

The influence of native populations' genetic history on the reconstruction of invasion routes: the case of a highly invasive aquatic species

Brazier Thomas ^{1,2}, Cherif Emira ^{3,*}, Martin Jean-François ⁴, Gilles André ⁵, Blanchet Simon ⁶, Zhao Yahui ⁷, Combe Marine ³, McCairns R. J. Scott ¹, Gozlan Rodolphe E. ³

¹ DECOD (Ecosystem Dynamics and Sustainability), INRAE, Institut Agro, IFREMER, Rennes, France

² University of Rennes, CNRS, ECOBIO (Ecosystems, Biodiversity, Evolution) - UMR 6553, Rennes, France

³ ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France

⁴ CBGP, Montpellier SupAgro, INRA, CIRAD, IRD, Université Montpellier, Montpellier, France

⁵ UMR 1467 RECOVER, INRAE, Aix Marseille University, Centre St Charles, 3 place Victor Hugo, 13331, Marseille, France

⁶ CNRS, Station d'Ecologie Théorique Et Expérimentale (SETE), Moulis, France

⁷ Institute of Zoology, Chinese Academy of Sciences, Chaoyang District, Beijing, 100101, China

* Corresponding author : Emira Cherif, email address : emira.cherif@ird.fr

Abstract :

Insufficient data on the origins of the first introduced propagule and the initial stages of invasion complicate the reconstruction of a species' invasion history. Phylogeography of the native area profoundly shapes the genomic patterns of the propagules on which subsequent demographic processes of the invasion are based. Thus, a better understanding of this aspect helps to disentangle native and invasive histories. Here, we used genomic data together with clustering methods, explicit admixture tests combined with ABC models and Machine Learning algorithms, to compare patterns of genetic structure and gene flow of native and introduced populations, and infer the most likely invasion pathways of the highly invasive freshwater fish *Pseudorasbora parva*. This species is the vector of a novel lethal fungal-like pathogen (*Sphaerothecum destruens*) that is responsible for the decline of several fish species in Europe. We found that the current genetic structuring in the native range of *P. parva* has been shaped by waves of gene flow from populations in southern and northern China. Furthermore, our results strongly suggest that the genetic diversity of invasive populations results from recurrent global invasion pathways of admixed native populations. Our study also illustrates how the combination of admixture tests, ABC and Machine Learning can be used to detect high-resolution demographic signatures and reconstruct an integrative biological invasion history.

Keywords : Invasion pathways, Aquatic biological invasion, Demographic inference, Population modeling, Population genomics, Approximate bayesian computation

69 **1 INTRODUCTION**

70 Invasive species are significant contributors to global change and often lead to biotic
71 homogenization and biodiversity losses (Chapin III et al. 2000; Clavero & García-Berthou
72 2005; Didham et al. 2005; Villéger et al. 2011; Simberloff 2013). Globalization of trade and
73 growth of worldwide transportation are the main drivers of non-native species introductions
74 (Hulme 2009). Major ecological impacts from biological invasions include destabilization of
75 trophic networks (Stiers et al. 2011; Gallardo et al. 2016), competition for resources and
76 habitats (Graebner et al. 2012; Perdereau et al. 2011), predation on native species (Salo et
77 al. 2007) and the transmission of novel infectious pathogens (Crowl et al. 2008). Invasions
78 can also have evolutionary consequences, such as genome introgression or altering
79 selection pressures on native species (Crispo et al. 2011; Mooney & Cleland 2001; Sinama
80 et al. 2013; Philips et al. 2006). Yet invasions offer an incredible framework to study
81 adaptation to new environmental conditions and to understand how small introduced
82 populations with a supposedly low level of genetic diversity manage to colonize large areas
83 (Bossdorf et al. 2005; Peischl & Excoffier 2015; Sax et al. 2007; Roman & Darling 2007;
84 Facon et al. 2006), as well as to understand the spread and emergence of pathogens co-
85 introduced together with their non-native hosts. Reconstructing the invasion history of a
86 species is challenging. Typically, few individuals may constitute the initial colonizing group,
87 but data on the amount of propagule pressure as well as its origin and time of first
88 introduction are often lacking. Historical records of introduction events, census data or
89 environmental monitoring projects often miss the first stages of invasion (Holsbeek et al.
90 2008; Mergeay et al. 2005). Molecular ecology allows the reconstruction of events that have
91 not been directly observed but have left a genomic signature (Cristescu 2015). Through the
92 use of new demographic inference methods that permit deviations from the assumptions of
93 the mutation-drift equilibrium model (which is often the case for recent colonization events), it
94 is now possible to identify source populations or even to model the short-term demo-genetic
95 processes underpinning invasion history (Beichman et al. 2018; Cabrera & Palsbøll 2017;
96 Shafer et al. 2015; Estoup & Guillemaud 2010). Recent population genetic studies on

97 invasion pathways have shown that invasion patterns may be more complex than previously
98 thought (Lombaert et al. 2014). A single introduction event is rare, and multiple introductions
99 and/or admixture between source populations are often the cornerstone of successful
100 invasions (Estoup et al. 2016; Roman & Darling 2007). Since introduced populations inherit
101 genetic variation primarily shaped by long-term evolutionary processes in the native area,
102 incorporating this information in the interpretation of the results is important to avoid
103 erroneous conclusions (Rius & Turon 2020).

104 The topmouth gudgeon, *Pseudorasbora parva* (Temminck & Schlegel), is a small freshwater
105 cyprinid fish. More importantly, it is an interesting study system for biological invasion for
106 several reasons: i) it is a highly successful invader that has spread and established very
107 quickly in diverse environments throughout the world, ii) it has a well-documented history
108 and iii) its invasion is associated with a pathogen co-introduction. It also has a very large
109 native distribution in Eastern Asia (Eastern China, Taiwan, Korea and Japan) with broad
110 environmental tolerance from continental climates to tropical ones (Gozlan 2012; Zhang &
111 Zhao 2016). It was initially accidentally released outside its native range via aquaculture
112 partnership exchanges of Chinese carp between China and former countries of the Soviet
113 Bloc (Gozlan et al. 2010). During the 1960s, multiple introductions of *P. parva* took place all
114 around the Black Sea area, followed by further introductions in the 80s in Eurasia and North
115 Africa. However, introduction records lack precise information on the source populations.
116 After these initial phases of human-made introductions, natural local colonization of entire
117 river networks occurred across major European rivers to the Middle East. Its life-history
118 traits, which include early maturity (1 year) coupled with nest guarding behavior that ensures
119 an increased likelihood of survival for juveniles and short longevity, have been identified as
120 key traits to explain its rapid establishment and spread (Gozlan et al. 2010; Gozlan 2012,
121 Gozlan et al. 2020), although its long-term invasion success depends on high genetic
122 diversity to drive adaptation to the new environment (Roman & Darling 2007). A major
123 biodiversity problem associated with *P. parva* is the fungus-like pathogen, the rosette agent
124 *Sphaerothecum destruens*, for which it acts as a vector, and that has subsequently spread

125 to invaded areas causing mortality in a large number of native freshwater fish species in
126 Europe (Andreou & Gozlan 2016; Combe & Gozlan 2018).
127 Genetic structure and phylogeography, based upon classical mitochondrial and nuclear
128 markers (microsatellites), have been extensively described in both native and invasive *P.*
129 *parva* populations, suggesting the existence of two genetic lineages within non-native
130 populations in Europe (Hardouin et al. 2018; Simon et al. 2011, 2015). This general pattern
131 has been recently confirmed by a study using 13,785 single nucleotide polymorphisms in
132 Slovakian and Turkish introduced populations (Baltazar-Soares et al. 2020). Yet, the fine-
133 scale genetic structure of the native populations and the gene flow that has shaped it, as
134 well as the demographic dynamics underpinning the invasion pathways, remain unresolved.
135 Assessing the source populations within the native range by extensive sampling, along with
136 high-throughput genotyping, is crucial to detect high-resolution signatures of demographic
137 history (Muirhead et al. 2008; Beichman et al. 2018; Shafer et al. 2015). Although such large
138 datasets require advanced statistical methods that are computationally demanding,
139 Approximate Bayesian Computation (ABC) and associated Machine Learning algorithms
140 take advantage of large sets of summary statistics, exploring huge parameter spaces with
141 reduced computational effort (Beaumont et al. 2002; Cabrera & Palsbøll 2017; Pudlo et al.
142 2016; Raynal et al. 2017; Rey et al. 2015).
143 Here we took advantage of genomic signatures, ABC models and Machine Learning
144 algorithms to i) characterize the genetic structure of native populations and test whether they
145 aggregate on homogeneous and coherent demes and ii) retrace *P. parva*'s introduction
146 history from Asia to Europe. To do so, we first applied genetic clustering methods and
147 explicit tests of admixture to describe the phylogeography of both native and invasive
148 ranges. We then identified putative source populations of introduced demes with population
149 assignment tests. We finally used ABC model-based procedures to infer the most probable
150 demographic scenarios of *P. parva* invasion and reconstruct an integrative biological
151 invasion history.

152 **2 MATERIALS AND METHODS**

153 **2.1 Sampling material**

154 We sampled *P. parva* from 21 discrete sites across its overall distribution in Asia, comprising
155 sixteen different river catchments in the historical Chinese native range (Figure S1, Table 1,
156 sites numbered from S1 to S18). Additional samples were obtained from Japan and
157 neighboring Asian invasive populations in Tibet and South-East China (sites: S19, S20). We
158 also sampled *P. parva* across thirteen locations within the invasive European and Middle-
159 Eastern range (Figure S1, Table 1). The sampling methods included fish traps, electric
160 fishing and micro-mesh seine netting where appropriate. In the field, the fish were
161 euthanized with an overdose of anesthetic, initially preserved in ethanol and subsequently
162 stored at -70 °C. Each fish was measured and a fin-clip taken and stored in ethanol.

163 **2.2 GBS sequencing**

164 In total, 858 DNA samples (746 individuals in total, with 75 individuals replicated across at
165 least one sequencing lane) were genotyped for single nucleotide polymorphism (SNP)
166 markers by first digesting genomic DNA with PstI, followed by genotyping-by-sequencing
167 (GBS), yielding an average of 2,702,000 raw sequencing reads per sample. SNP calling was
168 performed using programs comprising the Stacks (v1.46) bioinformatics *de novo* pipeline
169 (Catchen et al. 2013). Replicated individuals were used to estimating genotyping error and
170 only markers with less than 1% errors were retained. A minimum read depth of 20 was
171 required for each marker, ultimately yielding 3999 validated SNP markers. To reduce linkage
172 disequilibrium amongst markers, only a single SNP (first position) was called for each locus
173 stack. To prevent biases due to large proportions of missing data, populations with more
174 than 70% missing data, individuals with more than 60% missing data and loci with more than
175 45% of missing data were removed from the dataset. After trimming, the final dataset
176 contained 300 individuals from eighteen sites in the historic Asian range (including invasive
177 populations in Tibet and South-West China) and 168 individuals from eleven invasive sites
178 across Europe, Turkey and Iran, with on average 16 individuals per sampled site (Figure S1,
179 Table 1).

180 **2.3 Genetic diversity**

181 Statistics of genetic diversity were estimated with the 'adegenet' and 'hierfstat' R packages
182 (Goudet 2005; Goudet & Jombart 2015; Jombart 2008; Jombart & Ahmed 2011) for R
183 version 3.5.3 (R Core Team, 2019) (Table S1). Sensitivity analyses on summary statistics
184 showed that biases were minimized without loss of power when estimates were inferred from
185 2,000 to 3,000 loci with the lowest proportion of missing data (Figure S2). Consequently, the
186 dataset for inferring population assignment consisted of 2,112 SNPs to reduce
187 computational load (maximum 45% missing data per locus); however, up to 3000 SNPs
188 (maximum ~50% missing data) were retained for other analyses to keep most of the genetic
189 information. Simulations have shown that even loci with high levels of missing data (> 50%)
190 can retain meaningful information (Huang & Knowles 2016), as it has been empirically
191 assessed for population structure (Chan et al. 2017).

192 **2.4 Genetic clustering**

193 Genetic clustering of sampled sites was assessed independently in the Asian range (native
194 and non-native sites) and within the invasive European range. Putative demes were defined
195 as groups of individuals sharing a gene pool. Results from an iterative K-means method and
196 a model-based Bayesian clustering method were compared for cross-validation.
197 Discriminant Analysis of Principal Components (DAPC), implemented in the R package
198 'adegenet' (Jombart 2008; Jombart & Ahmed 2011) was first used to determine genetic
199 clusters within native and invasive regions and then to predict membership of individuals
200 within invasive sites to native clusters. The most probable number of genetic clusters, K, was
201 searched within the distribution of K among 1,000 independent clustering iterations, based
202 on the 'goodfit' and 'min' criteria of the Bayesian Information Criterion (BIC). Overfitting was
203 prevented by a cross-validation procedure to define the optimal number of Principal
204 Components required to discriminate amongst these K clusters.

205 Additionally, we used STRUCTURE 2.3 to infer the number of genetic clusters and to
206 estimate admixture between them (Pritchard et al. 2000). The admixture model with
207 correlated allele frequencies was parameterized with a fixed Lambda value (parameter of the

208 allele frequencies distribution) directly estimated from the data. Sampling location was set as
209 a prior to improve inferences on weak genetic structure. STRUCTURE was performed for
210 $K=1-21$ (number of sampled sites + 3) for Asian sites and $K=1-14$ for European and Middle-
211 Eastern sites. Twenty replicates were computed for each K value, with 100,000 sampling
212 iterations after a burn-in of 100,000. The most probable K was assessed after considering
213 the smallest value of K minimizing differences of likelihood (i.e. the plateau method of
214 Pritchard et al. 2000), the highest value of ΔK given using Evanno's method, computed with
215 STRUCTURE HARVESTER (Earl & vonHoldt 2012; Evanno et al. 2005) and the
216 Puechmaille statistics (Puechmaille 2016), implemented in the STRUCTURE SELECTOR
217 web interface (Li & Liu 2018). We then used CLUMPP to aggregate STRUCTURE replicates
218 to produce mean individual admixture proportions with a 'greedy' search algorithm over
219 1,000 repetitions (Jakobsson & Rosenberg 2007). Convergence among parameter sampling
220 chains was assessed via CLUMPP's H' statistic of similarity amongst several replicates.
221 Additionally, within-chain convergence for parameter estimates (i.e. ancestry coefficient
222 Alpha and Ln Likelihood) was assessed with diagnostics implemented in the R package
223 'coda' (Plummer et al. 2006).
224 Finally, based on the observed partitioning of genetic variance, sampled sites were pooled
225 into demes (also referred to as populations) that made sense biologically (i.e. continuous
226 gene flow among sites) and geographically. Admixture proportions inferred with
227 STRUCTURE helped to delineate putative demes and were computed as the mean of the
228 major ancestry coefficients (Q) in a given site. Sampled sites with a mean major ancestry
229 coefficient (Q) under 70% were considered as admixed sites. Sampled sites that did not
230 cluster well based on genetic markers were clustered into putative demes based on
231 geographical and historical data. The consistency of our final clustering of sampled sites in
232 genetic populations was assessed by a hierarchical AMOVA implemented in the 'poppr' R
233 package (Kamvar et al. 2014, 2015) and the significance of variance proportions was tested
234 with 1,000 random permutations (Table S3). Maps were drawn from the R package 'maps'
235 with the 'world' database (Becker et al. 2018). The main river network was drawn from

236 'RNaturalEarth' (South 2017).

237 **2.5 Assignment to source populations with supervised machine learning**

238 The R package 'AssignPOP' was used to assign invasive individuals to candidate source
239 populations (i.e. native demes) with a machine learning classification algorithm (Chen et al.
240 2018). The Support Vector Machine (SVM) classification algorithm was trained on a set of
241 individuals from known populations (i.e. candidate source populations in the native range),
242 and then individuals from unknown populations (i.e. invasive populations) were assigned to
243 one of the candidate source populations in the training dataset. Using a Monte Carlo
244 procedure with K-fold cross-validation, a set of known loci and individuals were randomly
245 sampled in multiple iterations to train the classification algorithm, while the remaining known
246 loci and individuals were used to iteratively test the accuracy of the predictive model. One
247 hundred training iterations were performed for different proportions of training individuals (i.e.
248 0.5, 0.7 and 0.9) and training loci (i.e. 0.25, 0.5 and 1) to select the best training sample size.
249 The training error rate was assessed as the reassignment success of known individuals to
250 their source population. Source populations of invasive individuals were predicted from
251 posterior assignment probabilities. Only individuals for whom the highest probability was at
252 least twice that of the second were retained as confident assignments.

253 **2.6 Inference of invasion history with ABC**

254 Approximate Bayesian Computation (ABC) was used to infer past demographic events
255 shaping contemporary genetic diversity (Supporting information). We simulated large
256 datasets under various invasion and admixture demographic scenarios and estimated the
257 probability that data were observed under a given demographic scenario (Estoup et al.
258 2012). Scenarios were designed in a two-step hierarchical procedure of increasing
259 complexity, with the second step derived from findings at step 1 (Figure S3, S4; see
260 supplementary method for details). In the first step, we tested 3 independent sets of
261 scenarios about the origins in the native range of 3 independent invasive demes (Western
262 Europe, Eastern Europe and Iran; see Results). For each independent invasive deme, we
263 tested if the source population was one of the three candidates or an admixture between the

264 three candidates. To avoid the trap of infinite combinations of exhaustive scenarios, source
265 populations predicted with population assignment served as initial candidate source
266 populations. Choice of candidate source populations was also cross-validated with a
267 Maximum-likelihood phylogenetic tree (Figure S8) inferred by the program TreeMix version
268 1.13 (Pickrell & Pritchard 2012). As the Italian sample's origin was ambiguous in previous
269 genetic clustering results, we performed the ABC analysis with and without Italian samples.
270 The comparison of both replicated scenarios allowed us to assess the sensitivity of the
271 selected scenario to Italian individuals (Table 2). The second step was dedicated to
272 resolving a more complex invasion pattern, comparing competing worldwide invasion
273 pathways encompassing all invasive demes. Source populations of invasive demes were
274 found to be admixed (results of step 1, described below). Therefore, plausible hypotheses to
275 test at step 2 were: (1) three independent introductions from three independent admixed
276 source populations; (2) three independent introductions from a single admixed native
277 population; or (3) a single continental introduction from an admixed native population.
278 Demo-genetic scenarios were simulated and summary statistics estimated, with DIYABC
279 version 2.1 (Cornuet et al. 2014). Prior probabilities of scenarios were set to uniform.
280 Parameter prior distributions were first set to a biologically reasonable range of values, then
281 confidence in priors checked (*via* DIYABC test of goodness of fit), with distributions refined
282 iteratively until the simulated scenarios fit the data (final parameter spaces given in
283 supplementary methods). For each scenario, 10,000 simulations were conducted, estimating
284 all available summary statistics in DIYABC for each (Cornuet et al. 2014). The best model
285 was selected using two machine-learning algorithms trained on the simulated datasets: a
286 Neural Network algorithm implemented in the R package 'abc' (Csilléry et al. 2012) and a
287 Random Forest algorithm in the 'abcrf' package (Pudlo et al. 2016). One thousand Neural
288 Networks were trained with 5 to 12 units of hidden neural layers. Parameter sets were
289 weighted by an Epanechnikov kernel. The tolerance rate was set to 0.2. Other configuration
290 values were set to the default value. Power of ABC model selection with Neural Network was
291 evaluated by leave-one-out cross-validation repeated 100 times. Independently, 1,000 trees

292 were grown in the Random Forest training set, with linear discriminant analysis scores
293 added to summary statistics when it reduced the prior error rate. Power of Random Forest
294 was evaluated by out-of-bag prior misclassification error rate. Lastly, the quality of the
295 selected scenario was checked by comparing the marginal posterior predictive distributions
296 to the observed values of the summary statistics. The marginal posterior predictive
297 distribution was computed from 10,000 simulations under the selected scenario,
298 parameterized with estimated posterior parameter distributions as priors.

299 The complexity of the selected scenario at step 2, combined with a restricted number of
300 markers (3,000 SNPs), reduced the power to infer demographic parameters jointly with a
301 reasonable confidence interval. Hence, demographic parameters of the selected scenarios
302 were estimated at step 1 with a regression-based method adjusted by local linear regression
303 (Blum & François 2010; Csilléry et al. 2012). Parameters were weighted by an Epanechnikov
304 kernel. One million simulations were produced under the selected scenario to explore
305 parameter space, but only simulations closest to the observed dataset were retained for
306 parameter estimation (i.e. tolerance rate). Confidence in parameter estimates was checked
307 with leave-one-out cross-validation repeated 1,000 times, thus estimating prediction errors.

308 **2.7 Genotype phasing and imputation**

309 The haplotype phase and missing data of the 468 sequenced individuals were inferred
310 using Beagle 5.1 (Browning & Browning 2007; Browning et al. 2018) to perform population
311 migration modeling and admixture tests. Beagle uses the localized haplotype-cluster model
312 and applies an iterative approach to infer the most likely haplotype pair for each individual.
313 At each iteration, phased input data are used to build a localized haplotype-cluster model.
314 Once the model is built, phased haplotypes for each individual are sampled from the induced
315 diploid HMM conditional on the individual's genotypes. The sampled haplotypes are the input
316 for the next iteration, and so forth. In the final iteration, the Viterbi algorithm selects the most
317 likely haplotypes for each individual, conditional on the diploid HMM and the individual's
318 genotype data. For each copy of each individual, missing alleles are randomly imputed
319 according to allele frequencies, and the data for each individual are phased by randomly

320 ordering the genotypes (Browning & Browning 2007).

321 **2.8 Native population migration modeling**

322 The modeling of the native population splits and mixtures was performed using TreeMix, a
323 statistical model inferring the patterns of population splits and mixtures in multiple
324 populations (Pickrell & Pritchard 2012). A Maximum-Likelihood tree was constructed using
325 genome-wide allele frequency and genetic drift approximation. One thousand bootstraps
326 were performed to assess the robustness of the inferred Maximum-Likelihood (ML) tree.
327 Migration edges were then added sequentially to connect pairs of populations when allele
328 frequency covariance excess was detected. For this analysis, we tested mainland China
329 populations. To strengthen the migration model, we added the genetically close-related Tibet
330 sample to the north-central China deme and considered Japan's population as an outgroup.
331 We ran TreeMix with migration events ranging from zero to four and the TreeMix composite
332 model incorporating known admixture with the '-cor_mig' option (Pickrell et al. 2012) based
333 on STRUCTURE admixture results. The robustness of the tree and the migration edges
334 were confirmed by 1,000 bootstraps using GNU Parallel (Tange 2011) and
335 'treemix.bootstrap' function implemented in the R package BITE (Milanesi et al. 2017).

336 **2.9 Native population admixture test**

337 We selected one deme from each mainland China region (North, North East, Central, South
338 East and South) to perform admixture tests. We used the four-population test of the D-
339 statistic (Green et al. 2010; Patterson et al. 2012) implemented in Popstats (Skoglund et al.
340 2015) to test for admixture and gene flow directionality within the native populations. The
341 notation used by Popstats for the D-statistic is $D(O, P3; P1, P2)$, where O is the outgroup,
342 P3 the test population and P1, P2 the sister populations. A significant negative D indicates
343 that P3 exchanged genes with P1; conversely, a positive D indicates that P3 exchanged
344 genes with P2 (Durand et al. 2011). The D-statistic was estimated for each combination of
345 demes. D-statistic significance was assessed by block jackknife of 5kb and the standard
346 error (SE) was used to estimate the Z-score (Skoglund et al. 2015).

347 **3 RESULTS**

348 **3.1 Complex genetic structure of the native range**

349 We searched for the genetic structure of the native range in order to aggregate sampled
350 sites into fewer homogeneous and consistent demes, hence simplifying further demographic
351 inferences based on ABC modeling. The highest level of structure estimated by
352 STRUCTURE was $K=5-6$ (Figure 1a, Table S3, Supplementary information). To select the
353 most plausible number of genetic clusters K, multiple criteria were assessed and two
354 clustering methods cross-validated (Table S3, Supplementary information). The modal value
355 of the distribution of ΔK suggested $K=3$ in the native area, whereas the shape of the plateau
356 of $\ln(\text{Pr}(X|K))$ suggested a value of K between 5 and 7. At $K=3$, Northern China was well
357 separated from Central China/Japan and Southern China. However, at $K=6-7$, a better
358 resolution on Central China (S4, S6, S10, S11) and Tibet was obtained, with signals of a
359 north-south admixture. The overall patterns of clustering and admixture were similar
360 between $K=6$ and $K=7$ with a spurious cluster at $K=7$. Convergence tests assessed that
361 most sampling chains were convergent, yet some of the sampling chains remained non-
362 convergent even after a burn-in of 100,000 and 100,000 sampling iterations. The highest
363 values of H' was for $K=3$ (0.99), though H' reached 0.81 for $K=6$. Moreover, the DAPC

364 approach yielded very similar results with a most plausible $K=5$. STRUCTURE and DAPC
365 clustering were congruent in revealing the same separation between Northern and Southern
366 China. The K that we finally selected was that meeting statistical, geographical and historical
367 compromises. Hence, $K=6$ was chosen as the better compromise between statistical
368 parsimony and the highest level of genetic structure.

369 **3.2 Definition of native putative demes**

370 The definition of native putative demes based on genetic clustering was crucial to building
371 invasion route scenarios (Supporting information). Most putative demes in the native region
372 were consistent (South China, North China, Japan), despite high uncertainty in some sites
373 (Figure 2a, Figure S6). S3 in particular had a small sample size and was strongly admixed,
374 causing high uncertainty for its assignment to a putative deme. As a consequence, S3 was
375 removed from the dataset for any further analysis. Two admixed demes, composed of sites
376 with a major ancestry coefficient lower than 70%, were created in Central China. These
377 groupings were also justified on the basis of previous studies reporting this region as a zone
378 of secondary contact between Northern and Southern populations (Hardouin et al. 2018;
379 Simon et al. 2011). S11 was assigned at 65% to the North East China deme, but its
380 geographic proximity and connectivity with S10 supported the constitution of an admixed
381 Central China deme encompassing S10 and S11. Lastly, the admixed S13 site was placed
382 into the North China deme because of its location within the same river basin (Figure 2a).
383 While the Tibetan population is genetically representative of the Central China deme
384 (ancestry coefficient $> 75\%$), for historical reasons (i.e. recent Tibetan introduction of *P.*
385 *parva* and few commercial exchanges with Europe), this population was not considered part
386 of the deme. On the other hand, S19 and S20 were genetically well clustered with S9 and
387 S18 (ancestry coefficient $>99\%$) and were considered part of the South China deme. Six
388 native demes were finally defined: North China, North-East China, Central China, South-
389 East China, South China and Japan (Figure 2a).

390 **3.3 Two main gene flow directions shaped the native populations**

391 In the obtained consensus tree (Figure 1b), almost all the nodes were well supported (75%-
392 100%). The Central-North East China node was supported by 47% of the bootstraps. The
393 model with four migration events showed stable migration edges after multiple runs of
394 TreeMix and TreeMix composite models (Figure 1b). The migration edges showed two main
395 southern and northern origins of gene flow: gene flow from the South mainly to the Central
396 and South East populations with 45% and 33% migration weights and gene flow from the
397 North to the North East China population with 9% migration weight (Figure 1b).

398 Admixture tests confirmed gene flow origins followed a directional pattern. The four-
399 population test assumed the population configuration D(O, P3; P1, P2), with Northern and
400 Southern demes as our test populations (P3), the remaining demes as P1 and P2 and Japan
401 as the outgroup O. D-statistics showed an excess of shared derived polymorphism,
402 highlighting genes exchanged between i) the South population and the South East and
403 Central ones and ii) between the North population and the Central and North East ones
404 (Figure 1c, Table S4).

405 **3.4 Genetic structure of the invasive range**

406 Genetic clustering in the invasive range was used to narrow the putative number of non-
407 native populations' origins to test (i.e. the number of scenarios). STRUCTURE results
408 indicated a clear genetic structure in the European invasive range with K=3, supported by
409 DAPC results, despite some uncertainty around the actual highest level of genetic clustering
410 (Table S3). Indeed, for all values of K between 2 and 9, Turkey and Bulgaria clustered
411 together without admixture; Iran formed a similarly distinct cluster (Figures S5a-S5b).
412 Conversely, Western Europe showed a pattern of admixture for all values of K and
413 admixture persisted unless K was equal to the number of sites (Figures S5b-S5c). This
414 indicated that Western Europe is most likely a single deme with a strong sub-structure.
415 Furthermore, hierarchical STRUCTURE analysis performed on the subset of West European
416 sites confirmed a substructure at this scale. Within West European sites (Aus, Bel, Hun, Ita,
417 Pol, Spa, UK), K=3-4 was the most probable grouping according to Evanno's method and

418 K=5 according to the plateau method (Figure S5c). K=5 was displayed for the Western
419 European sub-structure (Figure S5c) because spurious admixture indicated overfitting for
420 higher values (Figure S5b). H' was high (0.98) for K=3-4-5 in the STRUCTURE analysis of
421 the Western European deme. Under Geweke's diagnostic and trace plots, sampling chains
422 seemed convergent or close to convergence, yet some seemed to show departure from
423 stationarity. Heidelberg and Welsh's diagnostic supported that most of the Markov Chains
424 were a stationary distribution. Additionally, Gelman and Rubin's diagnostic assessed that
425 replicated chains converged on similar values. Finally, despite the strong sub-structuring
426 observed, the Western European deme was considered as a single consistent genetic
427 population in ABC scenarios. Hence, we retained three independent demes in the invasive
428 range formed by Western Europe, Eastern Europe and Iran.

429 **3.5 Population assignment to source populations**

430 'AssignPOP' training was efficient, with an assignment accuracy greater than 90% (Figure
431 S7) and all putative demes confidently discriminated by the training algorithm. Predictions
432 with the SVM algorithm confidently assigned 100 individuals (59%) to a source population,
433 with a relative posterior probability >2. Assignment to source populations showed multiple
434 origins in sampled sites, especially in Eastern Europe and Iran (Figure 2b), congruent with
435 DAPC. Posterior membership probabilities of invasive individuals assigned to native Asian
436 clusters with DAPC showed a putative origin in North East China for the Western Europe
437 deme (cluster) and South/South East China with admixture for Bulgaria and Turkey. The
438 Italian site showed a different origin than the rest of Western Europe, closer to Eastern
439 Europe. The Iranian population was separated from all others, linked to South East China or
440 Japan.

441 **3.6 ABC inference of source populations and invasion pathways**

442 ABC simulations were first used to infer the most probable source population in Asia of each
443 one of the three main demes identified in the non-native area (step 1), followed by
444 discrimination between different competing invasion pathways at a global scale (step 2).

445 *Step 1: Source populations of non-native populations*

446 Following prior calibration, all scenarios could be fitted to the data, with 112 summary
447 statistics estimated to compare scenarios. Prior error rates were low for both Neural Network
448 and Random Forest (Table 2), indicating a good predictive power. There was no confusion
449 between scenarios, and the marginal posterior predictive distribution improved the
450 goodness-of-fit of the selected model. Demographic inferences in Western Europe, Eastern
451 Europe and Iran (the three main demes identified from the clustering approaches) all
452 predicted introductions issued from admixture events (scenario 4, Figure S4), supported by
453 high posterior probabilities (Table 2). The choice of candidate source populations was
454 congruent with phylogenies inferred with Maximum Likelihood in Treemix (Figure S8). Each
455 invasive population formed a group with an Asian native population, with relatively short
456 distances between them for European demes. The Iranian population was particular,
457 exhibiting long drift from the most recent common ancestor shared with Japan.

458 *Step 2: Invasion pathways*

459 As in step one, prior fitting to the data was achieved, although model selection was based on
460 256 estimated summary statistics. Prior error rates were also low (Table 2), giving
461 confidence in subsequent inferences. Scenario 1, modelling three independent introductions
462 from independent admixed populations leading to the three observed invasive demes, was
463 selected with strong support from both Neural Network (99%) and Random Forest (86% of
464 votes) algorithms (Table 2, Figure 2c).

465 The three different invasive populations clearly formed three distinct groups with admixed
466 origins. Western Europe origins were in a Northern part of China (North-East China admixed
467 with Central China and North China). Removal of the ambiguous Italian sample did not
468 change the selected scenario. Eastern Europe origins were in a Southern part of China
469 (admixture between Central China, South East China and South China). Lastly, Iran origins
470 were an admixture between China (South East China, North East China) and Japan (Figure
471 2c).

472 **3.7 Invasion process and founding populations**

473 Parameters linked to invasion history (i.e. time of invasion and bottleneck severity of the
474 founding population) were estimated with reasonable confidence intervals by local linear
475 regression (tolerance rate of 0.05 for Eastern and Western Europe; 0.005 for Iran; Figure
476 S9). The Iranian invasion was estimated to have occurred 66 generations ago ($CI_{95\%} = 44;$
477 90), with an effective population size of 217 ($CI_{95\%} = 185; 253$) for the founding group. The
478 Eastern European invasion was estimated to be more recent at 39 generations ago ($CI_{95\%} =$
479 18; 62) and with a smaller effective population size 90 ($CI_{95\%} = 32; 148$). The effective
480 population size of the Western European founding group required two million simulations to
481 be estimated with a higher degree of confidence at 1,436 individuals ($CI_{95\%} = 1,001; 1,825$);
482 furthermore, the time of divergence was estimated at 40 generations ($CI_{95\%} = 20; 63$).

483 **4 DISCUSSION**

484 Population genomics approaches used to inform biological invasions can be challenged by
485 the discrepancy between the large timescale of genetic mutations and the smaller timescale
486 of human-mediated invasions, leading to potentially flawed analyses (Fitzpatrick et al. 2012).
487 Nevertheless, accurate assignments and estimates of geographic origin and number of
488 introductions can be obtained through extensive sampling of native and invasive ranges, the
489 use of high-resolution of molecular markers and a highly genetically structured source
490 population (Excoffier & Heckel 2006; Dlugosch & Parker 2008). Our extensive sampling
491 across both native and invasive areas, the genome-wide markers used in this
492 analysis, as well as previous knowledge on native genetic structuring (Hardouin et al.
493 2018), allowed us to highlight long-term phylogeography and recurrent gene flow. In
494 turn, this knowledge of complex genetic structure in the native range was an
495 essential element allowing us to successfully model probable invasion pathways. Our
496 results suggest that the genomic diversity of invaders was shaped long before introduction
497 by the presence of geographical barriers and by human-mediated gene flow in the native
498 range. The combination of appropriate and complementary methods for populations
499 departing from H-W equilibrium and appropriate clustering within demes allowed us to infer

500 source populations and an invasion scenario despite high admixture within native demes,
501 thus reconciling the genetic history of the native range with recent genomic patterns of
502 invasion.

503 **4.1 The Asian history of *P. parva* shaped by paleogeography and anthropization**

504 ***Paleogeographic influence***

505 Large-scale genetic structure was consistent with the known phylogeography of the species
506 (Hardouin et al. 2018; Simon et al. 2011). Likewise, the retained fine-scale structure in the
507 native range was consistent with non-genetic data (historical, geographical and
508 morphological), suggesting that the estimated value of K is representative of the actual
509 genetic structure. Asian phylogeography revealed a complex genetic structure involving
510 locally high gene flow and variable admixture. Results of STRUCTURE, TreeMix and D-
511 statistics taken together suggest that the current genetic structuring of the native *P. parva*
512 range has been shaped by waves of gene flow originating from southern and northern
513 populations (edges of the natural native distribution). Fish phylogeography in China has
514 been greatly influenced by geological events (Chiang et al. 2013; Li 1981) that can represent
515 major barriers to gene flow (Brandley et al. 2010). The complex geological history of the
516 South China landmass acted as a barrier to gene flows and induced vicariance in common
517 cyprinid species (Yang et al. 2016). In China, notable genetic divergence occurred between
518 the North and the South, with the Qinling Mountains acting as a strong biogeographic barrier
519 between the temperate climate of the North and the subtropical climate of the South (Yuan
520 et al. 2012; Dong et al. 2011) resulting in a wide range of local adaptations to various
521 ecological conditions that might facilitate establishment in the invaded area.

522 Yet within both biogeographical regions, gene flow was restricted and populations were
523 mainly differentiated per river basins. In China north to the Qinling Mountains, the two main
524 demes (i.e. North China and North East China; Figure 2a) were influenced by two different
525 major river networks, the Amur River and the Yellow River (S1, S2 and S3 are located on the
526 Haihe and Huaihe Rivers but both rivers were influenced by the Yellow River in history).
527 South China (sites S9, S18; Figure 2a) was structured by the Yangtze River. Indeed, most

528 native demes were congruent with the architecture of the main river networks. Historically,
529 five *Pseudorasbora* species, based on morphological characteristics and their association
530 with major river basins, were described as endemic to China (Nichols 1928). These five
531 species actually corresponded to the same *P. parva* species (Gozlan 2012), though genetic
532 clustering was largely in agreement with the distribution of historical morphotypes (Nichols
533 1928). Yet admixture was also observed in locations at the boundaries between river basins
534 (e.g. site S16 at the frontier between the Yellow and Amur river basins), indicating that
535 recurrent migration has happened between demes.

536 Wild populations in Japan split from Chinese continental populations 12.1 Mya (Hardouin et
537 al. 2018), consistent with Japan's separation from the Eurasian continent 15 Mya (Barnes
538 2003). Yet our results demonstrated significant gene flow with the South East deme. Two
539 mtDNA lineages co-exist in Japan, including one closely related to continental populations
540 (Watanabe et al. 2000), and recent hybridizations with Chinese populations have been
541 described (Hardouin et al. 2018). Across Honshū Island, *P. parva* rapidly expanded its
542 distribution with translocations of commercial cyprinids into ponds, which may have
543 facilitated the introduction and the spread of the Chinese lineage (Konishi et al. 2003, 2009).
544 However, resolving the origins and contemporary phylogeography of the Japanese
545 population within the native distribution of *P. parva* with sufficient power would require more
546 extensive sampling in Japan.

547 ***Anthropogenic influence***

548 Genetic structure with long-term admixture has been reported in South China for other
549 fishes, explained by coastal land and tributaries between river basins, especially for the
550 Yangtze and Pearl Rivers (Yang et al. 2016). Admixed populations may be the result of
551 anthropogenic modifications to the hydrological landscape, especially new dispersal
552 pathways such as canals. Recent secondary contacts between the North and South may
553 have been facilitated by the increased structural connectivity between river basins. The
554 construction of the Lingqu Canal 2,200 years ago connected the Yangtze River (sites S9,
555 S18 in South China and to a lesser extent S6 and S4 in South East China; Figure 2a) to the

556 Pearl River (Fengshu 1990). Moreover, the Beijing-Hangzhou Grand Canal connected the
557 Yangtze River to the Huai He, Hai He and Yellow rivers in the east of China (AD 581-618,
558 Sui Dynasty), a region at the core of the two admixed regions (Central China, North East
559 China and South East China: Figure 2a). Secondary contact due to human-induced corridors
560 can increase local admixture between populations (Crispo et al. 2011), and canals serving
561 as ecological corridors between two interbreeding cyprinid species have ultimately become
562 hybrid zones (Guivier et al. 2019). However, the most important vector for long-distance
563 dispersal in the native range is aquaculture expansion throughout central China. The two
564 most admixed demes correspond to areas of intensive aquaculture in China. In Central
565 China, many ponds and reservoirs were intensively stocked in the 1950s-60s, with *P. parva*
566 eventually becoming the dominant species (Zhao et al. 2015; Gong & Tu 1991). Massive
567 human translocations for aquaculture have often led to biotic homogenization (Olden et al.
568 2004) and increased pathogen dispersal into the pool of future invaders (Price et al. 2016).
569 Moreover, consecutive introductions represent a type of punctuated gene flow that produces
570 admixed genotypes with high genetic diversity but unpredictable evolutionary effects (Crispo
571 et al. 2011; Hasselman et al. 2014). In native locations, admixture is known for negative
572 effects such as outbreeding depression and loss of local adaptation (Côte et al. 2014; Huff et
573 al. 2011; Hufford et al. 2012), but in novel environments, admixture may increase adaptive
574 potential for translocated individuals (Verhoeven et al. 2011). For example, admixture in the
575 wild between divergent sculpin populations increased their genetic diversity, lineage
576 differentiation, and facilitated colonization of new habitats (Nolte et al. 2005, 2009;
577 Stemshorn et al. 2011).

578 **4.2 The invasion**

579 ***Invasion starts in Asia***

580 Invasive populations have been successfully established at the edges of the native
581 distribution. In South-West China, sites S19 and S20 are recently introduced populations
582 (1980s) from South-East China, and perfectly clustered with S9, S18 (Figure 2a; see also
583 Hardouin et al. 2018). The Tibetan population was also introduced, commercial exchange

584 being most probably the pathway for this invasion given that many living fishes are sold in
585 Tibetan markets and are traditionally released into rivers (Gozlan 2012). Although Tibet
586 shows similarities with Central China, a continuum of populations between them were not
587 sampled, making it difficult to determine if Tibet should be clustered with this deme or
588 another.

589 ***Out of Asia***

590 Three invasive demes were consistently discriminated as (1) Iran, (2) a small Balkan-
591 Anatolian deme (Eastern Europe) and (3) a large pan-European deme (Western Europe).
592 Population differentiation (F_{ST}) was overall higher between the three invasive demes, than
593 between the invasive demes and their source populations, suggesting independent
594 introductions with different genomic backgrounds (Figure 3, Table S5). In addition, clear
595 genetic sub-structuring was observed in the Western European deme. More populations
596 sampled in this region and more markers would then be essential to more accurately depict
597 fine-scale genetic structure of the West European deme. The Iranian population was
598 particular, forming a distinct deme completely isolated from all European populations,
599 consistent with observations of Hardouin et al. (2018), who revealed that Iran has a different
600 mtDNA lineage. Differentiation between Eastern and Western Europe may be imputed to the
601 former USSR era when on one side, the eastern introduction first opens and on the other
602 commercial exchanges with Western Europe were limited (Britton & Gozlan 2013). The
603 Danube river ensures structural connectivity which has favored colonization among a large
604 central European region (Weiss et al. 2002), from Germany to Bulgaria through Austria and
605 Hungary. However, regional biogeography of the Balkan region could also partially explain
606 the observed population structure (Oikonomou et al. 2014). In addition, rapid range
607 expansion of *P. parva* over long distances in the invasive range likely relied on human-
608 mediated dispersal (e.g. unintentional releases, aquaculture exchanges) rather than natural
609 dispersal, as *P. parva* favours lentic habitats. The observed genetic structure of invasive
610 locations may suggest multiple independent introductions without subsequent gene flow.
611 This assumption is strengthened by an invasive population genetic diversity similar to the

612 native populations (Table S1). A recent study showed a clear reduction in genetic diversity in
613 the French invasive deme of *P. parva* compared to native demes, highlighting that the
614 introduction was accompanied by a substantial loss of genetic diversity (Combe et al. 2022).
615 However, in a previous study, high genetic diversity among invasive *P. parva* has been
616 reported, first by Simon et al. (2015), and later attributed by Hardouin et al. (2018) to
617 unbalanced sampling. Baltazar-Soares et al. (2020) have since highlighted a high genomic
618 diversity attributed to a lack of admixture due to a recovery of the genetic bottleneck
619 associated with the introduction. Here, with more balanced sampling between invasive and
620 native populations, coupled with observed admixture in the invasive population (Figure 2b,
621 Figure S5), our results suggest an invasive genetic diversity shaped by past invasions of
622 admixed native populations.

623 It is common to observe strong genetic structure within demes for continental invasive fishes
624 when using nuclear markers. This is partly explained by human driven long-distance
625 dispersal and geographical barriers to gene flow (Sanz et al. 2013; Simon et al. 2011).
626 Although we detected signals of differentiation with nuclear SNP data, mtDNA has revealed
627 that Western European locations were characterized by the same mitochondrial lineage and,
628 thus, constitute a single genetic population sharing a common ancestor (Simon et al. 2011).
629 Low migration rates in a stepping-stone model induced by human-mediated secondary
630 introductions might explain the relative differentiation between Western European locations,
631 and might be a sign of an invasive bridgehead (van Boheemen et al. 2017). Invasive
632 populations could also act as source populations for a secondary invasion, even over long-
633 distances (Karsten et al. 2015; Lombaert et al. 2014).

634 **4.3 Origins of invasion: admixture and multiple introductions**

635 Our models described major demographic events, and gave congruent results
636 between the two independent Machine Learning algorithms, supporting three
637 independent introductions at the origins of the three invasive demes. Multiple
638 introductions would, thus, multiply the chances of successful establishment outside

639 the native range. Results point towards a southern Chinese origin of the Eastern European
640 deme, with alleles coming from and/or shared between the Central, South East and South
641 China demes (Figure 2c). Similar patterns have been demonstrated via population
642 assignment tests, revealing multiple origins in the same location indicative of multiple
643 introductions or admixture within invasive demes (Guillemaud et al. 2011). Studies have also
644 highlighted that admixture is likely a common characteristic for invasive species (Genton et
645 al. 2005; Kolbe et al. 2004; Rius & Darling 2014); however, it is often difficult to disentangle
646 multiple introductions from admixture before introduction. The reconstruction of admixture
647 history may depend on the genetic structure observed in the native area. Admixture after
648 introduction may be inferred when source populations are clearly differentiated and
649 admixture is demonstrated in the invasive area, as in the case of the mussel, *Mytella*
650 *charruana* (Gillis et al. 2009). However, this is not the case for *P. parva*, which shows high
651 rates of admixture within the inferred source populations, suggesting that human-mediated
652 gene flow within the native range shaped genetic variation in the pool of future invaders. The
653 Central deme corresponded to the Wuhan region, at the core of Chinese aquaculture, with
654 more than two million fishes produced annually, and frequent translocations from
655 surrounding regions to restock ponds (Zhao et al. 2015). Consequently, admixture in the
656 native range before the translocation into Europe may be at the origin of the Eastern
657 European deme. The original introduction probably took place in Romania, at the epicenter
658 of carp aquaculture in the 1960s, in the Nucet Fisheries Research Centre in 1961 (Gozlan et
659 al. 2010), although several other introductions took place at the same time throughout the
660 Black Sea region. Introduction was most likely accidental, as *P. parva* is a common
661 contaminant of carp stockings (Wolter & Röhr 2010). Assuming between one and two
662 generations per year among eastern European populations (Gozlan 2008), it complies with
663 our estimates of the time of invasion (ca. 18 to 62 generations), and the time of Asian fish
664 export increases. Likewise, development of aquaculture in this part of Europe may have
665 subsequently facilitated the spread of *P. parva* in Turkey as early as 1982 (Ekmekçi &
666 Kırankaya 2006; Britton & Gozlan 2013).

667 The Western Europe deme was probably colonized from a first introduction in Hungary,
668 noticed in the Paks Fisheries Farm in 1963 (Gozlan et al. 2010), which then spread
669 throughout Western Europe. The source populations mainly differed from those of the
670 Eastern Europe deme, although both share a connection with Central China. Western
671 Europe populations also include an admixed north Chinese origin to both the North East
672 China and North China demes. The invasion time was approximately the same as for
673 Eastern Europe, coincident with a trading cooperation period between countries of the
674 Eastern Bloc and China (Britton & Gozlan 2013). On average, genetic divergence amongst
675 source demes was relatively high (mean $F_{ST} = 0.27$; Figure 3, Table S5), suggesting limited
676 gene flow and a lower likelihood of native admixture, compared to that for Eastern Europe
677 source demes (mean $F_{ST} = 0.13$; Figure 3, Table S5). Moreover, individuals of the Hungarian
678 location were systematically assigned to two genetic clusters (Figure S5c), making it the
679 most admixed location of Western Europe, and suggested further that admixture most likely
680 happened from multiple introductions in the same invasive area. Temporal dynamics of
681 invasive mosquitoes in the US have shown the same admixture-like clustering pattern
682 (Fonseca et al. 2010). Range expansion across European countries most probably began in
683 the Hungarian admixed region, and its likely chronology has been reconstructed from census
684 data (Gozlan et al. 2010). *P. parva* colonized Austria as early as 1982 (Weber 1984), then
685 Germany from 1984 to 1987 (Arnold 1985; Lelek & Köhler 1989), and Belgium by 1992
686 (Vandelannoote & Yseboodt 1998). Austrian, Belgian and Polish sites clustered together,
687 suggesting a common ancestry. Their connectivity along the Danube and Rhine river
688 systems may have facilitated natural dispersal (Hegediš et al. 2007; Leuven et al. 2009).
689 However, natural dispersal cannot explain the observed high genetic structure, nor the
690 expansion over geographical barriers (e.g. the English Channel, the Alps), and so human-
691 mediated dispersal is more likely (Gozlan et al. 2002, Aparicio et al. 2012, Caiola & Sostoa
692 2002). Thus, Hungary could be the invasive bridgehead that initiated the colonization of
693 Western Europe, followed by secondary bottlenecks associated with a series of founding
694 events during westward range expansion. This invasive bridgehead effect scenario has also

695 been suggested by Combe et al. (2022) for a French deme, which probably originated from a
696 successful invasive population. Surprisingly, Italy showed different origins than other
697 Western European locations in population assignment tests (Central China instead of North-
698 East China; Figure 2b), but was nonetheless considered in the Western European deme
699 based on genetic clustering results. The origin of this small population was difficult to
700 explain, as it may be part of an unsampled fourth invasive population linked with Slovakia
701 that successfully spread across the Italian peninsula (Carosi et al. 2016; Záhorská & Kováč
702 2009). This uncertainty in the Italian origin underlies the necessity for the same extensive
703 sampling in Europe as in Asia, especially in the Southeastern Europe that was believed to
704 be the core of past introductions (e.g. Slovenia, Hungary, Romania, Bulgaria). This punctual
705 sampling represents an inherent limitation to the definition of putative demes, and
706 subsequently, to the number of invasion pathways inferred.

707 The Iranian population origin was complex to infer because of an ancient admixed origin
708 between South East China and Japan. Iran showed a pattern of assignment similar to that of
709 the Japanese admixture in nuclear markers between endemic and Chinese populations,
710 potentially explaining why it was not possible to disentangle China and Japan as putative
711 sources. In line with our results, a recent study (Ganjali et al. 2020), based on two mtDNA
712 markers, identified three Iranian matrilineal haplotypes belonging to two distinct lineages, an
713 older Japanese lineage and a Chinese lineage, corresponding to a recent natural dispersal
714 from Azerbaijan. They also highlighted an admixture of highly divergent Japanese and
715 Chinese lineages (Hardouin et al. 2018). Many questions remain, mainly because the history
716 of Japanese populations has not been fully resolved. The high number of private alleles in
717 Iran (Table S1) and the long drift separating the Iranian population from the most recent
718 common ancestor with Japan (Figure S8) might suggest that the true source population has
719 not been sampled. Thus, additional sampling would be required to effectively determine the
720 origins of Iranian *P. parva*.

721 **4.4 A successful invasion**

722 The surprisingly large effective population sizes in Western Europe likely enhanced the

723 probability of introduced populations survival, and may explain its invasiveness.
724 Furthermore, the co-introduction of non-native host-pathogen systems such as *P. parva*-*S.*
725 *destruens* may create new host-pathogen interactions that can be detrimental to naïve native
726 host species, hence providing a competitive advantage for the invaders (Price et al. 1986;
727 Andreou & Gozlan 2016; Combe & Gozlan 2018; Vilcinskis & Knoll 2015). The gene flow
728 resulting from the multiple introductions highlighted in Western Europe largely increased the
729 effective population size as well as genetic diversity. However, effective population sizes
730 were smaller in Iran and Eastern Europe, although the bottleneck they experienced was not
731 severe, with 100 to 200 individuals in the founder populations. *P. parva* probably overcame
732 the loss of diversity induced by bottlenecks due to admixture between native populations
733 within its new environment and prior to range expansion. Our results demonstrated multiple
734 source populations coming from a wide range of climatic conditions, the most flagrant being
735 the admixture in the Wuhan region between subtropical and temperate populations, which
736 represent different morphotypes (Nichols 1928; Gozlan et al. 2020). Multiple introductions
737 associated with admixture can either increase or decrease the adaptive potential of invasive
738 populations (Barker et al. 2019; Verhoeven et al. 2011). However, if from genetically distinct
739 sources, multiple introductions can mitigate the negative effects of bottlenecks associated
740 with invasion, as genetically diverse populations are less affected by the deleterious effects
741 of inbreeding depression (Verhoeven et al. 2011). Moreover, they can increase individual
742 fitness through heterosis, contributing to the invasion's success (Rius & Darling 2014;
743 Vallejo-Marín et al. 2021), which is most likely the case in *P. parva*. Hence, native
744 populations with broad and diverse biogeographic distribution may act as genetic and
745 ecological diversity reservoirs.

746 **5 CONCLUSION**

747 Our results shed light on the importance of grasping the genetic history of the native range
748 populations to better understand the effects of introductions and admixture on the success
749 and adaptive potential of invasive populations. Our study also draws a picture of the complex
750 demo-genetic history of an invasive species' source populations and its spread across

751 recurrent global invasion pathways. Reconstruction of such invasion pathways is crucial for
752 setting up conservation biology approaches and management to prevent further non-native
753 species introductions and potential associated pathogens.

754 **Acknowledgements**

755 We would like to thank all co-authors of Gozlan et al. 2020 for their help in collecting *P.*
756 *parva* samples across the native and invasive range as well as B. Haenfling. We would also
757 like to thank Eric Lombaert for kindly agreeing to discuss the best practices of Approximate
758 Bayesian Computation methods. This study has been conducted in accordance with the
759 relevant animal or human ethics approvals.

760 **References**

- 761 Andreou D, Gozlan, RE (2016) Associated disease risk from the introduced generalist
762 pathogen *Sphaerothecum destruens*: management and policy implications. *Parasitology*
763 143:1204-1210. <https://doi.org/10.1017/S003118201600072X>
- 764 Aparicio E, Peris B, Torrijos L, Prenda J, Nieva A, Perea S (2012) Expansion of the invasive
765 *Pseudorasbora parva* (Cyprinidae) in the Iberian Peninsula: first record in the Guadiana
766 River basin. *Cybiurn* 36:585-586
- 767 Arnold A (1985) *Pseudorasbora parva* (Schlegel 1842) nun auch in der DDR. *Z. Binnenfisch*
768 *DDR* 32:182-183
- 769 Baltazar Soares M, Blanchet S, Cote J, Tarkan AS, Záhorská E, Gozlan RE, Eizaguirre C
770 (2020) Genomic footprints of a biological invasion: Introduction from Asia and dispersal in
771 Europe of the topmouth gudgeon (*Pseudorasbora parva*). *Mol Ecol* 29:71-85
- 772 Barker BS, Cocio JE, Anderson SR, Braasch JE, Cang FA, Gillette HD et al (2019) Potential
773 limits to the benefits of admixture during biological invasion. *Mol Ecol* 28:100-113.
774 <https://doi.org/10.1111/mec.14958>
- 775 Barnes GL (2003) Origins of the Japanese Islands: The New" Big Picture". *Nichibunken*
776 *Japan Review* 3-50
- 777 Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in
778 population genetics. *Genetics* 162:2025-2035
- 779 Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A (2018) maps: Draw Geographical
780 Maps. R package version 3.3. 0. <https://CRAN.R-project.org/package=maps>
- 781 Beichman AC, Huerta-Sanchez E, Lohmueller KE (2018) Using genomic data to infer historic
782 population dynamics of nonmodel organisms. *Annu Rev Ecol Evol Syst* 49:433-456.
783 <https://doi.org/10.1146/annurev-ecolsys-110617-062431>
- 784 Blum MG, François O (2010) Non-linear regression models for Approximate Bayesian
785 Computation. *Stat Comput* 20:63-73. <https://doi.org/10.1007/s11222-009-9116-0>
- 786 Bossdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati D (2005) Phenotypic and
787 genetic differentiation between native and introduced plant populations. *Oecologia* 144:1-11.
788 <https://doi.org/10.1007/s00442-005-0070-z>
- 789 Brandley MC, Guiher TJ, Pyron RA, Winne CT, Burbrink FT (2010) Does dispersal across an
790 aquatic geographic barrier obscure phylogeographic structure in the diamond-backed
791 watersnake (*Nerodia rhombifer*)? *Mol Phylogenetics Evol* 57:552-560.

792 <https://doi.org/10.1016/j.ympcv.2010.07.015>

793 Britton JR, Gozlan RE (2013) Geo-politics and freshwater fish introductions: How the Cold
794 War shaped Europe's fish allodiversity. *Glob Environ Change* 23:1566-1574.
795 <https://doi.org/10.1016/j.gloenvcha.2013.09.017>

796 Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-
797 generation reference panels. *Am J Hum Genet* 103:338-348.
798 <https://doi.org/10.1016/j.ajhg.2018.07.015>

799 Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data
800 inference for whole-genome association studies by use of localized haplotype clustering. *Am*
801 *J Hum Genet* 81:1084-1097. <https://doi.org/10.1086/521987>

802 Cabrera AA, Palsbøll PJ (2017) Inferring past demographic changes from contemporary
803 genetic data: A simulation based evaluation of the ABC methods implemented in DIYABC.
804 *Mol Ecol Resour* 17:e94-e110. <https://doi.org/10.1111/1755-0998.12696>

805 Caiola N, De Sostoa A (2002) First record of the Asiatic cyprinid *Pseudorasbora parva* in the
806 Iberian Peninsula. *J Fish Biol* 4:1058-1060. <https://doi.org/10.1006/jfbi.2002.2103>

807 Carosi A, Ghetti L, Lorenzoni M (2016) Status of *Pseudorasbora parva* in the Tiber river
808 basin (Umbria, central Italy) 20 years after its introduction. *Knowl Manag Aquat Ecosyst*
809 417:22. <https://doi.org/10.1051/kmae/2016009>

810 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis
811 tool set for population genomics. *Mol Ecol* 22:3124-3140. <https://doi.org/10.1111/mec.12354>

812 Chan KO, Alexander AM, Grismer LL, Su Y-C, Grismer JL, Quah ESH, Brown RM (2017)
813 Species delimitation with gene flow: A methodological comparison and population genomics
814 approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Mol Ecol*
815 26:5435-5450

816 Chapin Iii FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL et al (2000)
817 Consequences of changing biodiversity. *Nature* 405:234-242.
818 <https://doi.org/10.1038/35012241>

819 Chen KY, Marschall EA, Sovic MG, Fries AC, Gibbs HL, Ludsins SA (2018) Assign POP: An r
820 package for population assignment using genetic, non-genetic, or integrated data in a
821 machine-learning framework. *Methods Ecol Evol* 9:439-446. <https://doi.org/10.1111/2041-210X.12897>

823 Chiang TY, Lin HD, Zhao J, Kuo PH, Lee TW, Hsu KC (2013) Diverse processes shape
824 deep phylogeographical divergence in *Cobitis sinensis* (*Teleostei: Cobitidae*) in East Asia. *J*
825 *Zool Syst Evol Res* 51:316-326. <https://doi.org/10.1111/jzs.12030>

826 Clavero M, García-Berthou E (2005) Invasive species are a leading cause of animal
827 extinctions. *Trends Ecol Evol* 20:110. <https://doi.org/10.1016/j.tree.2005.01.003>

828 Combe M, Gozlan RE (2018) The rise of the rosette agent in Europe: An epidemiological
829 enigma. *Transbound Emerg Dis* 65:1474-1481. <https://doi.org/10.1111/tbed.13001>

830 Combe M, Cherif E, Charrier A, Barbey B, Chague M, Carrel G et al (2022) Towards
831 unravelling the Rosette agent enigma: Spread and emergence of the co-invasive host-
832 pathogen complex, *Pseudorasbora parva-Sphaerothecum destruens*. *Sci. Total Environ.*
833 806:150427. <https://doi.org/10.1016/j.scitotenv.2021.150427>

834 Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R et al (2014)
835 DIYABC v2.0: a software to make approximate Bayesian computation inferences about
836 population history using single nucleotide polymorphism, DNA sequence and microsatellite
837 data. *Bioinformatics* 30:1187-1189. <https://doi.org/10.1093/bioinformatics/btt763>

838 Côte J, Roussel JM, Le Cam S, Evanno G (2014) Outbreeding depression in Atlantic
839 salmon revealed by hypoxic stress during embryonic development. *Evol Biol* 41:561-571.
840 <https://doi.org/10.1007/s11692-014-9289-0>

841 Crispo E, Moore JS, Lee Yaw JA, Gray SM, Haller BC (2011) Broken barriers: Human
842 induced changes to gene flow and introgression in animals: An examination of the ways in
843 which humans increase genetic exchange among populations and species and the
844 consequences for biodiversity. *BioEssays* 33:508-518.
845 <https://doi.org/10.1002/bies.201000154>

846 Cristescu ME (2015) Genetic reconstructions of invasion history. *Mol Ecol* 24:2212-2225.
847 <https://doi.org/10.1111/mec.13117>

848 Crowl TA, Crist TO, Parmenter RR, Belovsky G, Lugo AE (2008) The spread of invasive
849 species and infectious disease as drivers of ecosystem change. *Front Ecol Environ* 6:238-
850 246. <https://doi.org/10.1890/070151>

851 Csilléry K, François O, Blum MG (2012) abc: an R package for approximate Bayesian
852 computation (ABC). *Methods Ecol Evol* 3:475-479. [https://doi.org/10.1111/j.2041-
853 210X.2011.00179.x](https://doi.org/10.1111/j.2041-210X.2011.00179.x)

854 Didham RK, Tylianakis JM, Hutchison MA, Ewers RM, Gemmell NJ (2005) Are invasive
855 species the drivers of ecological change? *Trends Ecol Evol* 20:470-474.
856 <https://doi.org/10.1016/j.tree.2005.07.006>

857 Dlugosch KM, Parker IM (2008) Founding events in species invasions: Genetic variation,
858 adaptive evolution, and the role of multiple introductions. *Mol Ecol* 17:431-449

859 Dong Y, Zhang G, Neubauer F, Liu X, Genser J, Hauzenberger C (2011) Tectonic evolution
860 of the Qinling orogen, China: review and synthesis. *J Asian Earth Sci* 41:213-237.
861 <https://doi.org/10.1016/j.jseaes.2011.03.002>

862 Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between
863 closely related populations. *Mol Biol Evol* 28:2239-2252.
864 <https://doi.org/10.1093/molbev/msr048>

865 Earl DA, vonHoldt B.M (2012) STRUCTURE HARVESTER: a website and program for
866 visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet*
867 *Resour* 4:359-361. <https://doi.org/10.1007/s12686-011-9548-7>

868 Ekmekçi FG, Kirankaya ŞG (2006) Distribution of an invasive fish species, *Pseudorasbora*
869 *parva* (Temminck & Schlegel, 1846) in Turkey. *Turk J Zool* 30:329-334

870 Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data : Why,
871 how and so what? *Mol Ecol* 19:4113-4130. [https://doi.org/10.1111/j.1365-
872 294X.2010.04773.x](https://doi.org/10.1111/j.1365-294X.2010.04773.x)

873 Estoup A, Lombaert E, Marin JM, Guillemaud T, Pudlo P, Robert CP, Cornuet JM (2012)
874 Estimation of demographic genetic model probabilities with Approximate Bayesian Computation
875 using linear discriminant analysis on summary statistics. *Mol Ecol Resour* 12:846-855.
876 <https://doi.org/10.1111/j.1755-0998.2012.03153.x>

877 Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B (2016) Is there a genetic
878 paradox of biological invasion? *Annu Rev Ecol Evol Syst* 47:51-72.
879 <https://doi.org/10.1146/annurev-ecolsys-121415-032116>

880 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals
881 using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611-2620. [https://doi.org/
882 10.1111/j.1365-294X.2005.02553.x](https://doi.org/10.1111/j.1365-294X.2005.02553.x)

883 Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a
884 survival guide. *Nat Rev Genet* 7:745-758

885 Facon B, Genton BJ, Shykoff J, Jarne P, Estoup A, David P (2006) A general eco-
886 evolutionary framework for understanding bioinvasions. *Trends Ecol Evol* 21:130-135.
887 <https://doi.org/10.1016/j.tree.2005.10.012>

888 Fengshu Z (1990) Preliminary study of geological geomorphological conditions of ling canal.
889 *Carsologica Sinica* 1

890 Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about
891 biological invasions?. *Biol. Invasions* 14:245-253.

892 Fonseca DM, Widdel AK, Hutchinson M, Spichiger SE, Kramer LD (2010) Fine-scale
893 spatial and temporal population genetics of *Aedes japonicus*, a new US mosquito, reveal
894 multiple introductions. *Mol Ecol* 19:1559-1572. <https://doi.org/10.1111/j.1365-294X.2010.04576.x>

896 Gallardo B, Clavero M, Sánchez MI, Vilà M (2016) Global ecological impacts of invasive
897 species in aquatic ecosystems. *Glob Chang Biol* 22:151-163.
898 <https://doi.org/10.1111/gcb.13004>

899 Ganjali Z, Esmaeili HR, Zarei F, Sayyadzadeh G, Eagderi S, Gozlan RE (2020) West Asian
900 colonisation of topmouth gudgeon, *Pseudorasbora parva* (*Teleostei: Gobionidae*): Genetic
901 admixture at the crossroad of Europe and east Asia. *Freshw Biol* 66:699–715. <https://doi.org/10.1111/fwb.13671>

903 Genton BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive
904 populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of
905 introduction. *Mol Ecol* 14:4275-4285. <https://doi.org/10.1111/j.1365-294X.2005.02750.x>

906 Gillis NK, Walters LJ, Fernandes FC, Hoffman EA (2009) Higher genetic diversity in
907 introduced than in native populations of the mussel *Mytella charruana*: evidence of
908 population admixture at introduction sites. *Divers Distrib* 15:784-795.
909 <https://doi.org/10.1111/j.1472-4642.2009.00591.x>

910 Gong M, Tu F (1991) Fishery in contemporary China. Contemporary China Press, Beijing

911 Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol*
912 *Ecol Notes* 5:184-186. <https://doi.org/10.1111/j.1471-8278.2004.00828.x>

913 Goudet J, Jombart T (2015) hierfstat: estimation and tests of hierarchical F-statistics. R
914 package version 0.04-22. 10

915 Gozlan RE, Andreou D, Asaeda T, Beyer K, Bouhadad R, Burnard D et al (2010) Pan-
916 continental invasion of *Pseudorasbora parva*: towards a better understanding of freshwater
917 fish invasions. *Fish Fish* 11:315-340. <https://doi.org/10.1111/j.1467-2979.2010.00361.x>

918 Gozlan RE, Pinder AC, Shelley J (2002) Occurrence of the Asiatic cyprinid *Pseudorasbora*
919 *parva* in England. *J Fish Biol* 61:298-300. <https://doi.org/10.1006/jfbi.2002.2042>

920 Gozlan RE, Záhorská E, Cherif E, Asaeda T, Britton JR, Chang CH et al (2020) Native
921 drivers of fish life history traits are lost during the invasion process. *Ecol Evol* 10:8623-8633.
922 <https://doi.org/10.1002/ece3.6521>

923 Gozlan RE (2008) Introduction of non-native freshwater fish: is it all bad? *Fish Fish* 9:106-
924 115. <https://doi.org/10.1111/j.1467-2979.2007.00267.x>

925 Gozlan RE (2012) *Pseudorasbora parva* temminck and schlegel (topmouth gudgeon).
926 *Handb Glob Freshw Invasive Species*, Abingdon: Earthscan

927 Graebner RC, Callaway RM, Montesinos D (2012) Invasive species grows faster, competes
928 better, and shows greater evolution toward increased seed size and growth than exotic non-
929 invasive congeners. *Plant Ecol* 213:545-553. <https://doi.org/10.1007/s11258-012-0020-x>

930 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M et al (2010) A draft
931 sequence of the Neandertal genome. *Science* 328:710-722.
932 <https://doi.org/10.1126/science.1188021>

933 Guillemaud T, Ciosi M, Lombaert E, Estoup A (2011) Biological invasions in agricultural
934 settings: insights from evolutionary biology and population genetics. *C R Biol* 334:237-246.
935 <https://doi.org/10.1016/j.crv.2010.12.008>

936 Guivier E, Gilles A, Pech N, Dufлот N, Tissot L, Chappaz R (2019) Canals as ecological
937 corridors and hybridization zones for two cyprinid species. *Hydrobiologia* 830(1):1-16. <https://doi.org/10.1007/s10750-018-3843-1>

939 Hardouin EA, Andreou D, Zhao Y, Chevret P, Fletcher DH, Britton JR, Gozlan RE (2018)
940 Reconciling the biogeography of an invader through recent and historic genetic patterns: the
941 case of topmouth gudgeon *Pseudorasbora parva*. *Biol Invasions* 20(8):2157-2171.
942 <https://doi.org/10.1007/s10530-018-1693-4>

943 Hasselman DJ, Argo EE, McBride MC, Bentzen P, Schultz TF, Perez-Umphrey AA,
944 Palkovacs EP (2014) Human disturbance causes the formation of a hybrid swarm between
945 two naturally sympatric fish species. *Mol Ecol* 23:1137-1152.
946 <https://doi.org/10.1111/mec.12674>

947 Price PW, Westoby M, Rice B, Atsatt PR, Fritz RS, Thompson JN, Mobley K (1986) Parasite
948 mediation in ecological interactions. *Annu. Rev. Ecol. Syst.* 17:487–505.

949 Hegediš A, Lenhardt M, Mičković B, Cvijanović G, Jarić I, Gačić Z (2007) Amur sleeper
950 (*Percocottus glenii* Dubowski, 1877) spreading in the Danube River basin. *J Appl Ichthyol*
951 23:705-706. <https://doi.org/10.1111/j.1439-0426.2007.00867.x>

952 Holsbeek G, Mergeay J, Hotz H, Plötner J, Volckaert FAM, De Meester L (2008) A cryptic
953 invasion within an invasion and widespread introgression in the European water frog
954 complex: consequences of uncontrolled commercial trade and weak international legislation.
955 *Mol Ecol* 17:5023-5035. <https://doi.org/110.1111/j.1365-294X.2008.03984.x>

956 Huang H, Knowles LL (2016) Unforeseen Consequences of Excluding Missing Data from
957 Next-Generation Sequences: Simulation Study of RAD Sequences. *Syst Biol* 65:357–365.
958 <https://doi.org/10.1093/sysbio/syu046>

959 Huff DD, Miller LM, Chizinski CJ, Vondracek B (2011) Mixed-source reintroductions lead to
960 outbreeding depression in second-generation descendents of a native North American fish.
961 *Mol Ecol* 20:4246-4258. <https://doi.org/10.1111/j.1365-294X.2011.05271.x>

962 Hufford KM, Krauss SL, Veneklaas EJ (2012) Inbreeding and outbreeding depression in
963 *Styloidium hispidum*: implications for mixing seed sources for ecological restoration. *Ecol Evol*
964 2:2262-2273. <https://doi.org/10.1002/ece3.302>

965 Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an
966 era of globalization. *J Appl Ecol* 46:10-18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>

967 Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation
968 program for dealing with label switching and multimodality in analysis of population structure.
969 *Bioinformatics* 23:1801-1806. <https://doi.org/10.1093/bioinformatics/btm233>

970 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP
971 data. *Bioinformatics* 27:3070-3071. <https://doi.org/10.1093/bioinformatics/btr521>

972 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.
973 *Bioinformatics* 24:1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>

974 Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide
975 population genetic data with emphasis on clonality. *Front Genet* 6:208.
976 <https://doi.org/10.3389/fgene.2015.00208>

977 Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of
978 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.
979 <https://doi.org/10.7717/peerj.281>

980 Karsten M, Jansen van Vuuren B, Addison P, Terblanche JS (2015) Deconstructing
981 intercontinental invasion pathway hypotheses of the Mediterranean fruit fly (*Ceratitidis*
982 *capitata*) using a Bayesian inference approach: are port interceptions and quarantine
983 protocols successfully preventing new invasions? *Divers Distrib* 21:813-825. <https://doi.org/10.1111/ddi.12333>

984

985 Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB (2004) Genetic variation
986 increases during biological invasion by a Cuban lizard. *Nature* 431:177-181.
987 <https://doi.org/10.1038/nature02807>

988 Konishi M, Hosoya K, Takata K (2003) Natural hybridization between endangered and

989 introduced species of *Pseudorasbora*, with their genetic relationships and characteristics
990 inferred from allozyme analyses. *J Fish Biol* 63:213-231. <https://doi.org/10.1046/j.1095-8649.2003.00146.x>
991

992 Konishi M, Sakano H, Iguchi KI (2009) Identifying conservation priority ponds of an
993 endangered minnow, *Pseudorasbora pumila*, in the area invaded by *Pseudorasbora parva*.
994 *Ichthyol Res* 56:346. <https://doi.org/10.1007/s10228-009-0106-1>

995 Lelek A, Köhler C (1989) Zustandsanalyse der Fischartengemeinschaften im Rhein (1987–
996 1988) *Fischökologie* 1:47-64

997 Leuven RS, van der Velde G, Baijens I, Snijders J, van der Zwart C, Lenders HR, bij de
998 Vaate A (2009) The river Rhine: a global highway for dispersal of aquatic invasive species.
999 *Biol Invasions* 11:1989. <https://doi.org/10.1007/s10530-009-9491-7>

1000 Li S (1981) Studies on zoogeographical divisions for fresh water fishes of China. Science
1001 Press

1002 Li YL, Liu JX (2018) StructureSelector: A web-based software to select and visualize the
1003 optimal number of clusters using multiple methods. *Mol Ecol Resour* 18:176-177.
1004 <https://doi.org/10.1111/1755-0998.12719>

1005 Lombaert E, Guillemaud T, Lundgren J, Koch R, Facon B, Grez A et al (2014)
1006 Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the
1007 case of the Asian ladybird *Harmonia axyridis*. *Mol Ecol* 23:5979-5997.
1008 <https://doi.org/10.1111/mec.12989>

1009 Mergeay J, Verschuren D, De Meester L (2005) Cryptic invasion and dispersal of an
1010 American *Daphnia* in East Africa. *Limnol Oceanogr* 50:1278-1283.
1011 <https://doi.org/10.4319/lo.2005.50.4.1278>

1012 Milanesi M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, Colli L
1013 (2017) BITE: an R package for biodiversity analyses. *BioRxiv* 181610.
1014 <https://doi.org/10.1101/181610>

1015 Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. *Proc Natl Acad*
1016 *Sci U S A* 98:5446-5451. <https://doi.org/10.1073/pnas.091093398>

1017 Muirhead JR, Gray DK, Kelly DW, Ellis SM, Heath DD, Macisaac HJ (2008) Identifying the
1018 source of species invasions: sampling intensity vs. genetic diversity. *Mol Ecol* 17:1020-1035.
1019 <https://doi.org/10.1111/j.1365-294X.2008.03669.x>

1020 Nichols JT (1928) Chinese fresh-water fishes in the American Museum of Natural History's
1021 collections: a provisional check-list of the fresh-water fishes of China. *Bulletin of the AMNH*;
1022 v. 58, article 1

1023 Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, *Cottus*
1024 sp.(*Pisces, Teleostei*) in the Rhine with new habitat adaptations has originated from
1025 hybridization between old phylogeographic groups. *Proc R Soc Lond B Biol Sci* 272:2379-
1026 2387. <https://doi.org/10.1098/rspb.2005.3231>

1027 Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin
1028 hybrid zones suggest that genomic isolation differs among populations. *Mol Ecol* 18:2615-
1029 2627. <https://doi.org/10.1111/j.1365-294X.2009.04208.x>

1030 Oikonomou A, Leprieur F, Leonardos ID (2014) Biogeography of freshwater fishes of the
1031 Balkan Peninsula. *Hydrobiologia* 738:205-220. <https://doi.org/10.1007/s10750-014-1930-5>

1032 Olden JD, Poff NL, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and
1033 evolutionary consequences of biotic homogenization. *Trends Ecol Evol* 19:18-24.
1034 <https://doi.org/10.1016/j.tree.2003.09.010>

1035 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al (2012) Ancient admixture
1036 in human history. *Genetics* 192:1065-1093. <https://doi.org/10.1534/genetics.112.145037>

1037 Peischl S, Excoffier L (2015) Expansion load: recessive mutations and the role of standing

1038 genetic variation. *Mol Ecol* 24:2084-2094. <https://doi.org/10.1111/mec.13154>

1039 Perdereau E, Dedeine F, Christidès JP, Dupont S, Bagnères AG (2011) Competition
1040 between invasive and indigenous species: an insular case study of subterranean termites.
1041 *Biol Invasions* 13:1457-1470. <https://doi.org/10.1007/s10530-010-9906-5>

1042 Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in
1043 toads. *Nature* 439:803-803. <https://doi.org/10.1038/439803a>

1044 Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from Genome-
1045 Wide Allele Frequency Data. *PLoS Genet* 8:e1002967.
1046 <https://doi.org/10.1371/journal.pgen.1002967>

1047 Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T et al (2012) The
1048 genetic prehistory of southern Africa. *Nat Commun* 3(1):1-6.
1049 <https://doi.org/10.1038/ncomms2140>

1050 Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output
1051 analysis for MCMC. *R news* 6:7-11

1052 Price SJ, Garner TWJ, Cunningham AA, Langton TES, Nichols RA (2016) Reconstructing
1053 the emergence of a lethal infectious disease of wildlife supports a key role for spread
1054 through translocations by humans. *Proc R Soc B* 283:20160952.
1055 <https://doi.org/10.1098/rspb.2016.0952>

1056 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
1057 multilocus genotype data. *Genetics* 155:945-959. <https://doi.org/10.1093/sysbio/sys038>

1058 Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP (2016) Reliable ABC
1059 model choice via random forests. *Bioinformatics* 32:859-866.
1060 <https://doi.org/10.1093/bioinformatics/btv684>

1061 Puechmaille SJ (2016) The program structure does not reliably recover the correct
1062 population structure when sampling is uneven: subsampling and new estimators alleviate the
1063 problem. *Mol Ecol Resour* 16:608-627. <https://doi.org/10.1111/1755-0998.12512>

1064 Raynal L, Marin JM, Pudlo P, Ribatet M, Robert CP, Estoup A (2019) ABC random forests
1065 for Bayesian parameter inference. *Bioinformatics* 35:1720-1728.
1066 <https://doi.org/10.1093/bioinformatics/bty867>

1067 Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of
1068 colonising populations? *Trends Ecol Evol* 29:233-242.
1069 <https://doi.org/10.1016/j.tree.2014.02.003>

1070 Rius M, Turon X (2020) Phylogeography and the Description of Geographic Patterns in
1071 Invasion Genomics. *Front Ecol Evol* 8:595711. <https://doi.org/10.3389/fevo.2020.595711>

1072 Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic
1073 invasions. *Trends Ecol Evol* 22:454-464. <https://doi.org/10.1016/j.tree.2007.07.002>

1074 Salo P, Korpimäki E, Banks PB, Nordström M, Dickman CR (2007) Alien predators are
1075 more dangerous than native predators to prey populations. *Proc R Soc Lond B Biol Sci*
1076 274:1237-1243. <https://doi.org/10.1098/rspb.2006.0444>

1077 Sanz N, Araguas RM, Vidal O, Diez-del-Molino D, Fernández-Cebrián R, García-Marín JL
1078 (2013) Genetic characterization of the invasive mosquitofish (*Gambusia* spp.) introduced to
1079 Europe: population structure and colonization routes. *Biol Invasions* 15:2333-2346.
1080 <https://doi.org/10.1007/s10530-013-0456-5>

1081 Sax DF, Stachowicz JJ, Brown JH, Bruno JF, Dawson MN, Gaines SD et al (2007)
1082 Ecological and evolutionary insights from species invasions. *Trends Ecol Evol* 22:465-471.
1083 <https://doi.org/10.1016/j.tree.2007.06.009>

1084 Shafer AB, Gattepaille LM, Stewart RE, Wolf JB (2015) Demographic inferences using
1085 short-read genomic data in an approximate Bayesian computation framework: In silico
1086 evaluation of power, biases and proof of concept in Atlantic walrus. *Mol Ecol* 24:328-345.

1087 <https://doi.org/10.1111/mec.13034>

1088 Simberloff D (2013) Biological invasions: much progress plus several controversies.
1089 Contributions to Science 7-16. <https://doi.org/10.2436/20.7010.01.158>

1090 Simon A, Britton R, Gozlan R, Van Oosterhout C, Volckaert FA, Hänfling B (2011) Invasive
1091 cyprinid fish in Europe originate from the single introduction of an admixed source population
1092 followed by a complex pattern of spread. PLOS one 6:e18560.
1093 <https://doi.org/10.1371/journal.pone.0018560>

1094 Simon A, Gozlan RE, Britton JR, Van Oosterhout C, Hänfling B (2015) Human induced
1095 stepping-stone colonisation of an admixed founder population: the spread of topmouth
1096 gudgeon (*Pseudorasbora parva*) in Europe. Aquat Sci 77:17-25.
1097 <https://doi.org/10.1007/s00027-014-0374-3>

1098 Sinama M, Gilles A, Costedoat C, Corse E, Olivier JM, Chappaz R, Pech N (2013) Non-
1099 homogeneous combination of two porous genomes induces complex body shape trajectories
1100 in cyprinid hybrids. Front Zool 10:1-16. <https://doi.org/10.1186/1742-9994-10-22>

1101 Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML et al (2015)
1102 Genetic evidence for two founding populations of the Americas. Nature 525:104-108. <https://doi.org/10.1038/nature14895>

1104 South A (2017) Rnaturalearth: world map data from natural earth. R package version 0.1.0.
1105 <https://CRAN.R-project.org/package=rnaturalearth>.

1106 Stemshorn KC, Reed FA, Nolte AW, Tautz D (2011) Rapid formation of distinct hybrid
1107 lineages after secondary contact of two fish species (*Cottus* sp.). Mol Ecol 20:1475-1491.
1108 <https://doi.org/10.1111/j.1365-294X.2010.04997.x>

1109 Stiers I, Crohain N, Josens G, Triest L (2011) Impact of three aquatic invasive species on
1110 native plants and macroinvertebrates in temperate ponds. Biol Invasions 13:2715-2726.
1111 <https://doi.org/10.1007/s10530-011-9942-9>

1112 Tange O (2011) Gnu parallel-the command-line power tool. The USENIX Magazine 36:42-47

1113 Team RC (2019) R: A language and environment for statistical computing (version 3.1.2)
1114 Vienna, Austria. R Foundation for Statistical Computing; 2014

1115 Vallejo-Marín M, Friedman J, Twyford AD, Lepais O, Ickert-Bond SM, Streisfeld M et al
1116 (2021) Population genomic and historical analysis suggests a global invasion by bridgehead
1117 processes in *Mimulus guttatus*. Commun Biol 4:1-12. <https://doi.org/10.1038/s42003-021-01795-x>

1118

1119 van Boheemen LA, Lombaert E, Nurkowski KA, Gauffre B, Rieseberg LH, Hodgins KA
1120 (2017) Multiple introductions, admixture and bridgehead invasion characterize the
1121 introduction history of *Ambrosia artemisiifolia* in Europe and Australia. Mol Ecol 26:5421-
1122 5434. <https://doi.org/10.1111/mec.14293>

1123 Vandelannoote A, Yseboodt R (1998) Atlas van de Vlaamse beek-en riviervisserij. Water-
1124 Energik-vLario

1125 Verhoeven KJ, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological
1126 invasions and the balance between local adaptation and inbreeding depression. Proc R Soc
1127 Lond B Biol Sci 278:2-8. <https://doi.org/10.1098/rspb.2010.1272>

1128 Vilcinskis A (2015) Pathogens as Biological Weapons of Invasive Species. PLoS Pathog
1129 11:e1004714. <https://doi.org/10.1371/journal.ppat.1004714>

1130 Villéger S, Blanchet S, Beauchard O, Oberdorff T, Brosse S (2011) Homogenization
1131 patterns of the world's freshwater fish faunas. Proc Natl Acad Sci U S A 108:18003-18008.
1132 <https://doi.org/10.1073/pnas.1107614108>

1133 Watanabe K, Iguchi KI, Hosoya K, Nishida M (2000) Phylogenetic relationships of the
1134 Japanese minnows, *Pseudorasbora* (*Cyprinidae*), as inferred from mitochondrial 16S rRNA
1135 gene sequences. Ichthyol Res 47:43-50. <https://doi.org/10.1007/BF02674312>

1136 Weber E (1984) Die ausbreitung der pseudokeilfleckbarben im donauraum. Österreichs
1137 Fischerei 37:63-65

1138 Weiss S, Persat H, Eppe R, Schlötterer C, Uiblein F (2002) Complex patterns of
1139 colonization and refugia revealed for European grayling *Thymallus thymallus*, based on
1140 complete sequencing of the mitochondrial DNA control region. Mol Ecol 11:1393-1407.
1141 <https://doi.org/10.1046/j.1365-294x.2002.01544.x>

1142 Wolter C, Röhr F (2010) Distribution history of non-native freshwater fish species in
1143 Germany: how invasive are they? J Appl Ichthyol 26:19-27. <https://doi.org/10.1111/j.1439-0426.2010.01505.x>

1145 Yang JQ, Hsu KC, Liu ZZ, Su LW, Kuo PH, Tang WQ et al (2016) The population history of
1146 *Garra orientalis* (Teleostei: Cyprinidae) using mitochondrial DNA and microsatellite data with
1147 approximate Bayesian computation. BMC Evol Biol 16:73. <https://doi.org/10.1186/s12862-016-0645-9>

1149 Yuan JH, Cheng FY, Zhou SL (2012) Genetic structure of the tree peony (*Paeonia rockii*)
1150 and the Qinling Mountains as a geographic barrier driving the fragmentation of a large
1151 population. PLoS One 7:e34955. <https://doi.org/10.1371/journal.pone.0034955>

1152 Záhorská E, Kováč V (2009) Reproductive parameters of invasive topmouth gudgeon
1153 *Pseudorasbora parva* (Temminck and Schlegel, 1846) from Slovakia. J Appl Ichthyol 25:466-
1154 469. <https://doi.org/10.1111/j.1439-0426.2009.01190.x>

1155 Zhang C, Zhao Y (2016) Species diversity and distribution of inland fishes in China. Science
1156 Press, Beijing

1157 Zhao Y, Gozlan RE, Zhang C (2015) Current state of freshwater fisheries in China. Wiley-
1158 Blackwell, Oxford

1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180

1181 **List of Tables and Figures**

1182

1183 **Table 1.** Population properties with site specific sample, geographic range, geographical
1184 coordinates, and (N) number of individuals.

1185

1186 **Table 2.** Inference of the demographic model based on Neural Network and Random Forest.
1187 Selected models at each step are given for each algorithm (bold values are those of the
1188 selected model). 10,000 simulated datasets per scenario for training. Prior error rate
1189 estimated with leave-one-out cross-validation for Neural network and out-of-bag error rate
1190 for Random Forest. Random Forest does not provide posterior probability of each scenario,
1191 but rather the number of votes in favor to each one. Subsequently, posterior probability of
1192 the Random Forest corresponds to the posterior probability of the selected model.

1193

1194 **Fig. 1** Genetic structure and population modeling in the native range. **(a)** Posterior admixture
1195 proportions of individuals estimated with STRUCTURE for 18 sampled sites (n=300) in the
1196 native range (Asia) and K=6 genetic clusters (proportions aggregated with CLUMPP from 20
1197 independent replicates). **(b)** ML tree based on TreeMix with block size of 500 SNPs. The
1198 migration edges pointing from South China to Central China, South Central China, and North
1199 East China have migration weights of 45%, 33%, and 2,7%. The migration edge pointing
1200 from North China to North East China has a migration weight of 9%. **(c)** D-statistics for
1201 testing admixture in native populations. Populations are noted D(O, P3, P1, P2). Negative
1202 values indicate gene flow between P3 and P1 while positive values indicate gene flow
1203 between P3 and P2. Confidence interval at 95% of the D-statistic estimated by block
1204 jackknife of 5kb. Null expectation is a D-statistic of 0. Demes abbreviations are North=North
1205 China, South=South China, Central=Central China, NorthE=North East China,
1206 SouthE=South East China

1207

1208 **Fig. 2** Definition of demes in the native range, and source populations predicted for the
1209 invasive range. **(a) (Native range, Asia)** Mean admixture proportions estimated with
1210 STRUCTURE in the native range (Asia), for the chosen K=6. Native sampled sites are
1211 pooled into putative demes for subsequent analyses. Pie chart colors correspond to the
1212 genetic clusters inferred by STRUCTURE, as in Figure 1. **(Invasive range, Europe &
1213 Middle-East)** Assignment predictions of invasive individuals to native demes with
1214 AssignPOP's SVM algorithm with a relative posterior probability >2 (n=292 for training,
1215 n=100 for predictions). Pie chart colors correspond to the putative demes defined in the
1216 native range to which invasive individuals were assigned. **(b)** Posterior assignment
1217 probabilities to putative native demes estimated with AssignPop. **(c)** The demo-genetic
1218 scenario that was inferred with Approximate Bayesian Computation demonstrating three
1219 independent introductions from three independent admixed source populations. Branch
1220 lengths are not scaled. Bottleneck events are represented in thin red lines in branches.
1221 Colored branches correspond to invasive demes history.

1222

1223 **Fig. 3** Population differentiation. Pairwise F_{ST} (Weir & Cockerham's F_{ST}) between native
1224 and invasive demes. F_{ST} values were estimated from 3,000 loci.

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1236
1237

Table 1. Population properties with site specific sample, geographic range, geographical coordinates, and (N) number of individuals.

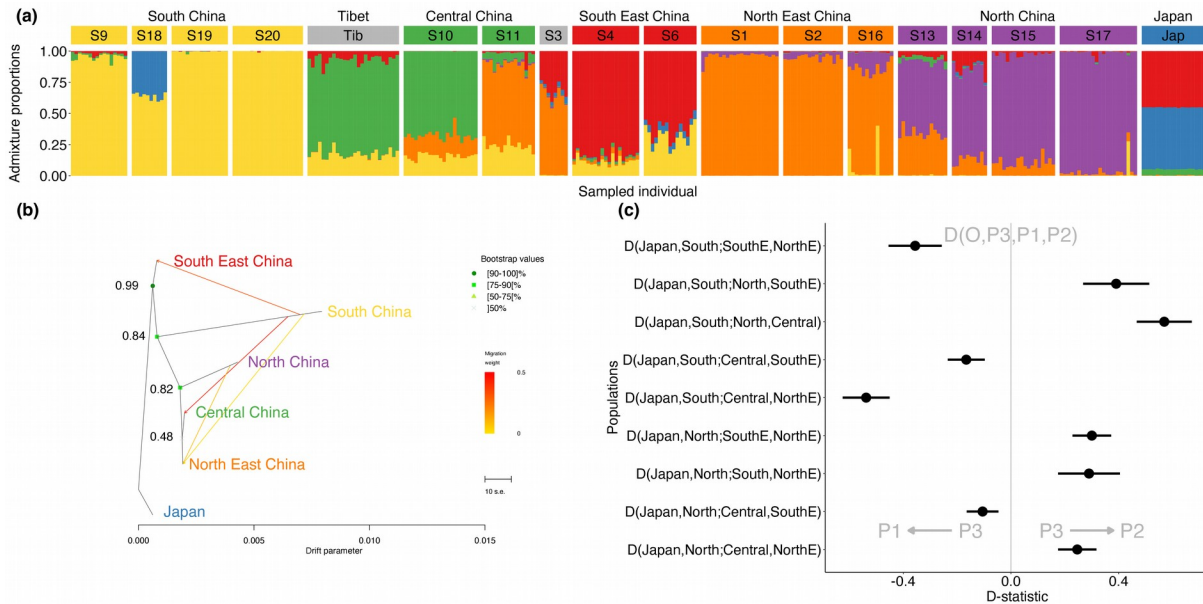
Country	Pop	Range	Coordinates		N
			Longitude	Latitude	
Japan	Jap	Asia	139.43	35.67	18
	S1	Asia	115.56	37.55	22
	S2	Asia	117.12	34.81	17
	S3	Asia	118.59	33.19	8
	S4	Asia	118.57	31.40	19
	S6	Asia	118.97	30.63	15
	S9	Asia	119.57	28.12	16
	S10	Asia	110.32	25.27	21
	S11	Asia	113.11	29.15	15
	S13	Asia	111.55	32.56	14
China	S14	Asia	110.99	34.62	10
	S15	Asia	122.52	40.10	18
	S16	Asia	124.99	45.03	13
	S17	Asia	122.93	42.64	22
	S18	Asia	118.27	40.9	10
	S19	Asia	116.89	43.3	16
	S20	Asia	115.17	24.74	20
	Tibet	Tib	Asia	99.94	23.56
Austria	Aus	Europe	99.53	23.35	17
Belgium	Bel	Europe	94.37	29.63	17
Bulgaria	Bul1	Europe	14.72	48.19	10
	Bul2	Europe	4.8	50.94	10
Hungary	Hun	Europe	26.85	44.06	22
Iran	Ira	Europe	26.7	43.99	10
Italy	Ita	Europe	18.87	46.63	18
Poland	Pol	Europe	54.78	37.05	12
Spain	Spa	Europe	10.52	44.77	19
Turkey	Tur	Europe	17.19	51.19	19
UK	UK	Europe	2.53	41.57	14

1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252

1253 Table 2. Inference of the demographic model based on Neural Network and Random Forest.
 1254 Selected models at each step are given for each algorithm (bold values are those of the
 1255 selected model). 10,000 simulated datasets per scenario for training. Prior error rate
 1256 estimated with leave-one-out cross-validation for Neural network and out-of-bag error rate
 1257 for Random Forest. Random Forest does not provide posterior probability of each scenario,
 1258 but rather the number of votes in favor to each one. Subsequently, posterior probability of
 1259 the Random Forest corresponds to the posterior probability of the selected model.

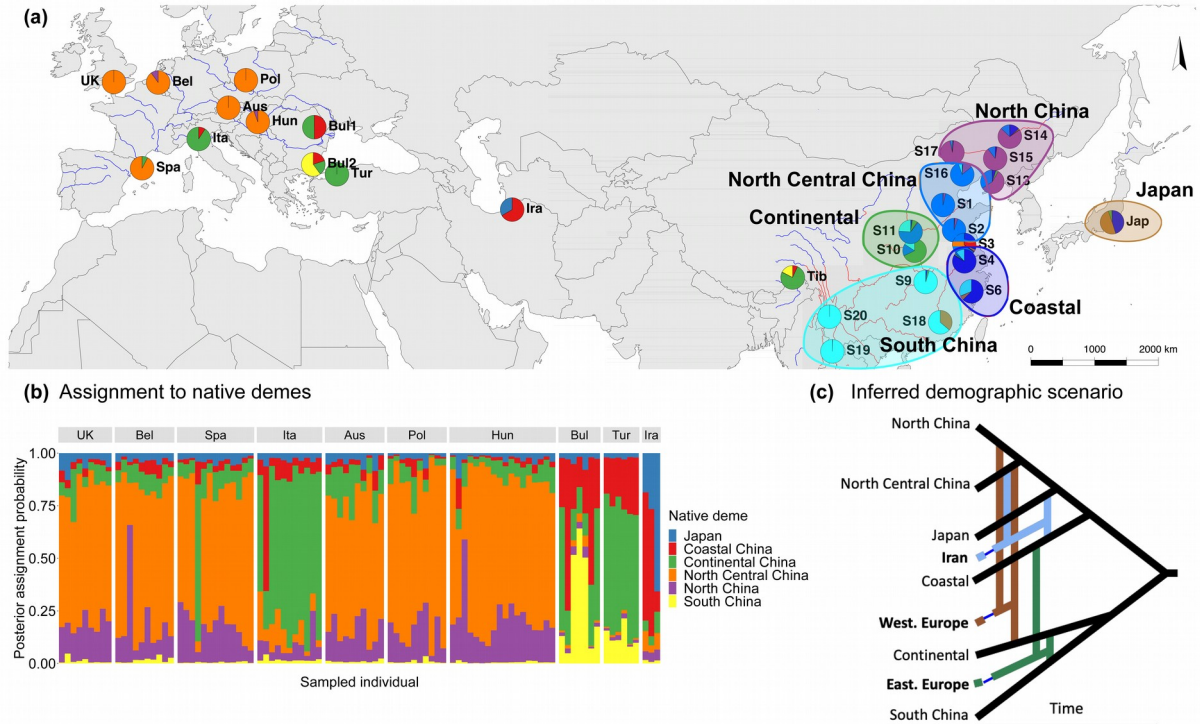
Scenario	Neural Network						Random Forest						
	Prior error rate	Posterior probabilities				Selected	Prior error rate	Number of votes				Posterior probability	Selected
		1	2	3	4			1	2	3	4		
Step 1. Source populations of invasive demes													
Western E.	0.03	0.03	0.02	0.1	0.85	4	0.05	175	175	192	458	0.87	4
Western E. (w/o Italy)	0.025	0.03	0.04	0.02	0.92	4	0.04	172	188	164	476	0.85	4
Eastern E.	0.04	0.02	0.03	0.03	0.92	4	0.06	104	121	136	639	0.79	4
Iran	0.04	0.04	0.12	0.04	0.80	4	0.07	60	210	36	694	0.8	4
Step 2. Global invasion pathways													
Global	0.09	0.99	0.006	0.004	-	1	0.03	859	124	17	-	0.47	1

1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286



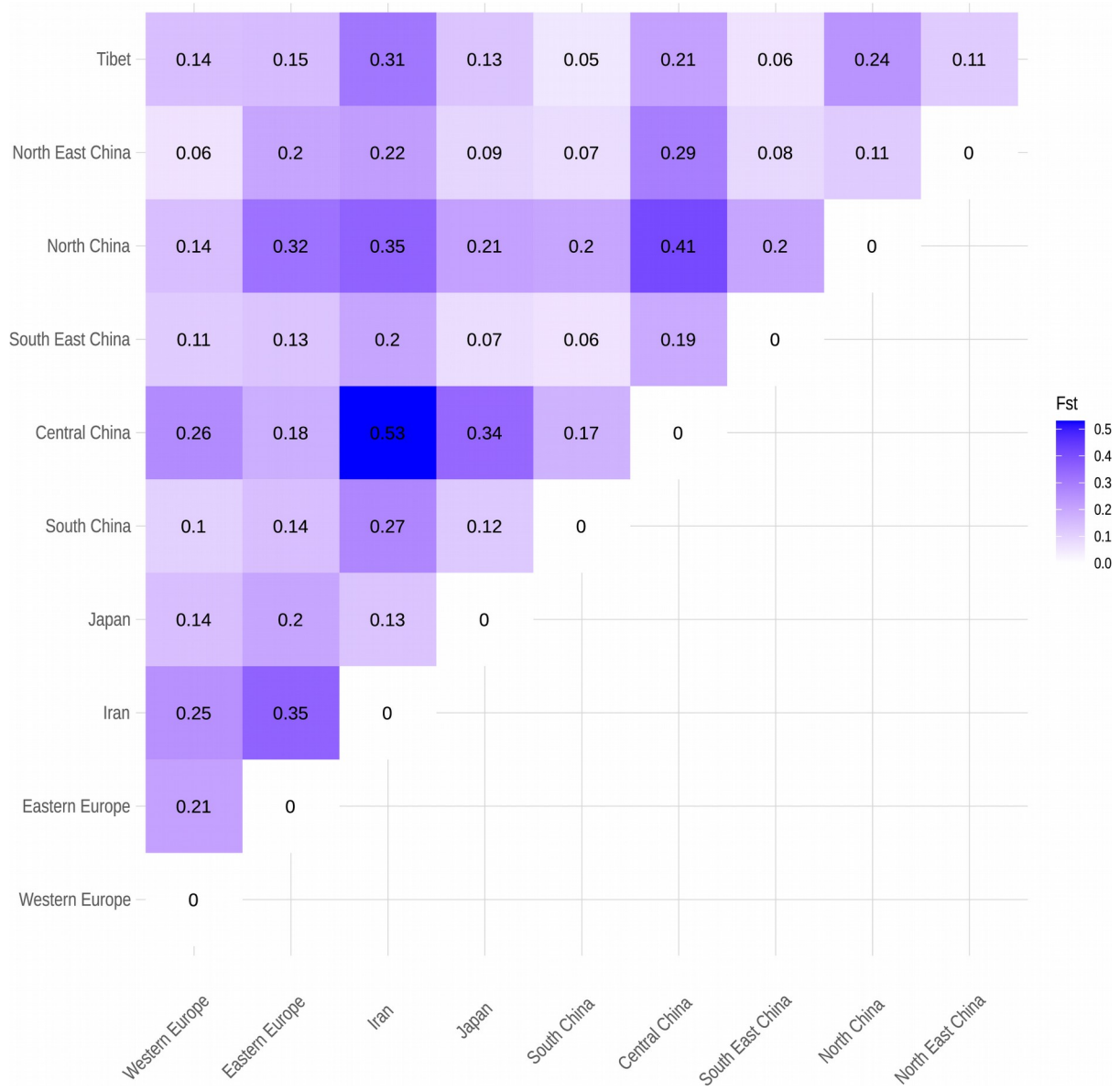
1287 **Fig. 1** Genetic structure and population modeling in the native range. **(a)** Posterior
 1288 admixture proportions of individuals estimated with STRUCTURE for 18 sampled sites
 1289 ($n=300$) in the native range (Asia) and $K=6$ genetic clusters (proportions aggregated
 1290 with CLUMPP from 20 independent replicates). **(b)** ML tree based on TreeMix with block
 1291 size of 500 SNPs. The migration edges pointing from South China to Central China, South
 1292 Central China, and North East China have migration weights of 45%, 33%, and 2,7%. The
 1293 migration edge pointing from North China to North East China has a migration weight of
 1294 9%. **(c)** D-statistics for testing admixture in native populations. Populations are noted
 1295 $D(O, P3, P1, P2)$. Negative values indicate gene flow between $P3$ and $P1$ while positive
 1296 values indicate gene flow between $P3$ and $P2$. Confidence interval at 95% of the D-
 1297 statistic estimated by block jackknife of 5kb. Null expectation is a D-statistic of 0. Demes
 1298 abbreviations are North=North China, South=South China, Central=Central China,
 1299 NorthE=North East China, SouthE=South East China

1300
 1301
 1302
 1303
 1304
 1305
 1306



1307 **Fig. 2** Definition of demes in the native range, and source populations predicted for the
 1308 invasive range. **(a) (Native range, Asia)** Mean admixture proportions estimated with
 1309 STRUCTURE in the native range (Asia), for the chosen $K=6$. Native sampled sites are
 1310 pooled into putative demes for subsequent analyses. Pie chart colors correspond to the
 1311 genetic clusters inferred by STRUCTURE, as in Figure 1. **(Invasive range, Europe &
 1312 Middle-East)** Assignment predictions of invasive individuals to native demes with
 1313 AssignPOP's SVM algorithm with a relative posterior probability >2 ($n=292$ for training,
 1314 $n=100$ for predictions). Pie chart colors correspond to the putative demes defined in the
 1315 native range to which invasive individuals were assigned. **(b)** Posterior assignment
 1316 probabilities to putative native demes estimated with AssignPop. **(c)** The demo-genetic
 1317 scenario that was inferred with Approximate Bayesian Computation demonstrating three
 1318 independent introductions from three independent admixed source populations. Branch
 1319 lengths are not scaled. Bottleneck events are represented in thin red lines in branches.
 1320 Colored branches correspond to invasive demes history.

1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340



1342 **Fig. 3** Population differentiation. Pairwise F_{ST} (Weir & Cockerham's F_{ST}) between
 1343 native and invasive demes. F_{ST} values were estimated from 3,000 loci.
 1344
 1345
 1346
 1347
 1348
 1349
 1350
 1351
 1352
 1353
 1354