species
```

```
ids <- unique(data2$Species)
```

```
### Randomly draw IDs from the species pool
```

```
draw <- ids %>% sample(90)
```

```
### Create dataframe containing only the randomly selected species
```

```
DendrogramData <- data2[data2$Species %in% draw, ]
```

```
### Here we will create a separate dataframes that only contains variables
```

```
### of relevance to the cluster analysis
```

```
dropa <- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "System", "ITOFF.Group",  
          "Status")
```

```
clustdata <- DendrogramData[,!(names(DendrogramData) %in% dropa)]
```

```
str(clustdata)
```

```
### We will also reformat our data so that it is appropriate for the
```

```
### requirements of the cluster package
```

```
rownames(clustdata) <- clustdata$Species
head(clustdata)
rownames(clustdata)
dropb <- c("Species")
clustdata = clustdata[!(names(clustdata) %in% dropb)]
head(clustdata)

### Perform divisive clustering using DIANA (Divisive ANALysis clustering algorithm)
### Divisive clustering based on: Age at maturity, Trophic level, and Temp range
clustdata <- scale(clustdata)
HC <- diana(clustdata)
HC

### ILLUSTRATING
ClustTree <- ggtree(HC,layout="circular", branch.length='none') + geom_tiplab()
ClustTree

### We can also construct a TAXONOMIC TREE using the same species pool

### Specifying the variables on which we will base our tree
frm <- ~Kingdom/Phylum/Class/Order/Family/Genus/Species

### Constructing the tree using the 'DendrogramData' dataframe
phyloree <- as.phylo(frm, data = DendrogramData, collapse=FALSE)
phyloree$edge.length <- rep(1, nrow(phyloree$edge))
phyloree

### Plotting the taxonomic tree
phyloree2 <- ggtree(phyloree,layout="circular") + geom_tiplab()
phyloree2
```

### Please see <https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html>

### For details on how to customise the aesthetics of dendrograms using

### the ggtree package