1 **i. Title Page**

2 **Article Title**

3 Historical, human and environmental drivers of genetic diversity in the red swamp crayfish

4 (*Procambarus clarkii*) invading the Iberian Peninsula.

5 **The full names of the authors**

6 Lucía Acevedo-Limón[1*#], Francisco J. Oficialdegui[2#], Marta I. Sánchez[3], Miguel Clavero[1]

7 **Correspondence:**

8 Lucía Acevedo-Limón. Estación Biológica de Doñana– CSIC. Avenida Américo Vespucio 26,

9 41092, Sevilla, Spain. Telephone: +34954232340. Fax: +34954621125

10 Email: *luciaacevedo294@gmail.com

11 # These authors contributed equally to this work

12 **The author's institutional affiliations where the work was carried out, with a footnote for the**

13 **author's present address if different from where the work was carried out**

14 1. Department of Conservation Biology, Doñana Biological Station (EBD-CSIC), Seville, Spain.

15 2. Department of Wetland Ecology, Doñana Biological Station (EBD-CSIC), Seville, Spain.

16 3. Department of Plant Biology and Ecology, Faculty of Biology, University of Seville, Seville,

17 Spain.

18 **E-mails:**

19 Francisco J. Oficialdegui; oficialdegui@ebd.csic.es; ORCID: 0000-0001-6223-736X

20 Marta I. Sánchez; msanchez85@us.es; ORCID: 0000-0002-8349-5410

21 Miguel Clavero; miguelclavero@ebd.csic.es; ORCID: 0000-0002-5186-0153

22 **Keywords**

23 environmental suitability; genetic structure; human-mediated dispersal; invasive species;

24 microsatellite; multiple introduction;

25

ii. Summary

1. Patterns of genetic diversity in invasive populations can be modulated by a range of factors acting at different stages of the invasion process, including the genetic composition of the source population(s), the introduction history (e.g. propagule pressure), the environmental suitability of recipient areas and the features of secondary introductions.
2. The North-American red swamp crayfish, *Procambarus clarkii*, is one of the most widely introduced freshwater species worldwide. It was legally introduced into Spain twice, near the city of Badajoz in 1973 and in the Guadalquivir marshes in 1974. Thereafter the species rapidly colonized almost the entire Iberian Peninsula.
3. We used seven nuclear microsatellites to describe the genetic diversity and structure of 28 locations distributed across the Iberian Peninsula and to explain the expansion process of the red swamp crayfish. Additionally, we analysed the relationship between environmental suitability and genetic diversity of the studied locations.
4. The red swamp crayfish had a clear spatial genetic structure in the Iberian Peninsula, probably determined by the two independent introduction events in the 1970s, which produced two main clusters separated spatially, one of which was dominant in Portugal and the other in Spain.
5. The human-mediated dispersal process seemed to have involved invasion hubs, hosting highly genetically diverse areas and acting as sources for subsequent introductions. Genetic diversity also tended to be higher in more suitable environments across the Iberian Peninsula.
6. Our results showed that the complex and human-mediated expansion of the red swamp crayfish in the Iberian Peninsula has involved several long- and short-distance movements and that both ecological and anthropogenic factors have shaped the genetic diversity patterns resulting from this invasion process. Early detection of potential invasion hubs may help to halt multiple short-distance translocations and thus the rapid expansion of highly prolific invasive species over non-native areas.

51 **iii. Main text**

52 **Introduction**

53 Biological invasions are one of the main threats to biodiversity globally (Bellard, Cassey &
54 Blackburn, 2016). The intensification of global trade and human movements, as well as the increase
55 of activities such as aquaculture, pet trade or gardening, have led to an acceleration of the global-
56 scale exchange of biota (Hulme et al., 2008; Ricciardi, 2007), which is blurring the traditionally
57 described biogeographical barriers (Capinha et al., 2015). The number of species introduced outside
58 their native ranges has been increasing in last decades and is expected to keep growing (Seebens et
59 al., 2017). Only a fraction of the introduced species is able to establish self-sustained populations,
60 thrive and spread, and only a fraction among them causes biodiversity losses, disruptions of
61 ecosystem functioning and economic impacts (Walsh et al., 2016). Understanding why some
62 introduced species succeed and become invasive, while other fail, is a central topic in invasion
63 science (Blackburn & Duncan, 2001; Blackburn, Prowse, Lockwood & Cassey, 2013; Facon et al.,
64 2006).

65 The genetic diversity of introduced populations can influence their ability to adapt to novel
66 environments and, thus, determine their invasiveness (Lavergne & Molofsky, 2007; but see
67 Bossdorf, Richards & Pigliucci, 2008; Hawes et al., 2018). Biological invasions are a multistep
68 process often described as a series of stages (transport, introduction, establishment and spread)
69 separated by different barriers that can impede the progress of an invasion (Blackburn et al., 2011).
70 Overcoming each of these barriers can generate population bottlenecks and alter the genetic
71 diversity patterns in invasive populations (Hardesty et al., 2012; Okada, Lyle & Jasieniuk, 2009).
72 Different factors can modulate the intensity of population bottlenecks in each barrier of the invasion
73 process, including genetic diversity of the source population, propagule pressure, environmental
74 suitability of the reciepient area and/or the characteristics of secondary introductions. Genetic
75 admixture (hereafter admixture) occurs when multiple divergent genetic lineages come into contact

76 and interbreed, increasing the genetic diversity of a population, as can occur in the source
77 population of the native range before the transport stage (Dlugosch & Parker, 2008; Oficialdegui et
78 al., 2019; Rius & Darling, 2014; van Boheemen et al., 2017). During the introduction stage,
79 propagule pressure (i.e., number of introduction events, inoculum size or both) modulates resulting
80 genetic diversity patterns since more introduction events and/or a large number of introduced
81 individuals promote higher genetic diversity in the introduced population (Blackburn, Prowse,
82 Loockwood & Cassey, 2013; Drolet & Locke, 2016). During the establishment stage, biotic (i.e.
83 niche competition) and abiotic (i.e. environmental suitability) factors can affect the genetic diversity
84 of an introduced population through modulation of survival and its associated population bottleneck
85 (Banks et al., 2013; Ellegren & Galtier, 2016). As such, the environmental suitability refers to the
86 climatic and physiographic variables of the introduced range. During the spread stage, founding
87 events, involving all the previous cited modulators of genetic diversity, take place whenever a
88 secondary introduction occurs (i.e., the source population being itself introduced). Therefore, range
89 expansions are generally associated with decreasing genetic diversity (i.e., allelic richness and
90 expected heterozygosity) along the expansion front (Austerlitz, Jung-muller, Godelle & Gouyon,
91 1997; Excoffier, Foll, & Petit, 2009).
92 The red swamp crayfish (*Procambarus clarkii*), native to North-Eastern Mexico and South-
93 Central United States, has been broadly introduced around the world, to the point that it is present in
94 up to 40 countries of four continents (Oficialdegui, Sánchez & Clavero, 2020). It was intentionally
95 introduced to southern Spain in the early 1970s, through two independent shipments from Louisiana
96 (U.S.A.). Both introductions had legal authorisations and were motivated by the high
97 socioeconomic value that crayfish was attaining in Spain (Clavero, 2016). The first introduction
98 took place in Badajoz (Spain) in 1973, and involved the release of around 300 individuals, the
99 survivors of an original batch of 500 crayfish (Habsburgo-Lorena, 1978). One year later, a larger
100 batch (around 500 kg) was imported to the marsh area of the Lower Guadalquivir River (Puebla del

101 Río, Seville), although only 100 kg (around 6,500 individuals) survived. The red swamp crayfish
102 immediately established self-sustained and abundant populations in the initial introduction areas
103 and rapidly spread over the Iberian Peninsula (Gutiérrez-Yurrita et al., 1999; Oficialdegui, Sánchez
104 & Clavero, 2020), aided by both intrinsic traits (e.g., short life cycle, high fecundity, high
105 environmental tolerance; Geiger, Alcorlo, Baltanas & Montes, 2005) and by multiple (arguably,
106 thousands) and uncontrolled secondary introductions (Clavero, 2016; Oficialdegui et al., 2019).
107 Shortly after the introduction, by 1982, there were already reports of the red swamp crayfish in the
108 Tablas de Daimiel National Park and the Ebro Delta (some 320 and 730 km straight-line,
109 respectively, from the Lower Guadalquivir introduction site) (Clavero 2016). Once introduced and
110 established, this species often becomes dominant in the occupied freshwater habitats, producing
111 severe ecological impacts and losses of ecosystem services (Gherardi, 2006; Souty-Grosset et al.
112 2016).

113 In this study, we examine the spatial patterns of genetic diversity of the red swamp crayfish in
114 the Iberian Peninsula to test various hypotheses about the invasion history. Based on the analysis of
115 nuclear microsatellites, we aim to analyse the present-day genetic structure of the red swamp
116 crayfish, to then explore the drivers that may have modulated the dynamics of genetic diversity
117 during the invasion process. The main questions we address are: (i) is there a relationship between
118 the number and size of initial introduction events and the present-day genetic diversity?; (ii) how
119 was the pattern of spread of the swamp crayfish among the Iberian Peninsula?; (iii) is there a
120 relationship between environmental suitability and genetic diversity? We hypothesize that
121 populations originated from Lower Guadalquivir would have a higher genetic diversity than the
122 ones originated from Badajoz, as the inoculum size was around 20 times larger in the Lower
123 Guadalquivir. Besides, genetic patterns within the Iberian Peninsula would be mainly explained by
124 human-mediated dispersal, with a negligible influence of natural dispersal. We also consider two
125 dispersion processes: the jump-dispersal and the invasion hub scenario. The jump-dispersal scenario

126 assumes that the spread has occurred through successive small-scale secondary introductions,

127 supposing that genetic diversity would tend to diminish with increasing distances to the initial foci,

128 due to the accumulation of genetic bottlenecks at each secondary introduction. The invasion hub

129 scenario involves also large-scale translocations and relevant sources other than the initial foci (i.e.

130 the invasion hubs). It supposes that the high genetic diversity in the invasion hubs could enhance

131 genetic diversity in neighbouring introduced populations. Finally, we hypothesize that suitable

132 environmental conditions would reduce the intensity of population bottlenecks, so that crayfish

133 introduced in suitable areas would present higher genetic diversity than the ones in unsuitable areas.

134

135 Methods

136 *Sample collection, DNA extraction and microsatellite genotyping*

137 A total of 903 adult red swamp crayfish were collected from 28 locations distributed across the

138 Iberian Peninsula (Table 1; Fig. 1). A piece of abdominal muscle tissue was extracted from each

139 crayfish and stored in 96% ethanol at room temperature until subsequent analyses.

140 Total genomic DNA was extracted from approximately 10 mg of dried muscle tissue using a

141 modified DNA salt-extraction protocol (Aljanabi, 1997) containing NaCl 25 mM, Tris 12.5 mM

142 (pH 8.0), EDTA 12.5 mM (pH 8.0), 31.5 µL SDS 10%, 230 µL deionized water and Proteinase K.

143 After overnight incubation at 34 °C, DNA samples were extracted with a Tecan robot, Freedom Evo

144 model. Resulting DNA was diluted 1:10 and preserved at -20ºC for genotyping analyses. We

145 designed two multiplex PCRs for fragment analysis, with Mix 1 (PCSH0002, PCSH0006, PclG-17,

146 PclG-29) and Mix 2 (PCSH0038, PCSH0065, PclG-15, PclG-48) containing microsatellite loci

147 previously developed by Belfiore and May (2000) and Jiang et al. (2015). A multiplex polymerase

148 chain reaction (PCR) was performed on both Mix 1 and Mix 2 (Table S1). All PCR amplifications

149 were performed in 15 µL reactions containing 4 µL of template DNA, 3 µL buffer 5x PROMEGA,

150 2.5 mM dNTP, 25 mM $MgCl_2$, 2 µL of Primer Mix (forward primer endlabelled with $[^{32}P]\gamma$ ATP),

151 0.75 U Taq polymerase PROMEGA, and deionized water up to the final volume of 15 µL. The
152 thermocycling regime of the Mix 1 consisted of an initial denaturation step at 95 ºC for 3 min,
153 followed by 8 cycles of denaturing at 95 ºC for 30 s, annealing at 60 ºC (decreasing 1 ºC for each
154 cycle) for 30 s, and extension at 72 ºC for 30 s, followed by 23 cycles of denaturing at 95 ºC for 30
155 s, annealing at 52 ºC for 30 s and 72 ºC for 30 s with a final extension at 72 ºC for 10 min.
156 Thermocycling conditions of the Mix 2 were 95 ºC for 3 min followed by 10 cycles of 95 ºC for 30
157 s, 60 ºC (decreasing 1 ºC for each cycle) for 30 s, 72 ºC for 30 s, followed by 23 cycles of 95 ºC for
158 30 s, 50 ºC for 30 s, 72 ºC for 30 s with a final extension at 72 ºC for 10 min. Genotyping of
159 amplified products were performed by using an ABI3130xl Genetic Analyser (Applied Biosystem,
160 UK) and allele size was determined using the Genescan 500-LIZ size standard and
161 electrophoretograms were scored in Genemapper version 4.0 (Applied Biosystems). All peaks were
162 manually verified by the lead author to ensure genotyping accuracy.
163
164 *Genetic structure and diversity*
165 MICROCHECKER v.2.2.3 was used to assess the presence of null alleles, large allele drop-outs and
166 scoring errors due to stuttering (van Oosterhout, Hutchinson, Wills & Shipley, 2004). GENEPOP
167 v.4.7.0 software (Rousset, 2008) was used to detect deviation from Hardy-Weinberg equilibrium
168 (HWE) and linkage disequilibrium (LD) between pairs of loci and each locus across locations.
169 While HWE test provided possible departures from equilibrium in our locations, which may
170 indicate systematic genotyping errors and other biases (Salanti, Amountza, Ntzani & Joannidis,
171 2005); LD test was used to assess the independence between analysed loci. Exact tests were used
172 with specified Markov chain parameters of 10,000 dememorization steps, followed by 5,000
173 batches of 5,000 iterations per batch. Statistical significance levels were adjusted according to
174 Bonferroni’s procedure to counteract the problem of multiple testing in HWE and linkage
175 disequilibrium (Rice, 1989).

176 In order to characterise the genetic diversity of the red swamp crayfish in the Iberian Peninsula,
177 we estimated the total number of observed alleles (N_A), the effective number of alleles (N_E), the
178 expected and the observed heterozygosity (H_E and H_O, respectively) and the inbreeding coefficients
179 (F_{IS}) for each locus in each location by using GENALEX v.6.503 software (Peakall & Smouse,
180 2012). The allelic richness (A_R) and the number of private alleles (P_A) were calculated with ADZE
181 software (Szpiech, Jakobsson & Rosenberg, 2008), a rarefaction method to be able to compare
182 locations with different sampling sizes. In order to infer the genetic differentiation among locations,
183 pairwise F_{ST} values were calculated by using ARLEQUIN v.3.1 (Excoffier, Laval & Schneider,
184 2005). Bonferroni's correction was performed to adjust the significance for multiple pairwise
185 comparisons in F_{ST} values (Rice, 1989).
186 BOTTLENECK v.1.2.02 was used to identify locations that have recently experienced a
187 significant reduction in effective population size (Piry, Luikart & Cornuert, 1999). This software
188 performs a test of heterozygosity based on the assumption that the number of alleles decreases
189 faster than the heterozygosity when a population experience a bottleneck. The stepwise-mutation
190 (SMM) and two-phased (TPM) models with 10,000 replicates were used to test population
191 bottlenecks. Variance for TPM was set to 30 and the proportion of SMM in TPM was set to 80%.
192 The Wilcoxon's test was used to establish whether the number of loci showing heterozygosity
193 excess was significantly greater than expected in locations at equilibrium.
194 Isolation by distance (IBD) analysis was used to evaluate the relationship between genetic (F_{ST})
195 and geographic (based on X-Y coordinates) distances among pairs of locations (Wright, 1943). A
196 Mantel test with 100,000 replicates was performed using *ade4* package in R software (Dray &
197 Dufour, 2007). To calculate the geographic distances among Iberian locations we used the
198 *geosphere* (Hijmans, Williams, & Vennes, 2017) and *Imap* (Wallace, 2015) packages in R v3.2.3 (R
199 Development Core Team, 2014).

200 STRUCTURE v.2.3.4 was used to characterize the genetic structure of red swamp crayfish in
201 the Iberian Peninsula, and particularly to test whether the two introduction foci can explain the
202 present-day observed genetic structure (Pritchard, Stephens, & Donnelly, 2000). This Bayesian
203 clustering method assigns individuals to a given number of genetic clusters (K) based on their
204 genotypes. In order to identify the number of clusters, we first analysed the likelihood of models
205 with a number of clusters ranging from K = 1 to 27 (n-1). Due to the large number of clusters, we
206 performed 20 independent runs for each K, each run involving a Markov Chain Monte Carlo using
207 2000 burn-in followed by 10,000 iteration steps. Once preliminary results were obtained and to get
208 more accuracy, another analysis was performed from K = 1 to 8 with 20 independent runs for each
209 K, each run involving a Markov Chain Monte Carlo using 200,000 burn-in followed by 1,000,000
210 iteration steps. Admixture ancestry models and correlated allele frequencies (with default
211 parameters) were considered in all cases. The most likely value of real number of clusters in the
212 genetic dataset was estimated by examining the log probability of data [Ln Pr(X|K)] and the ΔK
213 method (Evanno et al., 2005) using STRUCTURE HARVESTER (Earl & vonHoldt, 2012). We
214 summarized the clustering results of multiple runs for each K value and these were visually
215 evaluated in CLUMPAK (http://clumpak.tau.ac.il) (Kopelman, Mayzel, Jakobsson, Rosenberg &
216 Mayrose, 2015). Additionally, a discriminant analysis of principal components (DAPC) was
217 performed to identify the number of different clusters without assuming marker linkage neither
218 HWE (Jombart, Devillard & Balloux, 2010). This multivariate method consists of a two-step
219 procedure to characterize population subdivision, being a Principal Component Analysis (PCA) as a
220 prior step to Discriminant Analysis (DA) (Jombart, Devillard & Balloux, 2010). DAPC was
221 performed using *adegenet* version 2.1.1 (Jombart, 2008) in the R environment.
222
223 *Historical, human and environmental drivers of genetic diversity*

224 To introduce the historical factor in our models, we firstly used the grouping of locations resulting
225 from STRUCTURE (Fig. 1; Table S4) to generate a new categorical variable (genetic group) with
226 two levels (Badajoz and Lower Guadalquivir). Genetic group thus identifies the original
227 introduction foci that originated each of the present-day populations.
228 We then evaluated two alternative scenarios of human-driven spread through secondary
229 introductions: the jump-dispersal scenario and the invasion hub scenario. To test the jump-dispersal
230 scenario (Fig. 2A), we calculated the linear distance (in km) of each location to its corresponding
231 introduction foci (BDJ and LGQ, based on STRUCTURE results) and used this variable (step
232 distance) as a continuous predictor of genetic diversity. This scenario assumes that transport
233 distances are relatively constant among secondary introductions, resulting in an increasing number
234 of jumps for increasing distances. By contrast, to test the invasion hub scenario, we selected the
235 Ebro Delta (DEB) and Valencia Albufera (ABF) (northeast and east coast of Iberian Peninsula,
236 respectively; see Fig. 2B) as plausible invasion hubs, because both are large coastal wetlands with
237 vast areas devoted to rice cropping (similar to the two original introduction foci), which received an
238 important amount of the red swamp crayfish soon after the initial introduction (1978 in the Albufera
239 and 1979 in the Ebro Delta; Gutierrez-Yurrita et al., 1999) and where the species have reached high
240 densities (e.g. Clavero et al., 2015). To test the plausibility of the invasion hub scenario, we
241 calculated the distance (in km) of each location to their associated introduction foci (same as for
242 step distance) and to the candidate invasion hubs (DEB and ABF), and selected the minimum value
243 among these distances to generate a new continuous predictor (hub distance) of genetic diversity.
244 Finally, we characterised the environmental suitability for the red swamp crayfish in each
245 location, in order to test whether higher levels of genetic diversity were related to higher suitability
246 values. We obtained the estimated suitability based on the results of the species distribution model
247 presented by Capinha & Anastácio (2011). These authors collected red swamp crayfish records
248 worldwide, including native and non-native areas, and used six climatic (annual mean temperature,

249 mean temperature of warmest quarter, mean temperature of coldest quarter, annual precipitation,

250 precipitation of wettest quarter, and precipitation of driest quarter) and four physiographic (altitude,

251 slope, distance to ocean, and a compound topographical index) variables to predict the species

252 occurrence in the Iberian Peninsula with a cell resolution of 1×1 km. The environmental suitability,

253 which we used as a continuous predictor of genetic diversity, was calculated as the average value of

254 the 1×1 km cells included within a 5-km buffer constructed around each of our locations, excluding

255 sea surface whenever it was included inside the buffer (see Fig. S1).

256

257 *Statistical analyses*

258 We used generalized linear models (GLMs) to test the influence of the historical, human and

259 environmental factors on the spatial patterns of genetic diversity of red swamp crayfish in the

260 Iberian Peninsula. We ran GLMs using two genetic diversity indices [the allelic richness (A_R) and

261 the expected heterozygosity (H_E)] as dependent variables and genetic group, step distance, hub

262 distance, environmental suitability and sampling size (i.e. number of individuals analysed in each

263 location) as predictors. GLMs used normal error distribution and identity link function for both

264 dependent variables (A_R, Shapiro-Wilk, $W = 0.98$, $P = 0.89$; H_E, Shapiro-Wilk, $W = 0.97$, $P = 0.60$).

265 We first ran univariate GLMs testing the influence of each of the five predictors on each of the two

266 dependent variables. Then, we ran multivariate GLMs and selected final models following a

267 backward stepwise procedure, through which predictors were sequentially excluded from the

268 models attending at the significance of their effects (i.e. higher P-values excluded first) until all

269 predictors had either significant or marginally significant P-values (i.e. equal or lower than 0.1).

270 Backward stepwise procedures for variable selection in multiple regression-type models often use P

271 = 0.1 as a threshold to retain or remove variables. When forward and backward procedures are

272 combined, a common strategy is to use $P < 0.05$ to enter and $P > 0.1$ to remove (e.g. Swartz et al.,

273 2019). GLMs were conducted with the *lme4* package (Bates et al., 2015) in R v3.2.3 (R

274 Development Core Team, 2014).

275

276 **Results**

277 *Genetic diversity*

278 We genotyped eight polymorphic microsatellite loci for 903 red swamp crayfish specimens from 28

279 locations distributed across the Iberian Peninsula. The PclG-29 locus was discarded from our

280 dataset because it had evidence of null alleles and scoring errors due to stuttering. We thus carried

281 out subsequent analyses with the remaining seven loci. The seven microsatellite loci exhibited

282 moderate to high levels of polymorphism (H_E between 0.56 and 0.79) across all locations (Table 2).

283 Most microsatellite loci were found to be in HWE, except locus PCSH02 in Ança (ANC) and

284 Badajoz (BDJ) locations, PCSH06 in BDJ, PCSH65 in ANC and BDJ, PclG-15 in the BDJ and

285 Guadiamar (GUA) and PclG-17 in the BDJ location (Table S2). Badajoz location presented five out

286 of seven loci in HW disequilibrium and also six locus comparisons with significant linkage

287 disequilibrium.

288 We found a total of 98 alleles in the seven microsatellite loci genotyped, with polymorphism

289 ranging from 21 (PclG-15) to eight alleles (PCSH38). At population level, the average number of

290 alleles per locus (N_A) ranged between 4.71 ± 0.78 SE in Valle location (VAL), and 10.57 ± 1.54 SE

291 in Lower Guadalquivir (LGQ), where large number of crayfish were introduced. Overall, measures

292 of genetic diversity such as allelic richness (A_R), observed (H_O) and expected heterozygosity (H_E)

293 were relatively high (Table 2). The A_R ranged from 4.14 (VAL) to 7.93 (LGQ), H_O ranged from

294 0.51 (JAE) to 0.79 (LGQ) and H_E varied from 0.60 (JAE) to 0.80 (LGQ). While Lower

295 Guadalquivir (LGQ) location had the highest genetic diversity among all locations, with the highest

296 allelic richness (7.93 ± 2.65, mean ± SE), observed (0.79 ± 0.04) and expected heterozygosity

297 values (0.80 ± 0.04), the BDJ location had low levels of genetic diversity and the highest F_{IS} value
298 (0.22 ± 0.06) across all locations.
299 We tested whether any of 28 locations had recently experienced a population bottleneck.
300 According to the Wilcoxon's test with two tails, while the TPM model showed that 14 locations had
301 probably experienced bottleneck ($P < 0.05$), the SMM model indicated that only BDJ location could
302 have experienced bottleneck ($P < 0.05$), though other four locations were marginally significant
303 ($0.05 < P > 0.10$) (Table S3). The BDJ location was the only one that had probably experienced a
304 bottleneck using both models.
305 The BDJ location showed the lowest genetic diversity values and highest F_{IS} values ($F_{IS} = 0.22$)
306 among all locations, lower observed than expected heterozygosity, frequent significant Hardy-
307 Weinberg deviations and likely genetic bottleneck (Table 2 and S3). This fact could be explained by
308 the origin of these samples, which were collected from aquaculture ponds (where crayfish were
309 originally introduced) that could have been largely isolated from free-ranging crayfish for an
310 unknown period of time. Because of this, we used the BDJ location for the analyses related to the
311 genetic structure, but not for analyses of genetic diversity patterns.
312
313 *Genetic structure*
314 Clustering genetic structure analysis showed that the most solid structure of red swamp crayfish in
315 the Iberian Peninsula is the one assuming three distinct genetic clusters, K = 3 (Fig. 1 and Fig. S2).
316 Six locations were mainly assigned to cluster 1 (orange), which mostly included Portuguese
317 locations, as well as the Spanish BDJ and JAE locations. This cluster arguably corresponds to the
318 group of locations originated by the spread of the crayfish introduced to Badajoz in 1973, and we
319 henceforth refer to it as the Badajoz group (Table S4). Sixteen locations were assigned to cluster 2
320 (blue) and the remaining six locations were mainly included in cluster 3 (purple). Clusters 2 and 3
321 grouped several and widespread Iberian locations including the Lower Guadalquivir area, where a

322 large batch of red swamp crayfish was introduced in 1974. The identification of cluster 1 (i.e., the
323 Badajoz group) remained constant for different K values (see Fig. 1 for K = 2, K = 3 and K = 6),
324 while clusters 2 and 3 were grouped in a single cluster for K = 2. As we found no clear geographical
325 structure between cluster 2 and 3, we consider them together as the Lower Guadalquivir group. The
326 DAPC analysis grouped the locations in concordance with the STRUCTURE results, with a first
327 axis separating the locations belonging either to the Lower Guadalquivir or Badajoz groups and a
328 second axis that separates mainly the two clusters of the Lower Guadalquivir group (cluster 2 and 3)
329 (Fig. S3). STRUCTURE and DAPC analyses both supporting an admixed origin of the MUN
330 location, which was intermediate between Badajoz and Lower Guadalquivir group. However, it is
331 noteworthy that, unlikely STRUCTURE analysis, the BRU location was not grouped with Lower
332 Guadalquivir in cluster 2.
333 For the IBD analyses, we did not find any relationship between geographic and genetic distances
334 among red swamp crayfish locations in the Iberian Peninsula, neither when considering all analysed
335 locations nor when analysing independently the Badajoz and Lower Guadalquivir group (Mantel
336 tests, $P > 0.5$ in all cases).
337
338 *Drivers of genetic diversity*
339 The GLMs showed that the hub distance (i.e., the minimum distance of one location to its
340 corresponding introduction foci or invasion hub) had a negative influence on the two descriptors of
341 genetic diversity used in the analyses (Table 3; Fig. 3). These relationships were significant in all
342 univariate models and were kept in all multivariate ones (Table S5). Moreover, they were evident
343 for both Badajoz and Lower Guadalquivir genetic groups (Fig. 3). No other predictor was
344 consistently maintained in the multivariate models. For instance, the step distance (i.e., the distance
345 of each location to its corresponding introduction foci) had a weaker, often non-significant, effect
346 on univariate models than hub distance and it was not included in any of the multivariate models

347 (Table S5). These patterns support the existence of genetically diverse invasion hubs other than the
348 original introduction foci, which would have served as sources for secondary introductions (i.e.,
349 invasion hub scenario, Fig. 2B).

350 Genetic diversity figures tended to be higher in locations belonging to the Lower Guadalquivir
351 group than in those of the Badajoz group and to be higher in areas with higher environmental
352 suitability (Fig. 3), though none of these effects were significant in univariate models (Table 3).
353 However, the non-significance of environmental suitability could be related with the relatively high
354 genetic diversity values of the Jiloca location (JIL), the one with the lowest suitability values (see
355 Table 1; Fig. S1).

356

357 Discussion

358 *Introduction history, genetic structure and political boundaries*

359 The genetic patterns observed for the red swamp crayfish in the Iberian Peninsula are associated
360 with a complex human-mediated dispersal process involving both short- and long-distance
361 translocations and give insight to the importance of invasion hubs, which have resulted in a lack of
362 relationship between genetic and spatial distances. This scenario may differ from that described for
363 other invasive freshwater organisms, when expansion after an initial introduction is due to unaided
364 dispersal, to human-driven dispersal involving only short-distance transport or to a combination of
365 both processes. For example, Díez-del-Molino et al. (2013) reported a positive relationship between
366 genetic and spatial distances for Spanish eastern mosquitofish (*Gambusia holbrooki*) populations.
367 However, long-distance human-mediated spread of invasive species is currently a frequent feature
368 of aquatic invasions (Audzijonyte, Baltrūnaitė, Väinölä & Arbačiauskas, 2017; Dias et al., 2018;
369 Wilke et al., 2015) and has been already described for the red swamp crayfish at the global scale
370 (Oficialdegui et al., 2019).

371 We identified two robust genetic groups among red swamp crayfish locations in the Iberian
372 Peninsula, which arguably derive from the quasi-independent expansion of the two crayfish batches
373 introduced to Spain (1973 in Badajoz and 1974 in Lower Guadalquivir). Despite the sources of
374 crayfish for both introductions are arguably close areas in Louisiana (i.e., the native range), the lack
375 of a strong genetic structure and the large degree of genetic admixture in Louisiana (Oficialdegui et
376 al., 2019) could have favoured a random genetic distinction between the two transported batches
377 due to founder effect. Once in the Iberian Peninsula, the Badajoz group would have expanded
378 mainly westward into Portugal, but also, though less intensely, eastward (JAE location). The Lower
379 Guadalquivir group comprised most of the red swamp crayfish Spanish range, including also a
380 location in North-eastern Portugal (VLR location). The probability of belonging to a given group
381 was very high around the introduction foci of both groups (i.e., near Badajoz or around the Lower
382 Guadalquivir, Table S4), a pattern that strengthen the assumption that the observed genetic groups
383 clusters correspond to those initial introduction events. Similarly, in a previous work based on the
384 mtDNA, Oficialdegui et al. (2019) found one haplotype (Hap_06) that was present in most
385 Portuguese locations, but was not detected in the Lower Guadalquivir basin.
386 Since most alien species in inland waters are dispersed by human vectors (Cerri, Ciappelli,
387 Lenuzza, Zaccaroni & Nocita, 2018; Strayer, 2010), clear spatial structuring of genetic variability is
388 often lacking among populations of invasive freshwater species (Audzijonyte et al., 2017; Blakeslee
389 et al., 2017). However, we observed a strong genetic structure among red swamp crayfish
390 populations in the Iberian Peninsula, likely generated by the two expansion ways from both
391 introduction foci. A similar pattern was observed in the invasion process of the European green crab
392 in North America, in which two separately introduction events led to two genetically distinct groups
393 (FitzGerald et al., 2017; Jeffery et al., 2017). Although natural and artificial barriers in rivers can
394 define the spatial distribution or expansion of freshwater invasive species (see Teixeira et al., 2020*)*,
395 the limited admixture in both genetic groups of red swamp crayfish (Badajoz and Lower

396 Guadalquivir) suggests an effect related to the political border between Spain and Portugal. The
397 border has apparently favoured the existence of two quasi-independent expansion processes, despite
398 both countries share several river basins through which natural dispersion of invasive species may
399 occur (Gago, Anastácio, Gkenas, Banha & Ribeiro, 2016). In fact, the red swamp crayfish may had
400 entered in Portugal through natural dispersion since the first record in the country is very near to the
401 introduction area in Badajoz (Ramos & Pereira, 1981), although short-distance human transport
402 cannot be discarded. However, present-day genetic patterns suggest that most subsequent human-
403 driven translocations have remained within the political border of Portugal, with few additional
404 introductions from Spain (even though at least one other did occur, see VLR). Contrastingly, the
405 expansion of the red swamp crayfish across Spain relied on the transport of individuals belonging
406 mainly to the Guadalquivir group (except JAE location). The genetic structure within the Lower
407 Guadalquivir group did not follow any clear spatial pattern (absence of isolation by distance and
408 spatial distribution of clusters 2 and 3), fitting well with the patterns reported for other widely
409 spread freshwater species (see above). The political border between Spain and Portugal may thus
410 act as an actual ecological barrier, as has been already described for other taxa (Arrondo et al.,
411 2018; García et al., 2018). In the red swamp crayfish case this barrier does not seem to be related to
412 policy differences between countries (as reported Arrondo et al., 2018), but to the behaviour of the
413 people stocking crayfish, who apparently tended to remain within national limits. We thus
414 emphasize the importance of political boundaries as invisible barriers that can determine the
415 structure of wild populations, especially so for those species translocated by humans, calling for an
416 international coordination management measures accordingly (Dresser, Pierson, & Fitzpatrick,
417 2018; Rollins, Woolnough, Wilton, Sinclair & Sherwin, 2009).

418

419 *Drivers of genetic diversity*

420 Propagule pressure is a key factor modulating the probability of establishment in introduced
421 populations and their dynamics of genetic diversity (Lockwood, Cassey & Blackburn, 2005).
422 Overall, genetic diversity indices were lower in locations of the Badajoz group than in those of the
423 Lower Guadalquivir group, a pattern probably related to the higher propagule size of the
424 introduction event into the Lower Guadalquivir. In fact, the number of crayfish involved in the
425 Guadalquivir introduction was around 20 times larger than in the one taking place in Badajoz
426 (Habsburgo-Lorena, 1978).

427 The nature of transport events (short-distance, long-distance or both) and the existence of one
428 or several invasion hubs acting as genetically diverse sources of individuals (e.g., Fig. 2A and 2B)
429 may influence the genetic diversity spatial patterns of an already established invasive species. In the
430 Iberian Peninsula, it seems that large stocks of red swamp crayfish specimens were long-distance
431 translocated without intermediary bottlenecks (Gutierrez-Yurita et al., 1999), generating high
432 genetically diverse invasion hubs that subsequently acted as source for multiple secondary
433 introduction events. Accordingly, we found that genetic diversity tended to decrease in locations
434 that were farther to either its respective introduction focus or invasion hub. This steady decline in
435 genetic diversity along the invasion process corroborate the genetic consequences described for
436 population expansion range (Austerlitz et al., 1997; Excoffier et al., 2009). For example, White et
437 al. (2013) found a significant decline in genetic diversity across the expansion range of the bank
438 vole (*Myodes glareolus*) in Ireland. Based on historical information and environmental
439 characteristics, we had selected *a priori* the Ebro Delta and the Valencia Albufera as plausible
440 invasion hubs, but additional invasion hubs could have existed. Candidate areas for this role could
441 be the Tablas de Daimiel National Park, where the red swamp crayfish was introduced in 1982, or
442 the northern Spanish plateau, were a long-lasting tradition around crayfish consumption could have
443 favoured the occurrence of several introduction events (Clavero, 2016). Detecting possible invasion

444 hubs is a determining factor to predict and prevent potential range expansion of invasive species
445 over non-native territories (Muirhead & MacIsaac, 2005).
446 How the genetic diversity is affected by the distance to the introduction focus and by the
447 accumulation of bottleneck effects have been well studied in invasion biology (van Boheemen et al.,
448 2017). However, our results also showed a positive relationship between genetic diversity and
449 environmental suitability for the red swamp crayfish in the Iberian Peninsula, highlighting the
450 importance of ecological factors in shaping the genetic patterns of invasive species. Similar patterns
451 have been reported for other invertebrate species (Ortego, Aguirre, Noguerales & Cordero, 2015).
452 They support the influence of environmental suitability on evolutionary processes through
453 demographic mechanisms that affect the effective population size (Wang, 2012) and ultimately
454 module the genetic patterns of populations. Furthermore, the relevance of the suitability-genetic
455 diversity relationships for the management of biological invasions can be modulated by climate
456 change (Tilman, Balzer, Hill & Befort, 2011). Capinha et al. (2012) predicted that suitable areas for
457 the red swamp crayfish in the Iberian Peninsula would show moderate changes (including both
458 suitability increases and decreases) from the present situation. But warmer future environments will
459 enhance climate suitability for the red swamp crayfish across several European areas (Zhang et al.,
460 2019), which according to our results, could end up hosting viable and more genetically diverse red
461 swamp crayfish populations.
462
463 *Conclusions*
464 We have described clear patterns in genetic structure of red swamp crayfish in the Iberian
465 Peninsula, determined by the two introduction events that took place in the 1970s, a complex
466 human-mediated dispersal process involving the presence of invasion hubs, and a tendency of
467 higher levels of genetic diversity occurring in more suitable environments. These results help to
468 comprehend the invasion history of the red swamp crayfish in the Iberian Peninsula and how the

469 natural and anthropogenic factors modulate it. Our study thus highlights the importance of
470 analysing patterns of genetic variability to understand the invasion processes, a knowledge that can
471 be applied to manage current invasions and prevent plausible future ones.
472
473 **iv. Acknowledgements**

474 We thank R. López, M. A. Bravo, and Á. Vidal for field and lab assistance, as well as the many
475 colleagues who kindly provided samples from Spain [A. Pradilla (CCEDCV - EL PALMAR), L. M.
476 Álvarez (Asturias Government), N. Franch (Parc Natural del Delta de l'Ebre), L. Valls, F. J.
477 Galindo (Andalusian Government), F. Unzu and L. Lopo (La Rioja Government), J. Bosch, F. J.
478 Oliva, I. Vedia] and Portugal (F. Ribeiro and A. F. Filipe). We also thank Laboratorio de Ecología
479 Molecular (LEM-EBD) at Estación Biológica de Doñana (CSIC) as well as C. Lejeusne and C.
480 Daguin-Thiébaut for providing logistical support. We are especially grateful to César Capinha for
481 sharing with us his dataset. We thank two anonymous reviewers who gave helpful comments to
482 improve the final version of this manuscript. F.J.O. was supported by an Andalusian Government
483 grant. The study was funded by the Andalusian Government (RNM-936) and PICS (PIC2015FR4).
484
485 **Conflicts of Interest**

486 The authors declare no conflict of interest.
487
488 **v. References**

489 Aljanabi, S. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-
490 based techniques. *Nucleic Acids Research*, 25(22), 4692–4693.
491 https://doi.org/10.1093/nar/25.22.4692
492 Arrondo, E., Moleón, M., Cortés-Avizanda, A., Jiménez, J., Beja, P., Sánchez-Zapata, J. A., &
493 Donázar, J. A. (2018). Invisible barriers: Differential sanitary regulations constrain vulture

494 movements across country borders. *Biological Conservation*, 219, 46–52.
495 https://doi.org/10.1016/j.biocon.2017.12.039
496 Audzijonyte, A., Baltrūnaitė, L., Väinölä, R., & Arbačiauskas, K. (2017). Human-mediated lineage
497 admixture in an expanding Ponto-Caspian crustacean species *Paramysis lacustris* created a
498 novel genetic stock that now occupies European waters. *Biological Invasions*, 19(8), 2443–
499 2457. https://doi.org/10.1007/s10530-017-1454-9
500 Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P. H. (1997). Evolution of coalescence
501 times, genetic diversity and structure during colonization. Theoretical Population
502 Biology, 51(2), 148-164.
503 Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., … Peakall, R.
504 (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology and*
505 *Evolution*, 28(11), 670–679. https://doi.org/10.1016/j.tree.2013.08.005
506 Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Rcpp, L.
507 (2015). Package ‘lme4’. *Convergence*, 12(1).
508 Belfiore, N. M., & May, B. (2000). Variable microsatellite loci in red swamp crayfish,
509 *Procambarus clarkii*, and their characterization in other crayfish taxa. *Molecular Ecology*, 9,
510 2231–2234. https://doi.org/10.1046/j.1365-294X.2000.105339.x
511 Bellard, C., Cassey, P., & Blackburn, T. M. (2016). Alien species as a driver of recent
512 extinctions. *Biology letters*, 12(2), 20150623. https://doi.org/10.1098/rsbl.2015.0623
513 Blackburn, T. M., & Duncan, R. P. (2001). Determinants of establishment success in introduced
514 birds. *Nature*, 414(6860), 195–197. https://doi.org/10.1038/35102557
515 Blackburn, T. M., Prowse, T. A. A., Lockwood, J. L., & Cassey, P. (2013). Propagule pressure as a
516 driver of establishment success in deliberately introduced exotic species: Fact or artefact?
517 *Biological Invasions*, 15(7), 1459–1469. https://doi.org/10.1007/s10530-013-0451-x
518 Blackburn, T. M., Pyšek, P., Bacher, S., Carlton, J. T., Duncan, R. P., Jarošík, V., … Richardson,

519 D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology and*
520 *Evolution*, 26(7), 333–339. https://doi.org/10.1016/j.tree.2011.03.023
521 Blakeslee, A. M. H., Kamakura, Y., Onufrey, J., Makino, W., Urabe, J., Park, S., … Miura, O.
522 (2017). Reconstructing the invasion history of the Asian shorecrab, *Hemigrapsus sanguineus*
523 (De Haan 1835) in the Western Atlantic. *Marine Biology*, 164:47. doi: 10.1007/s00227-017-
524 3069-1.
525 Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology letters*,
526 11(2), 106-115. https://doi.org/10.1111/j.1461-0248.2007.01130.x
527 Capinha, C., & Anastácio, P. (2011). Assessing the environmental requirements of invaders using
528 ensembles of distribution models. *Diversity and Distributions*, 17(1), 13–24.
529 https://doi.org/10.1111/j.1472-4642.2010.00727.x
530 Capinha, C., Anastácio, P., & Tenedório, J. A. (2012). Predicting the impact of climate change on
531 the invasive decapods of the Iberian inland waters: An assessment of reliability. *Biological*
532 *Invasions*, 14(8), 1737–1751. https://doi.org/10.1007/s10530-012-0187-z
533 Capinha, C., Essl, F., Seebens, H., Moser, D., & Pereira, H. M. (2015). The dispersal of alien
534 species redefines biogeography in the Anthropocene. *Science*, 348(6240), 1248–1251.
535 https://doi.org/10.1126/science.aaa8913
536 Cerri, J., Ciappelli, A., Lenuzza, A., Zaccaroni, M., & Nocita, A. (2018). Recreational angling as a
537 vector of freshwater invasions in Central Italy: perceptions and prevalence of illegal fish
538 restocking. *Knowledge & Management of Aquatic Ecosystems*, (419), 38.
539 https://doi.org/10.1051/kmae/2018028
540 Clavero, M. (2016). Species substitutions driven by anthropogenic positive feedbacks: Spanish
541 crayfish species as a case study. *Biological Conservation*, 193, 80–85.
542 https://doi.org/10.1016/j.biocon.2015.11.017
543 Clavero, M., López, V., Franch, N., Pou-Rovira, Q., & Queral, J. M. (2015). Use a seasonally

544 flooded rice fields by fish and crayfish in a Mediterranean wetland. *Agriculture*, *Ecosystems & Environment*, 213, 39-46. https://doi.org/10.1016/j.agee.2015.07.022

546 Dias, P. J., Gilg, M. R., Lukehurst, S. S., Kennington, W. J., Huhn, M., Madduppa, H. H., ... McDonald, J. I. (2018). Genetic diversity of a hitchhiker and prized food source in the Anthropocene: the Asian green mussel *Perna viridis* (Mollusca, Mytilidae). *Biological invasions*, 20(7), 1749-1770.

550 Díez-del-Molino, D., Carmona-Catot, G., Araguas, R. M., Vidal, O., Sanz, N., García-Berthou, E., & García-Marín, J. L. (2013). Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). *PLoS ONE*, 8(12), e82501. https://doi.org/10.1371/journal.pone.0082501

554 Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, 17(1), 431–449. https://doi.org/10.1111/j.1365-294X.2007.03538.x

557 Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4), 1-20. https://doi.org/10.18637/jss.v022.i04

559 Dresser, C. M., Pierson, T. W., & Fitzpatrick, B. M. (2018). Isolation by distance, local adaptation, and fortuitous coincidence of geo-political boundaries with spatial-genetic clusters in southern Bog Turtles. *Global Ecology and Conservation*, 16, e00474. https://doi.org/10.1016/j.gecco.2018.e00474

563 Drolet, D., & Locke, A. (2016). Relative importance of propagule size and propagule number for establishment of non-indigenous species: A stochastic simulation study. *Aquatic Invasions*, 11(1), 101–110. https://doi.org/10.3391/ai.2016.11.1.11

566 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

569 Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*,
570 17(7), 422–433. https://doi.org/10.1038/nrg.2016.58
571 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using
572 the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620.
573 https://doi.org/10.1111/j.1365-294X.2005.02553.x
574 Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software
575 package for population genetics data analysis. *Evolutionary bioinformatics*, 1, 47-50.
576 https://doi.org/10.1177/117693430500100003
577 Excoffier, L., Foll, M., & Petit, J. (2009). Genetic Consequences of Range Expansions. *Annual*
578 *Review of Ecology, Evolution, and Systematics*, 40, 481-501.
579 https://doi.org/10.1146/annurev.ecolsys.39.110707.173414
580 Facon, B., Genton, B. J., Shykoff, J., Jarne, P., Estoup, A., & David, P. (2006). A general eco-
581 evolutionary framework for understanding bioinvasions. *Trends in Ecology and Evolution*,
582 21(3), 130–135. https://doi.org/10.1016/j.tree.2005.10.012
583 FitzGerald, J., Bradbury, I. R., Nadukkalam Ravindran, P., DiBacco, C., Stanley, R. R. E.,
584 Matheson, K., … Van Wyngaarden, M. (2017). RAD sequencing reveals genomewide
585 divergence between independent invasions of the European green crab (*Carcinus maenas*) in
586 the Northwest Atlantic. *Ecology and Evolution*, 7(8), 2513–2524.
587 https://doi.org/10.1002/ece3.2872
588 Gago, J., Anastácio, P., Gkenas, C., Banha, F., & Ribeiro, F. (2016). Spatial distribution patterns of
589 the non-native European catfish, *Silurus glanis*, from multiple online sources–a case study for
590 the River Tagus (Iberian Peninsula). *Fisheries Management and Ecology*, 23(6), 503-509.
591 https://doi.org/10.1111/fme.12189
592 García, E., López-Bao, J. V., Ferrand, N., Blanco, J. C., Palacios, V., Cortés, Y., … Rio-Maior, H.
593 (2018). Cryptic population structure reveals low dispersal in Iberian wolves. *Scientific Reports*,

594 8(1), 1–14. https://doi.org/10.1038/s41598-018-32369-3
595 Geiger, W., Alcorlo, P., Baltanas, A., & Montes, C. (2005). Impact of an introduced Crustacean on
596 the trophic webs of Mediterranean wetlands. *Biological Invasions*, 7(1), 49-73.
597 https://doi.org/10.1007/1-4020-3870-4_6
598 Gherardi, F. (2006). Crayfish invading Europe: The case study of *Procambarus clarkii*. *Marine and*
599 *Freshwater Behaviour and Physiology*, 39(3), 175–191.
600 https://doi.org/10.1080/10236240600869702
601 Gutiérrez-Yurrita, P. J., Martinez, J. M., Ilhéu, M., Bravo-Utrera, M. A., Bernardo, J. M., &
602 Montes, C. (1999). The status of crayfish populations in Spain and Portugal. *Crustacean*
603 *Issues*, 11, 161-192.
604 Habsburgo-Lorena, A. S. (1978). Present situation of exotic species of crayfish introduced into
605 Spanish continental water. *Freshwater Crayfish*, 4, 175–184.
606 Hardesty, B. D., Le Roux, J. J., Rocha, O. J., Meyer, J. Y., Westcott, D., & Wieczorek, A. M.
607 (2012). Getting here from there: Testing the genetic paradigm underpinning introduction
608 histories and invasion success. *Diversity and Distributions*, 18(2), 147–157.
609 https://doi.org/10.1111/j.1472-4642.2011.00832.x
610 Hawes, N. A., Fidler, A. E., Tremblay, L. A., Pochon, X., Dunphy, B. J., & Smith, K. F. (2018).
611 Understanding the role of DNA methylation in successful biological invasions: a review.
612 *Biological Invasions*, 20(9), 2285-2300.
613 Hijmans, R. J., Williams, E., & Vennes, C. (2017). Package ‘geosphere’ version 1.5-7.
614 https://CRAN.R-project.org/package=geosphere
615 Hulme, P. E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., … Vilà, M. (2008). Grasping
616 at the routes of biological invasions: A framework for integrating pathways into policy.
617 *Journal of Applied Ecology*, 45(2), 403–414. https://doi.org/10.1111/j.1365-
618 2664.2007.01442.x

619 Jeffery, N. W., Dibacco, C., Wyngaarden, M. Van, Hamilton, L. C., Stanley, R. R. E., Bernier, R.,
620 … Bradbury, I. R. (2017). ORIGINAL RESEARCH RAD sequencing reveals genomewide
621 divergence between independent invasions of the European green crab (*Carcinus maenas*) in
622 the Northwest Atlantic. *Ecology and evolution*, *7*(8), 2513-2524.
623 https://doi.org/10.1002/ece3.2872
624 Jiang, H., Qian, Z., Lu, W., Xing, Z., Yu, H., & Li, J. (2015). Microsatellite marker identification
625 from transcriptome derived sequences of the red swamp crawfish, *Procambarus clarkii*.
626 *Conservation Genetics Resources*, 7(3), 729–731. https://doi.org/10.1007/s12686-015-0447-1
627 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic
628 markers. *Bioinformatics*, *24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129
629 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a
630 new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 94.
631 https:// doi.org/10.1186/1471-2156-11-94.
632 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A
633 program for identifying clustering modes and packaging population structure inferences across
634 K. *Molecular Ecology Resources*, 15(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387
635 Lavergne, S., & Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive
636 the success of an invasive grass. *Proceedings of the National Academy of Sciences*, 104(10),
637 3883–3888. https://doi.org/10.1073/pnas.0607324104
638 Lockwood, J. L., Cassey, P., & Blackburn, T. (2005). The role of propagule pressure in explaining
639 species invasions. *Trends in Ecology and Evolution*, 20(5), 223–228.
640 https://doi.org/10.1016/j.tree.2005.02.004
641 Muirhead, J. R., & MacIsaac, H. J. (2005). Development of inland lakes as hubs in an invasion
642 network. *Journal of Applied Ecology*, 42, 80–90. https://doi.org/10.1111/j.1365-
643 2664.2004.00988.x

644 Oficialdegui, F. J., Green, A. J., Clavero, M., Sánchez, M. I., Boyero, L., Michot, T. C., …
645 Lejeusne, C. (2019). Unravelling the global invasion routes of a worldwide invader, the red
646 swamp crayfish (*Procambarus clarkii*). *Freshwater Biology*, 64(8), 1382–1400.
647 https://doi.org/10.1111/fwb.13312
648 Oficialdegui, F. J., Sánchez, M. I., Clavero, M. (2020) One century away from home: how the red
649 swamp crayfish took over the world. *In Press*. *Reviews in Fish Biology and Fisheries*.
650 https://doi.org/10.1007/s11160-020-09594-z
651 Okada, M., Lyle, M., & Jasieniuk, M. (2009). Inferring the introduction history of the invasive
652 apomictic grass *Cortaderia jubata* using microsatellite markers. *Diversity and Distributions*,
653 15(1), 148–157. https://doi.org/10.1111/j.1472-4642.2008.00530.x
654 Ortego, J., Aguirre, M. P., Noguerales, V., & Cordero, P. J. (2015). Consequences of extensive
655 habitat fragmentation in landscape-level patterns of genetic diversity and structure in the
656 Mediterranean esparto grasshopper. *Evolutionary Applications*, 8(6), 621–632.
657 https://doi.org/10.1111/eva.12273
658 Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic
659 software for teaching and research-an update. *Bioinformatics*, *28*(19), 2537–2539.
660 https://doi:10.1093/bioinformatics/bts460
661 Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: a computer program for detecting
662 recent reductions in the effective population size using allele frequency data. *Journal of*
663 *heredity*, 90, 502-503. https://doi.org/10.1093/jhered/90.4.502
664 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
665 multilocus genotype data. *Genetics*, 155(2), 945–959. https://doi.org/10.1111/j.1471-
666 8286.2007.01758.x
667 Ramos, M. A., & Pereira, T. G. (1981). Um novo Astacidae para a Fauna Portuguesa *Procambarus*
668 *clarkii* (Girard, 1852). *Boletim Inst. Nacional de Investigac¸a˜o e Pescas*, *Lisboa*. 6, 37-47.

669 Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global change?
670 *Conservation Biology*, 21(2), 329–336. https://doi.org/10.1111/j.1523-1739.2006.00615.x
671 Rice, W. R. (1989). Analyzing tables of statistical test. *Evolution*, 43(1), 223–225.
672 https://doi.org/10.1111/j.0014-3820.2001.tb00731.x
673 Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success
674 of colonising populations? *Trends in Ecology and Evolution*, 29(4), 233–242.
675 https://doi.org/10.1016/j.tree.2014.02.003
676 Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R., & Sherwin, W. B. (2009). Invasive
677 species can't cover their tracks: Using microsatellites to assist management of starling (*Sturnus*
678 *vulgaris*) populations in Western Australia. *Molecular Ecology*, 18(8), 1560–1573.
679 https://doi.org/10.1111/j.1365-294X.2009.04132.x
680 Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for
681 Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106.
682 https://doi.org/10.1111/j.1471-8286.2007.01931.x
683 Salanti, G., Amountza, G., Ntzani, E. E., & Ioannidis, J. P. A. (2005). Hardy-Weinberg equilibrium
684 in genetic association studies: An empirical evaluation of reporting, deviations, and power.
685 *European Journal of Human Genetics*, 13(7), 840–848.
686 https://doi.org/10.1038/sj.ejhg.5201410
687 Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., … Essl, F.
688 (2017). No saturation in the accumulation of alien species worldwide. *Nature*
689 *Communications*, 8, 1–9. https://doi.org/10.1038/ncomms14435
690 Souty-Grosset, C., Anastácio, P. M., Aquiloni, L., Banha, F., Choquer, J., Chucholl, C., &
691 Tricarico, E. (2016). The red swamp crayfish *Procambarus clarkii* in Europe: Impacts on
692 aquatic ecosystems and human well-being. *Limnologica*, 58, 78–93.
693 https://doi.org/10.1016/j.limno.2016.03.003

694 Strayer, D. L. (2010). Alien species in fresh waters: Ecological effects, interactions with other
695 stressors, and prospects for the future. *Freshwater Biology*, *55*, 152-174.
696 https://doi.org/10.1111/j.1365-2427.2009.02380.x
697 Swartz, L. K., Hossack, B. R., Muths, E., Newell, R. L., & Lowe, W. H. (2019). Aquatic
698 macroinvertebrate community responses to wetland mitigation in the Greater Yellowstone
699 Ecosystem. Freshwater Biology, 64(5), 942-953.
700 Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: A rarefaction approach for
701 counting alleles private to combinations of populations. *Bioinformatics*, 24(21), 2498–2504.
702 https://doi.org/10.1093/bioinformatics/btn478
703 Teixeira, D. F., Neto, F. R. A., Gomes, L. C., Beheregaray, L. B., & Carvalho, D. C. (2020)
704 Invasion dynamics of the white piranha (*Serrasalmus brandtii*) in a Neotropical river basin.
705 Biological Invasions, in press. https://doi.org/10.1007/s10530-019-02138-y
706 Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). *Climate change 2014: synthesis report*.
707 *Proceedings of the National Academy of Sciences* (Vol. 108).
708 https://doi.org/10.1073/pnas.1116437108
709 van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins,
710 K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the
711 introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology*,
712 26(20), 5421–5434. https://doi.org/10.1111/mec.14293
713 Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER:
714 Software for identifying and correcting genotyping errors in microsatellite data. *Molecular*
715 *Ecology Notes*, 4(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x
716 Wallace, J. R. (2015). Interactive Mapping: Package 'IMAP'. version 1.32. https://CRAN.R-
717 project.org/package=Imap.
718 Walsh, J. R., Carpenter, S. R., & Vander Zanden, M. J. (2016). Invasive species triggers a massive

719 loss of ecosystem services through a trophic cascade. *Proceedings of the National Academy of*
720 *Sciences*, 113(15), 4081–4085. https://doi.org/10.1073/pnas.1600366113
721 Wang, I. J. (2012). Environmental and topographic variables shape genetic structure and effective
722 population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18(10), 1033–
723 1041. https://doi.org/10.1111/j.1472-4642.2012.00897.x
724 White, T. A., Perkins, S. E., Heckel, G., & Searle, J. B. (2013). Adaptive evolution during an
725 ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular*
726 *Ecology*, *22*(11), 2971-2985. https://doi.org/10.1111/mec.12343
727 Wilke, T., von Oheimb, P. V., Albrecht, C., Marescaux, J., Van Doninck, K., Etoundi, E., & von
728 Oheimb, K. C. M. (2015). Unravelling the invasion pathways of the quagga mussel (*Dreissena*
729 *rostriformis*) into Western Europe. *Biological Invasions*, 18(1), 245–264.
730 https://doi.org/10.1007/s10530-015-1005-1
731 Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.
732 Zhang, Z., Capinha, C., Usio, N., Weterings, R., Liu, X., Li, Y., Landeria, J.M., Zhou, Q., Yokota,
733 M. (2019). Impacts of climate change on the global potential distribution of two notorious
734 invasive crayfishes. *Freshwater biogy, 00*, 1-13. https://doi.org/10.1111/fwb.13429
735
736
737
738

739 Table 1. Information on the 28 locations of red swamp crayfish (*Procambarus clarkii*) surveyed in
740 the Iberian Peninsula, including name, code (as in Table 2 and Fig. 1), number of collected
741 individuals (N), geographical coordinates, type of habitat, environmental suitability (as reported by
742 Capinha & Anastácio, 2011), distance (km) to the original introduction focus (step distance) and
743 minimum distance between those to the introduction focus or to the nearest invasion hub (hub
744 distance) (see Fig. 2D).

Locations	**Code**	**N**	**Lat**	**Lon**	**Habitat**	**Enviromental suitability**	**Step dist**	**Hub dist**
Albufera	ABF	26	39.184	-0.192	Rice Field	0.873	566	0
Ança	ANC	30	40.160	-8.306	River	0.803	206	206
Arreo	ARE	20	42.779	-2.991	Lake	0.764	683	382
Brugent	BRU	30	42.021	2.362	River	0.808	927	213
Cidacos	CID	29	42.274	-1.373	River	0.825	708	274
Delta del Ebro	DEB	28	40.783	0.690	Rice Field	0.891	718	0
Badajoz	BDJ	31	38.899	-6.871	Ponds	0.833	0	0
Guadalporcún	GDP	50	36.565	-5.213	River	0.731	75	75
Gijón	GIJ	13	43.321	-5.382	Pond	0.886	713	611
Guadiamar	GUA	50	37.392	-6.134	River	0.801	59	59
Hueznar	HUE	40	37.556	-5.415	River	0.767	98	98
Jaén	JAE	30	37.494	-3.441	River	0.836	301	301
Jiloca	JIL	15	40.544	-1.293	River	0.538	568	171
Leza	LEZ	30	42.263	-2.184	Stream	0.790	676	316
Lower Guadalquivir	LGQ	49	37.755	-6.959	Rice Field	0.894	0	0
Lousal	LOU	30	38.014	-8.255	River	0.857	164	164
Madrid	MAD	30	40.400	-4.056	Pond	0.646	435	350
Mundo	MUN	21	38.273	-1.462	Stream	0.678	415	157
Olivargas	OLI	50	37.471	-6.486	River	0.826	93	93
Reguengos	REG	30	38.284	-7.312	River	0.836	71	71
Requeixo	REQ	30	40.353	-8.313	River	0.786	233	233
Rocina	ROC	30	37.101	-6.372	Stream	0.826	41	41
Sopetón	SOP	20	36.573	-6.266	Lagoon	0.874	32	32
Sotogrande	STG	50	37.097	-6.463	Lake	0.883	27	27
Valle	VAL	50	36.050	-5.414	River	0.737	124	124
Villar	VIL	32	37.412	-6.433	River	0.825	79	79
Valoria la buena	VLB	29	41.801	-4.588	Stream	0.729	536	454
Vila-Rica	VLR	30	41.135	-7.055	Stream	0.744	462	462
Total		903						

745

746

747 Table 2. Summary of genetic diversity values for seven microsatellite loci in the 28 sampled

748 locations of red swamp crayfish distributed across the Iberian Peninsula. N = average number of

749 crayfish in each location/loci, N_A = mean number of alleles observed, N_E = mean number of

750 effective alleles, P_A = mean number of private alleles corrected by the sampling size, A_R = mean

751 allelic richness, H_O = mean observed heterozygosity, H_E = mean expected heterozygosity, and F_{IS} =

752 mean fixation index. Values are shown by mean of the seven microsatellite markers and standard

753 error (SE).

Sampled locations		**N**	**N_A**	**N_E**	**P_A**	**A_R**	**H_O**	**H_E**	**F_{IS}**
ABF	Mean	25.429	8.571	4.947	0.261	7.138	0.736	0.728	-0.022
	± SE	0.149	0.644	0.592	0.137	1.112	0.022	0.026	0.024
ANC	Mean	30.000	6.143	3.606	0.103	5.364	0.605	0.687	0.114
	± SE	0.000	0.400	0.294	0.091	0.684	0.032	0.020	0.044
ARE	Mean	19.857	4.857	2.632	0.000	4.472	0.568	0.590	0.015
	± SE	0.071	0.442	0.146	0.000	0.710	0.016	0.023	0.033
BDJ	Mean	31.000	5.286	2.497	0.000	4.444	0.465	0.560	0.219
	± SE	0.000	0.340	0.152	0.000	0.466	0.051	0.027	0.060
BRU	Mean	30.000	4.857	3.280	0.071	4.554	0.619	0.624	0.001
	± SE	0.000	0.335	0.355	0.063	0.647	0.028	0.030	0.015
CID	Mean	29.000	7.571	4.682	0.013	6.347	0.768	0.749	-0.028
	± SE	0.000	0.697	0.388	0.012	0.980	0.021	0.019	0.017
DEB	Mean	28.000	8.714	5.046	0.037	7.184	0.760	0.783	0.034
	± SE	0.000	0.531	0.340	0.035	0.863	0.025	0.012	0.021
GDP	Mean	49.857	7.857	4.263	0.127	6.062	0.714	0.714	-0.002
	± SE	0.071	0.694	0.455	0.123	0.911	0.021	0.022	0.012
GIJ	Mean	12.571	6.000	3.726	0.000	5.921	0.758	0.709	-0.066
	± SE	0.101	0.267	0.202	0.000	0.525	0.029	0.018	0.025
GUA	Mean	49.857	9.429	4.753	0.045	7.045	0.667	0.764	0.120
	± SE	0.071	0.625	0.357	0.034	0.813	0.012	0.015	0.020
HUE	Mean	39.571	7.143	4.139	0.046	5.978	0.741	0.750	0.016
	± SE	0.149	0.369	0.151	0.046	0.500	0.019	0.010	0.014
JAE	Mean	30.000	5.143	2.712	0.059	4.381	0.505	0.603	0.161
	± SE	0.000	0.335	0.140	0.059	0.511	0.027	0.023	0.036
JIL	Mean	15.000	6.571	4.781	0.000	6.416	0.724	0.759	0.045
	± SE	0.000	0.434	0.364	0.000	0.842	0.029	0.019	0.030

LEZ	Mean	30.000	8.286	4.589	0.027	6.846	0.714	0.755	0.052
	± SE	0.000	0.508	0.384	0.026	0.762	0.016	0.014	0.016
LGQ	Mean	47.286	10.571	6.079	0.126	7.932	0.787	0.795	0.009
	± SE	0.237	0.770	0.600	0.089	1.043	0.020	0.019	0.014
LOU	Mean	30.000	6.714	3.541	0.000	5.685	0.686	0.693	0.009
	± SE	0.000	0.459	0.203	0.000	0.682	0.020	0.018	0.019
MAD	Mean	30.000	6.286	3.740	0.001	5.467	0.743	0.701	-0.068
	± SE	0.000	0.389	0.223	0.001	0.637	0.019	0.023	0.015
MUN	Mean	21.000	5.286	3.227	0.002	4.950	0.633	0.629	-0.019
	± SE	0.000	0.322	0.283	0.002	0.605	0.036	0.033	0.034
OLI	Mean	48.714	5.857	3.100	0.034	4.643	0.633	0.645	0.026
	± SE	0.143	0.442	0.203	0.034	0.576	0.026	0.021	0.018
REG	Mean	30.000	6.857	4.099	0.000	5.995	0.695	0.730	0.051
	± SE	0.000	0.550	0.282	0.000	0.792	0.021	0.016	0.009
REQ	Mean	29.857	5.286	3.438	0.000	4.765	0.646	0.679	0.043
	± SE	0.071	0.340	0.234	0.000	0.581	0.021	0.019	0.025
ROC	Mean	30.000	8.571	4.581	0.001	6.817	0.714	0.737	0.035
	± SE	0.000	0.635	0.415	0.001	0.953	0.026	0.021	0.010
SOP	Mean	20.000	5.857	3.318	0.000	5.400	0.621	0.649	0.041
	± SE	0.000	0.442	0.240	0.000	0.753	0.040	0.031	0.042
STG	Mean	50.000	9.714	5.086	0.054	6.737	0.777	0.745	-0.054
	± SE	0.000	0.696	0.564	0.046	1.093	0.019	0.023	0.027
VAL	Mean	49.857	4.714	3.008	0.004	4.141	0.604	0.626	0.016
	± SE	0.071	0.389	0.240	0.004	0.647	0.028	0.023	0.042
VIL	Mean	31.429	6.429	3.854	0.000	5.434	0.629	0.666	0.062
	± SE	0.101	0.606	0.415	0.000	0.891	0.035	0.032	0.012
VLB	Mean	29.000	7.286	4.223	0.001	6.162	0.739	0.726	-0.027
	± SE	0.000	0.574	0.324	0.001	0.830	0.020	0.022	0.020
VLR	Mean	29.714	6.143	3.404	0.000	5.034	0.654	0.659	0.005
	± SE	0.092	0.631	0.271	0.000	0.792	0.024	0.026	0.009
TOTAL	**Mean**	**32.030**	**6.857**	**3.941**	**0.036**	**5.761**	**0.675**	**0.695**	**0.028**
	± SE	**0.748**	**0.213**	**0.134**	**0.011**	**0.192**	**0.011**	**0.009**	**0.010**

754

755

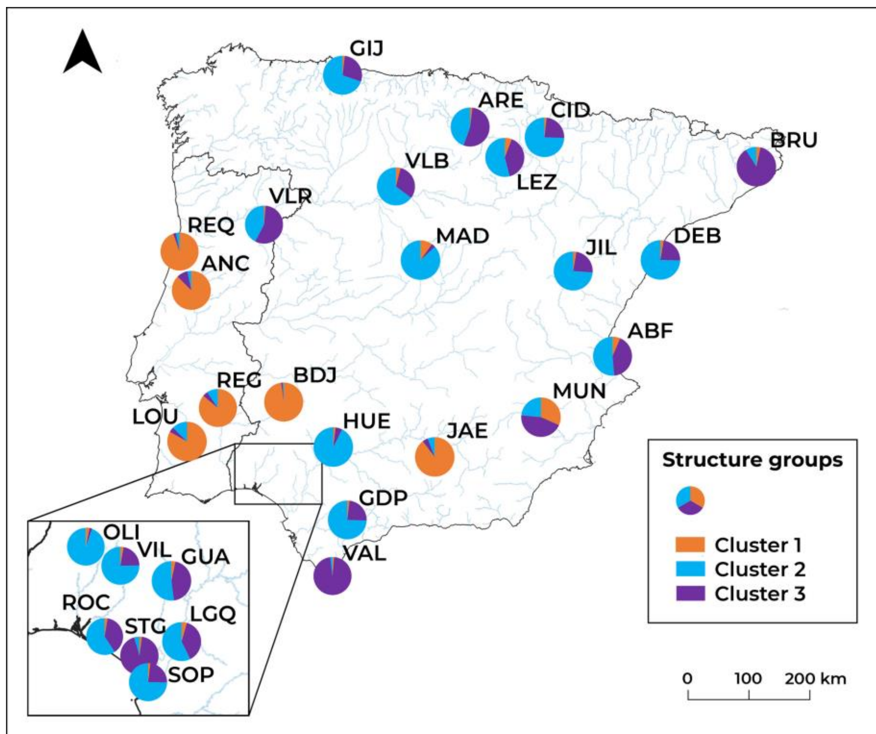
756 Table 3. Univariate and multivariate general linear models assessing the influence of different
757 predictors on the estimators of genetic diversity (allelic richness and expected heterozygosity).
758 Results are provided in terms of the direction of the relationship (positive “POS” or negative
759 “NEG”) for continuous predictors and comparing Lower Guadalquivir (LGQ) and Badajoz (BDJ)
760 groups, for the genetic group factor, with an indication of the statistical significance. The coefficient
761 of determination of the final multivariate models (selected following a backward procedure) is also
762 showed (for full models see Table S5).

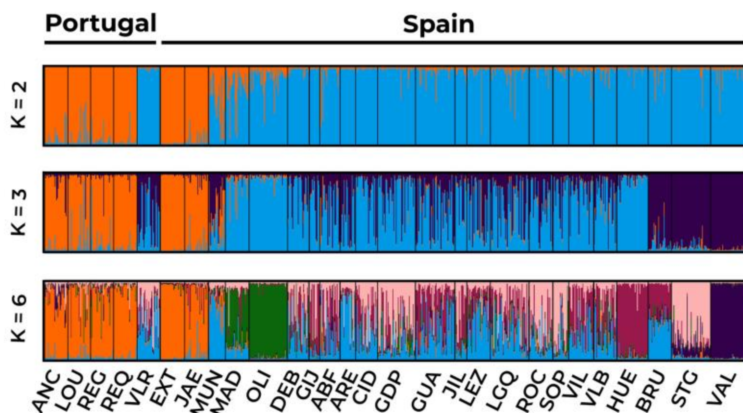
		Genetic group	**N**	**Step distance**	**Hub distance**	**Environmental suitability**
Allelic richness	Univariate	LGQ>BDJ	POS **	NEG *	NEG**	POS
	Multivariate (R^2= 0.54)		POS		NEG*	
Expected heterozygosity	Univariate	LGQ>BDJ	POS	NEG	NEG *	POS
	Multivariate (R^2= 0.13)				NEG *	

763 (*) $p < 0.05$; (**) $p < 0.01$
764

765 **vi. Figure captions**

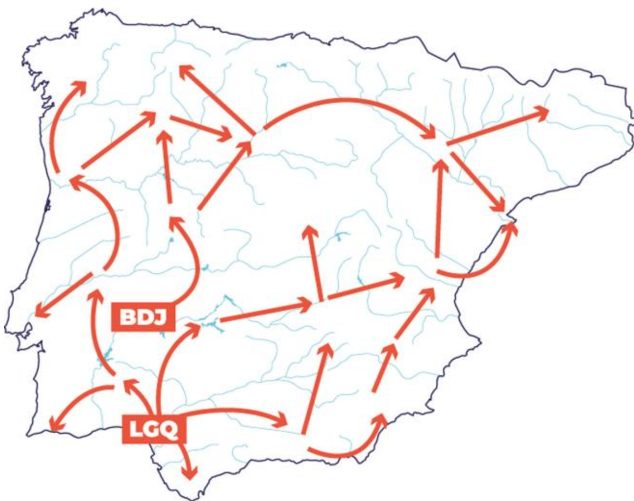
766 Figure 1. Genetic structure of the red swamp crayfish in the Iberian Peninsula, as resulting from
767 STRUCTURE outputs. The upper map shows the spatial distribution of the 28 locations and the
768 proportion of association to each of the genetic clusters defined for the most plausible K value (K =
769 3). Lower panels show the probability of assignment of red swamp crayfish individuals to the
770 genetic clusters defined for plausible K values (K = 2, K = 3 and K = 6, after the ΔK method, Fig.
771 S2). In these panels, each vertical line represents an individual, with individuals being grouped by
772 locations (codes as in Table 1), and genetic clusters are represented by different colours.

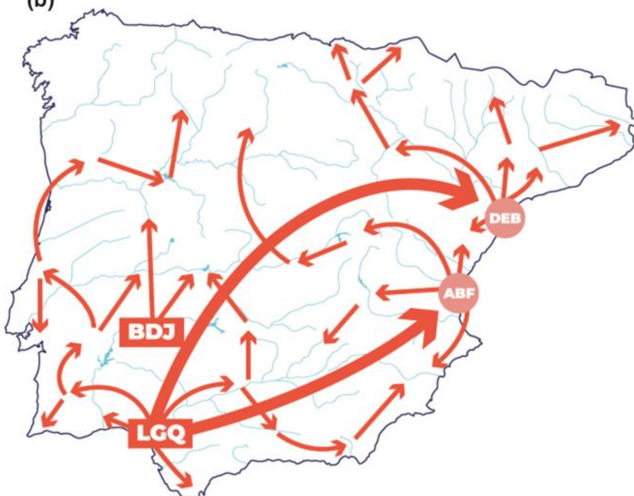




773
774

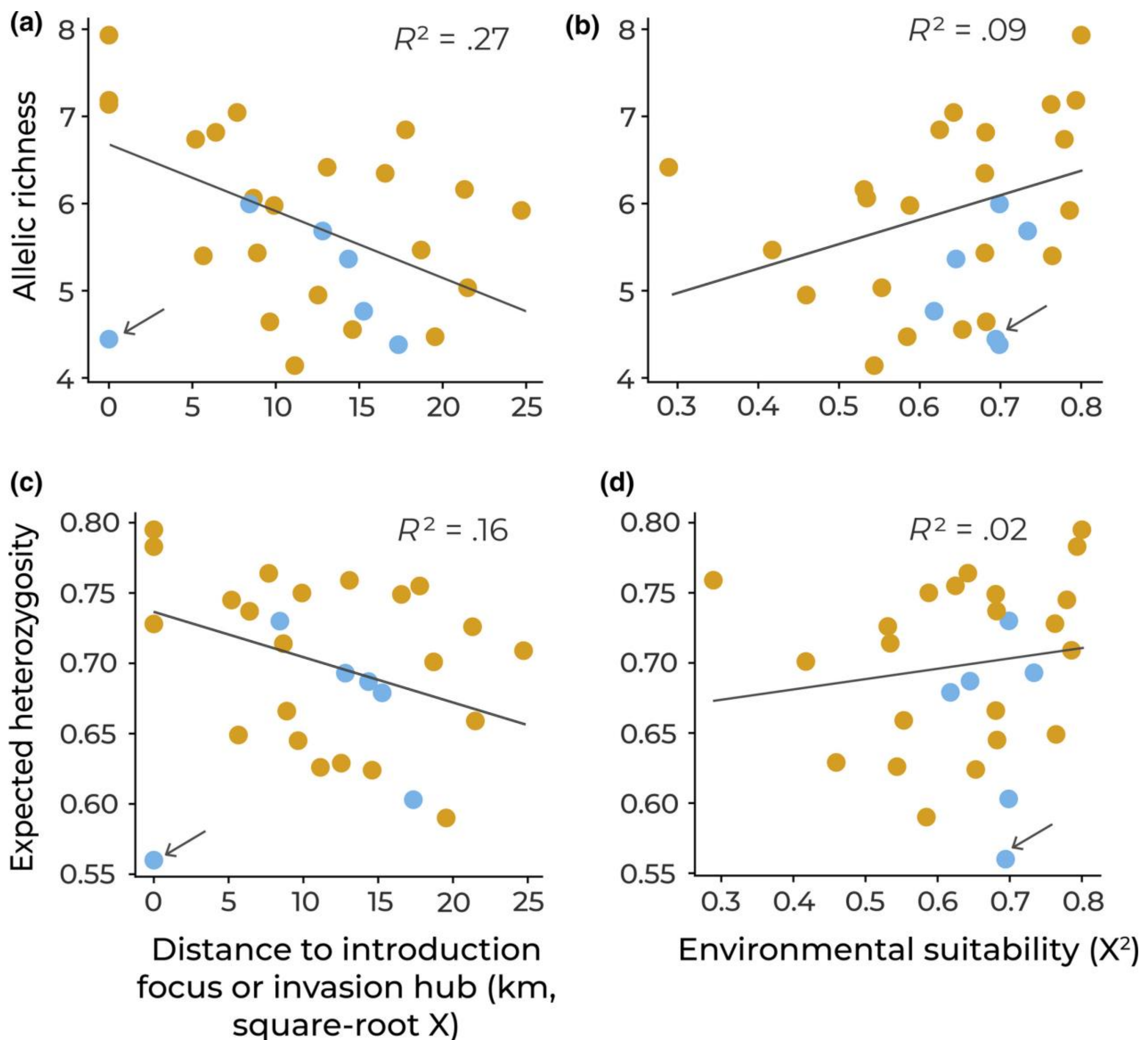
775 Figure 2. Schematic representation of plausible dispersal patterns of the red swamp crayfish
776 (*Procambarus clarkii*) across the Iberian Peninsula. In the jump-dispersal scenario (A), the
777 accumulation of bottlenecks due to successive introduction events would involve a reduction of
778 genetic diversity with increasing distance from the introduction focus (BDJ, LGQ). Contrastingly,
779 in the invasion hub scenario (B), long-distance transport of genetically diverse crayfish batches
780 (e.g., due to high propagule pressure, which is represented by arrow thickness) could have
781 generated invasion hubs (ABF, DEB), acting as additional sources for secondary introductions.
782 Thus, genetic diversity would decrease with increasing distances to either original introduction foci
783 or to invasion hubs.





784

785 Figure 3. Genetic diversity indices (allelic richness and expected heterozygosity) of the 28 red
786 swamp crayfish locations in relation to: i) the minimum distance of each location to its respective
787 introduction focus (Badajoz or Lower Guadalquivir, see Figure 1) or the closest invasion hub
788 (Valencia Albufera or Ebro Delta) (left panels); and ii) the environmental suitability of the area
789 occupied by those locations (right panels). Each location is assigned to one of the two genetic
790 groups identified (Badajoz group, blue dots; Lower Guadalquivir group, orange dots). Linear
791 regression lines and associated coefficients of determination for the two genetic groups pooled and
792 with the Badajoz population excluded (marked in all panels with an arrow, see results) are also
793 shown.



794

795 **Supplementary Material**

796 Table S1. Characteristics of the 8 polymorphic microsatellite loci for the red swamp crayfish. Ta =
797 annealing Temperature (ºC); N_A = number of observed alleles, H_O = observed heterozygosity and
798 H_E = expected heterozygosity.

Locus	Ta (ºC)	Primer concentration	N_A	H_O/H_E	GeneBank accession no.
Multiplex 1					
PCSH38	60	0.4	8	0.606/0.613	KJ607979
PCSH65	55	0.4	14	0.629/0.673	KJ607985
PclG-15	60	0.8	21	0.807/0.823	AF290227
PclG-48	51	0.8	14	0.659/0.685	AF290241
Multiplex 2					
PCSH02	60	0.8	18	0.739/0.812	KP675952
PCSH06	60	0.4	9	0.666/0.682	KP675956
PclG-17	50	0.4	14	0.674 /0.687	AF290229
PclG-29	50	0.4	10	0.311/0.641	AF290934

799

800

801 Table S2. Genetic diversity values of the 28 red swamp crayfish locations for seven microsatellite

802 loci. N_A = number of alleles; N_E = number of effective alleles; H_O = observed heterozygosity; H_E =

803 expected heterozygosity; HW = P-values for deviation of Hardy-Weinberg equilibrium; F_{IS} =

804 fixation index (positive value indicates homozygosity excess); A_R = mean allelic richness and N =

805 number individuals. Significant values of deviation of HW after Bonferroni correction are indicated

806 in bold.

		Microsatellite loci						
Locations		**PCSH02**	**PCSH06**	**PCSH38**	**PCSH65**	**PclG-15**	**PclG-17**	**PclG-48**
ABF	N_A	13	5	6	6	13	10	7
	N_E	9.324	2.295	2.370	3.173	9.398	4.881	3.189
	H_O	0.769	0.615	0.708	0.600	0.920	0.846	0.692
	H_E	0.893	0.564	0.578	0.685	0.894	0.795	0.686
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.138	-0.090	-0.225	0.124	-0.030	-0.064	-0.009
ANC	N_A	6	4	4	9	9	5	6
	N_E	3.061	2.946	3.704	6.818	4.045	2.145	2.525
	H_O	0.300	0.533	0.700	0.700	0.833	0.600	0.567
	H_E	0.673	0.661	0.730	0.853	0.753	0.534	0.604
	HW	**0.000**	1.000	1.000	**0.000**	1.000	1.000	1.000
	F_{IS}	0.554	0.193	0.041	0.180	-0.107	-0.124	0.062
ARE	N_A	7	4	3	3	9	3	5
	N_E	3.374	2.036	1.802	2.532	3.587	1.831	3.265
	H_O	0.474	0.500	0.550	0.600	0.650	0.500	0.700
	H_E	0.704	0.509	0.445	0.605	0.721	0.454	0.694
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.327	0.017	-0.236	0.008	0.099	-0.102	-0.009
BDJ	N_A	8	3	4	4	6	5	7
	N_E	3.331	1.495	2.364	1.783	1.981	3.599	2.925
	H_O	0.484	0.194	0.613	0.161	0.258	0.839	0.710
	H_E	0.700	0.331	0.577	0.439	0.495	0.722	0.658
	HW	**0.000**	**0.000**	1.000	**0.02**	**0.000**	**0.000**	1.000
	F_{IS}	0.309	0.415	-0.062	0.633	0.479	-0.161	-0.078
BRU	N_A	5	5	2	4	8	5	5
	N_E	3.435	3.579	1.965	3.152	7.171	1.861	1.796

	H_O	0.667	0.800	0.467	0.633	0.800	0.500	0.467
	H_E	0.709	0.721	0.491	0.683	0.861	0.463	0.443
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.060	-0.110	0.050	0.072	0.070	-0.080	-0.053
CID	N_A	11	7	6	4	14	7	4
	N_E	7.313	6.395	2.930	3.617	6.810	2.717	2.993
	H_O	0.828	0.897	0.793	0.759	0.862	0.586	0.655
	H_E	0.863	0.844	0.659	0.724	0.853	0.632	0.666
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.041	-0.063	-0.204	-0.048	-0.010	0.072	0.016
DEB	N_A	12	7	6	6	13	9	8
	N_E	6.588	3.613	3.409	3.970	8.340	4.854	4.545
	H_O	0.929	0.536	0.679	0.714	0.893	0.821	0.750
	H_E	0.848	0.723	0.707	0.748	0.880	0.794	0.780
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	-0.095	0.259	0.040	0.045	-0.014	-0.035	0.038
GDP	N_A	11	5	5	4	14	9	7
	N_E	5.405	2.539	2.303	3.347	9.182	3.010	4.052
	H_O	0.860	0.660	0.560	0.700	0.857	0.680	0.680
	H_E	0.815	0.606	0.566	0.701	0.891	0.668	0.753
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	-0.055	-0.089	0.010	0.002	0.038	-0.018	0.097
GIJ	N_A	8	5	5	4	7	7	6
	N_E	3.388	3.414	4.174	2.043	5.541	4.173	3.347
	H_O	0.667	0.923	0.833	0.500	0.923	0.692	0.769
	H_E	0.705	0.707	0.760	0.510	0.820	0.760	0.701
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.054	-0.305	-0.096	0.020	-0.126	0.089	-0.097
GUA	N_A	13	8	6	7	15	9	8
	N_E	6.964	4.227	2.804	3.597	7.859	3.870	3.953
	H_O	0.700	0.740	0.580	0.640	0.592	0.700	0.720
	H_E	0.856	0.763	0.643	0.722	0.873	0.742	0.747
	HW	1.000	1.000	1.000	1.000	**0.000**	1.000	1.000
	F_{IS}	0.183	0.031	0.099	0.114	0.322	0.056	0.036
HUE	N_A	10	5	5	6	9	8	7
	N_E	5.452	3.674	2.875	3.879	4.520	4.161	4.414
	H_O	0.897	0.675	0.575	0.750	0.816	0.725	0.750
	H_E	0.817	0.728	0.652	0.742	0.779	0.760	0.773
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	F_{IS}	-0.099	0.073	0.118	-0.011	-0.048	0.046	0.030
JAE	N_A	7	4	2	5	7	5	6
	N_E	3.000	2.233	1.980	3.303	3.442	3.377	1.651
	H_O	0.567	0.467	0.233	0.667	0.567	0.600	0.433
	H_E	0.667	0.552	0.495	0.697	0.709	0.704	0.394
	HW	1.000	1.000	0.764	1.000	1.000	1.000	1.000
	F_{IS}	0.150	0.155	0.529	0.044	0.201	0.148	-0.099
JIL	N_A	8	5	4	4	10	8	7
	N_E	7.143	3.600	2.432	3.261	7.500	4.639	4.891
	H_O	0.733	0.800	0.600	0.533	0.933	0.867	0.600
	H_E	0.860	0.722	0.589	0.693	0.867	0.784	0.796
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.147	-0.108	-0.019	0.231	-0.077	-0.105	0.246
LEZ	N_A	12	7	7	5	12	7	8
	N_E	5.422	3.550	3.141	3.352	8.824	4.423	3.409
	H_O	0.667	0.667	0.700	0.733	0.900	0.667	0.667
	H_E	0.816	0.718	0.682	0.702	0.887	0.774	0.707
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.183	0.072	-0.027	-0.045	-0.015	0.139	0.057
LGQ	N_A	16	7	8	6	16	10	11
	N_E	11.267	5.284	2.853	3.137	9.596	5.251	5.166
	H_O	0.917	0.771	0.653	0.739	0.913	0.826	0.688
	H_E	**0.911**	0.811	0.650	0.681	0.896	0.810	0.806
	HW	1.000	1.000	1.000	1.000	1.000	1.000	0.412
	F_{IS}	-0.006	0.049	-0.005	-0.085	-0.019	-0.020	0.147
LOU	N_A	10	4	4	5	9	8	7
	N_E	5.028	2.317	2.503	3.719	4.027	4.557	2.635
	H_O	0.667	0.533	0.633	0.700	0.867	0.767	0.633
	H_E	0.801	0.568	0.601	0.731	0.752	0.781	0.621
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.168	0.062	-0.055	0.043	-0.153	0.018	-0.021
MAD	N_A	8	6	4	5	10	5	6
	N_E	4.800	3.854	2.568	3.758	5.263	1.883	4.054
	H_O	0.867	0.733	0.667	0.800	0.833	0.567	0.733
	H_E	0.792	0.741	0.611	0.734	0.810	0.469	0.753
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	-0.095	0.010	-0.092	-0.090	-0.029	-0.209	0.027
MUN	N_A	8	3	5	5	5	4	7

	N_E	6.083	1.411	3.885	2.321	2.818	2.513	3.556
	H_O	0.857	0.333	0.571	0.571	0.857	0.524	0.714
	H_E	0.836	0.291	0.743	0.569	0.645	0.602	0.719
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	-0.026	-0.144	0.231	-0.004	-0.329	0.130	0.006
OLI	N_A	8	4	4	5	6	4	10
	N_E	4.259	2.889	2.948	2.348	2.495	1.888	4.873
	H_O	0.792	0.625	0.714	0.612	0.571	0.375	0.740
	H_E	0.765	0.654	0.661	0.574	0.599	0.470	0.795
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	-0.035	0.044	-0.081	-0.066	0.046	0.203	0.069
REG	N_A	10	4	3	5	10	7	9
	N_E	3.782	3.651	2.965	3.516	6.618	5.678	2.483
	H_O	0.700	0.700	0.567	0.667	0.833	0.833	0.567
	H_E	0.736	0.726	0.663	0.716	0.849	0.824	0.597
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.048	0.036	0.145	0.068	0.018	-0.011	0.051
REQ	N_A	8	4	3	4	7	6	5
	N_E	5.663	2.353	2.605	2.323	3.750	4.369	3.000
	H_O	0.724	0.567	0.700	0.533	0.533	0.833	0.633
	H_E	0.823	0.575	0.616	0.569	0.733	0.771	0.667
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.121	0.014	-0.136	0.063	0.273	-0.081	0.050
ROC	N_A	12	8	5	4	13	8	10
	N_E	7.531	4.157	2.597	2.476	7.595	2.985	4.724
	H_O	0.800	0.767	0.567	0.533	0.900	0.633	0.800
	H_E	0.867	0.759	0.615	0.596	0.868	0.665	0.788
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.078	-0.010	0.079	0.105	-0.036	0.048	-0.015
SOP	N_A	8	5	3	3	9	6	7
	N_E	4.255	2.941	1.512	2.228	5.063	2.899	4.324
	H_O	0.750	0.550	0.400	0.300	0.900	0.750	0.700
	H_E	0.765	0.660	0.339	0.551	0.803	0.655	0.769
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.020	0.167	-0.181	0.456	-0.121	-0.145	0.089
STG	N_A	12	6	6	7	16	11	10
	N_E	8.961	2.836	2.374	3.030	9.560	5.086	3.751
	H_O	0.860	0.580	0.760	0.760	0.840	0.880	0.760
	H_E	0.888	0.647	0.579	0.670	0.895	0.803	0.733

	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.032	0.104	-0.313	-0.134	0.062	-0.095	-0.036
VAL	N_A	4	3	3	8	7	3	5
	N_E	2.629	2.162	2.251	3.909	5.482	1.888	2.734
	H_O	0.520	0.640	0.531	0.440	0.900	0.580	0.620
	H_E	0.620	0.537	0.556	0.744	0.818	0.470	0.634
	HW	1.000	1.000	1.000	**0.000**	1.000	1.000	1.000
	F_{IS}	0.161	-0.191	0.045	0.409	-0.101	-0.234	0.022
VIL	N_A	10	5	5	4	12	5	4
	N_E	7.813	4.029	2.860	1.816	5.636	3.013	1.812
	H_O	0.871	0.710	0.625	0.438	0.806	0.581	0.375
	H_E	0.872	0.752	0.650	0.449	0.823	0.668	0.448
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.001	0.056	0.039	0.026	0.020	0.131	0.163
VLB	N_A	11	6	5	4	12	7	6
	N_E	5.533	3.482	1.970	3.318	7.250	4.335	3.672
	H_O	0.690	0.690	0.552	0.759	0.828	0.862	0.793
	H_E	0.819	0.713	0.492	0.699	0.862	0.769	0.728
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.158	0.033	-0.121	-0.086	0.040	-0.121	-0.090
VLR	N_A	7	6	3	4	13	4	6
	N_E	5.751	4.604	1.795	2.293	4.083	2.571	2.735
	H_O	0.800	0.800	0.467	0.567	0.724	0.567	0.655
	H_E	0.826	0.783	0.443	0.564	0.755	0.611	0.634
	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	F_{IS}	0.032	-0.022	-0.054	-0.005	0.041	0.073	-0.033
MEAN	N_A	31.964	32.107	32.071	32.036	31.857	32.036	32.179
	N_E	9.393	5.179	4.500	5.000	10.393	6.607	6.929
	H_O	5.591	3.270	2.641	3.143	5.980	3.516	3.446
	H_E	0.727	0.643	0.607	0.611	0.790	0.686	0.663
	HW	0.795	0.656	0.600	0.655	0.800	0.677	0.682
	F_{IS}	0.091	0.023	-0.017	0.074	0.018	-0.016	0.024

807

808

809 Table S3. P-value for Wilcoxon's test for heterozygosity excess conducted in Bottleneck for 28 red

810 swamp crayfish locations in the Iberian Peninsula. TPM: two-phased model of mutation, SMM:

811 stepwise mutation model.

812

Locations (code)	**P-value of Wilcoxon test**	
	T.P.M.	**S.M.M.**
Albufera (ABF)	0.3750	0.3750
Ança (ANC)	0.1094	0.4688
Arreo (ARE)	0.4688	0.5781
Brugent (BRU)	0.0078 **	1.0000
Cidacos (CID)	0.0078 **	0.0781
Delta del Ebro (DEB)	0.2969	0.8125
Badajoz (BDJ)	**0.0078 ****	**0.0078 ****
Guadalporcun (GDP)	0.0156 *	0.6875
Gijón (GIJ)	0.0781	0.5781
Guadiamar (GUA)	0.0547	0.8125
Hueznar (HUE)	0.0391 *	0.3750
Jaen (JAE)	0.0078 **	0.2969
Jiloca (JIL)	0.2969	0.8125
Leza (LEZ)	0.0156 *	0.4688
Lower Guadalquivir (LGQ)	0.2969	0.4688
Lousal (LOU)	0.5781	0.0781
Madrid (MAD)	0.0391 *	0.9375
Mundo (MUN)	0.0078 **	0.0547
Olivargas (OLI)	0.8125	0.1094
Reguengos (REG)	0.8125	0.3750
Requeixo (REQ)	0.0234 *	0.9375
Rocina (ROC)	0.0781	0.3750
Sopetón (SOP)	0.5781	0.8125
Sotogrande (STG)	0.0078 **	1.0000
Valle (VAL)	0.0156 **	1.0000
Villar (VIL)	0.3750	0.1094
Valoria la buena (VLB)	0.0234 *	0.4688
Vila-Rica (VLR)	0.0078 **	0.0781

813 (*) $p < 0.05$; (**) $p < 0.01$

814

815 Table S4. Proportion of association of the red swamp crayfish studied locations to each of the three
816 genetic clusters identified by the Bayesian analysis method implemented in STRUCTURE software.
817 Populations are ordered by proportion of association to each cluster and at the end is Mundo, an
818 admixed population. Populations are assigned to genetic group taking into account the introduction
819 focus and the genetic cluster.

Location	Code	Cluster 1	Cluster 2	Cluster 3	Gentic group
Ança	ANC	0.875	0.033	0.092	1
Badajoz	BDJ	0.98	0.009	0.011	1
Jaen	JAE	0.896	0.059	0.045	1
Lousal	LOU	0.832	0.127	0.041	1
Reguengos de Monzaraz	REG	0.861	0.095	0.044	1
Requeixo	REQ	0.947	0.031	0.023	1
Cidacos	CID	0.018	0.745	0.237	2
Delta del Ebro	DEB	0.027	0.746	0.227	2
Guadalporcun	GDP	0.019	0.745	0.236	2
Gijón	GIJ	0.019	0.697	0.284	2
Guadiamar	GUA	0.033	0.520	0.446	2
Hueznar	HUE	0.019	0.926	0.055	2
Jiloca	JIL	0.027	0.738	0.235	2
Leza	LEZ	0.058	0.540	0.403	2
Lower Guadalquivir	LGQ	0.044	0.573	0.383	2
Madrid	MAD	0.099	0.868	0.033	2
Olivargas	OLI	0.031	0.946	0.023	2
Rocina	ROC	0.026	0.592	0.381	2
Sopetón	SOP	0.021	0.747	0.231	2
Villar	VIL	0.028	0.753	0.219	2
Valoria la buena	VLB	0.04	0.647	0.312	2
Albufera	ABF	0.065	0.514	0.421	2
Arreo	ARE	0.019	0.445	0.536	2
Brugent	BRU	0.034	0.089	0.877	2
Sotogrande	STG	0.022	0.045	0.932	2
Valle	VAL	0.014	0.019	0.967	2
Vila-Rica	VLR	0.014	0.423	0.563	2
Mundo	MUN	0.316	0.235	0.449	2

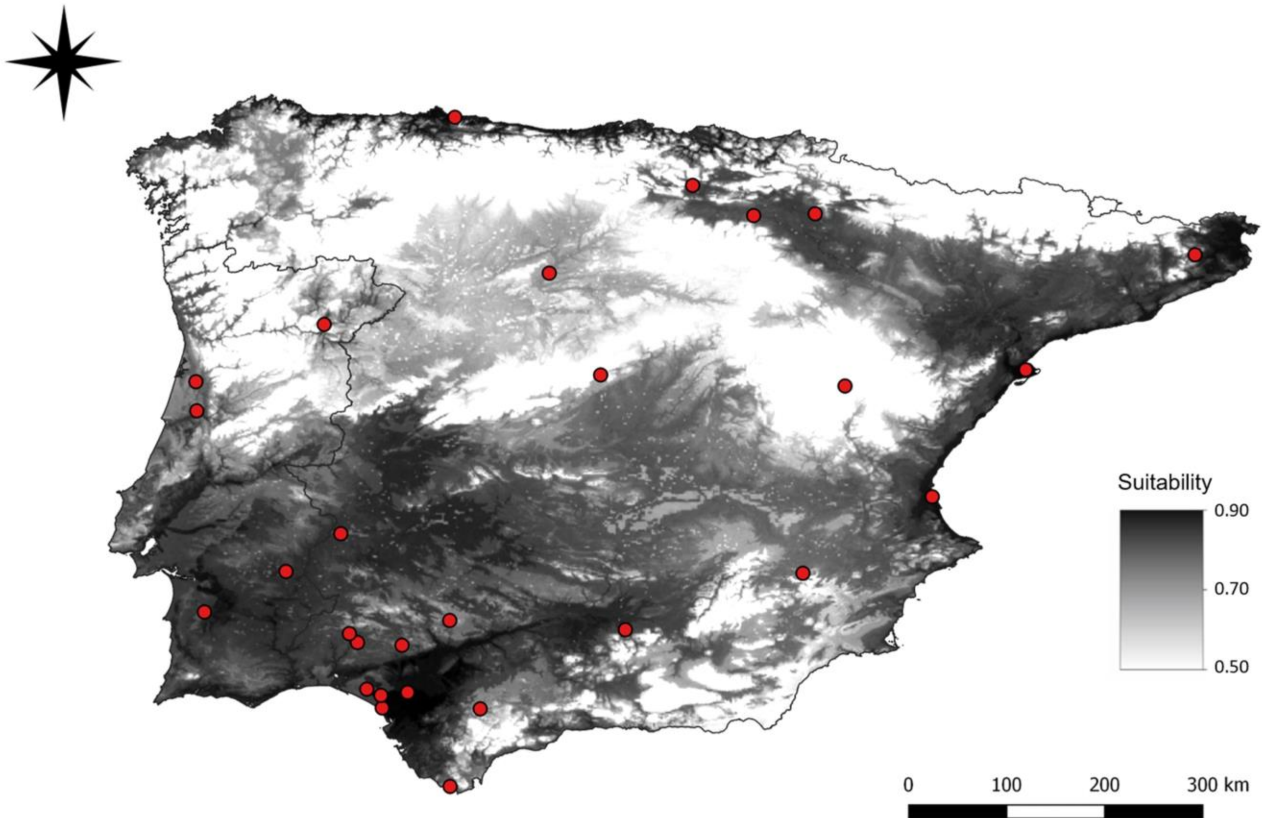
820
821

822 Table S5. The full univariate (a,b) and multivariate (c,d) general linear models assessing the
823 influence of different predictors on the estimators of genetic diversity (allelic richness and expected
824 heterozygosity). The coefficient, p-value as an indication of the statistical significance and
825 coefficient of determination are shown for each model.

Candidate model	Coefficient	p-value	R^2
	Univariate model		
a) Allelic richness			
Genetic group	-0.351	0.160	0.08
N	0.012	0.530	0.02
Step distance	-0.020	0.400	0.03
Hub distance	**-0.078**	**0.005**	**0.27**
Suitability	2.437	0.130	0.09
b) Expected heterozygosity			
Genetic group	-0.013	0.360	0.03
N	0.001	0.370	0.03
Step distance	-0.001	0.480	0.02
Hub distance	**-0.003**	**0.036**	**0.16**
Suitability	0.072	0.440	0.02
	Multivariate model		
c) Allelic richness			
Genetic group	-0.305		
N	-0.008		
Step distance	0.007		
Hub distance	**-0.074**	**0.001**	**0.54**
Suitability	1.286		
d) Expected heterozygosity			
Genetic group	-0.008		
N	0.000		
Step distance	0.001		
Hub distance	**-0.004**	**0.007**	**0.13**
Suitability	0.008		

826
827
828

829 Figure S1. Suitability model for the red swamp crayfish in the Iberian Peninsula provided by
830 Capinha & Anastácio, 2011. To better visualization of the difference in environmental suitability
831 across the Peninsula, the intensity of grey colour was adjusted to the range of suitability values
832 present in our data (from 0.50 to 0.90). The 28 locations are depicted in red dots, including the 5 km
833 buffer.



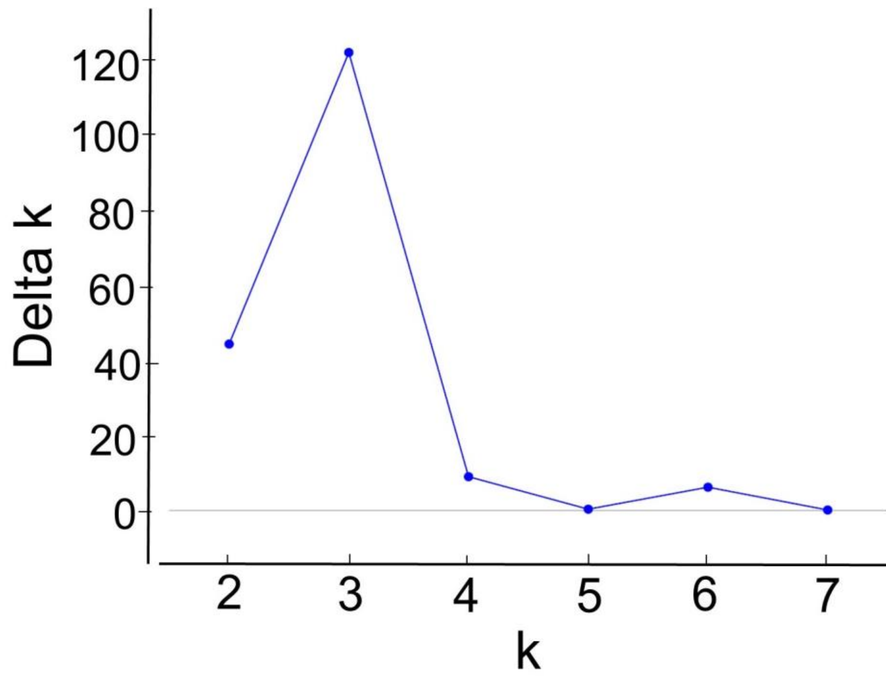
834

835

836 Figure S2. The most likely value of real population clusters in the genetic dataset estimated by the

837 ΔK method (Evanno et al., 2005) after STRUCTURE results. Values of ΔK, showing peaks at K =

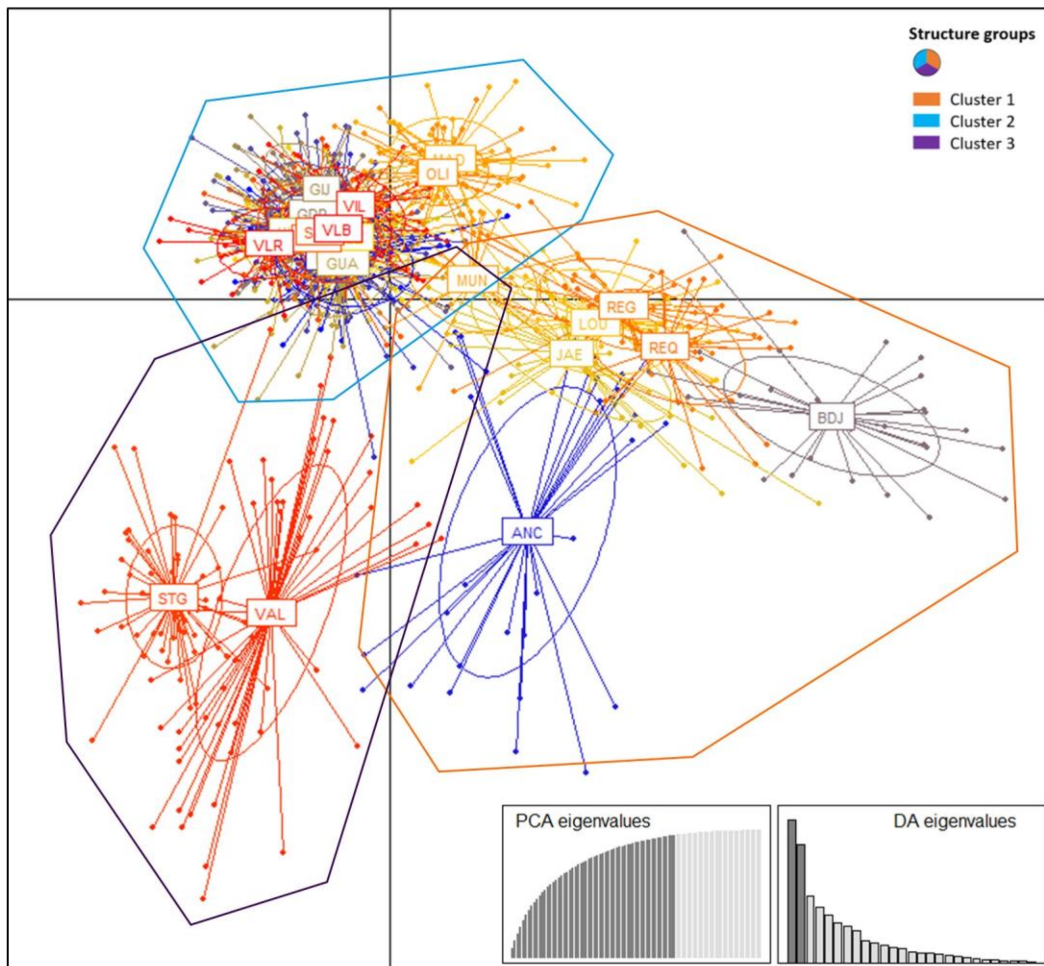
838 2, 3 and 6.



839

840

841 Figure S3. Discriminant analysis of principal components (DAPC) for the red swamp crayfish
842 locations in the Iberian Peninsula. The graph represents the individuals as dots and the populations
843 as inertia ellipses. PCA and DA eigenvalues are displayed in the inset where the retained
844 eigenvalues are depicted in black and the number of bars represents the number of discriminant
845 functions that retained in the analysis, respectively. See Table 1 for location.



846