

A tale of an invader: Reconstructing the genomic history of invasive topmouth gudgeon (*Pseudorasbora parva*) populations.

Thomas Brazier¹, Emira CHERIF², Jean-François Martin³, André Gilles⁴, Simon Blanchet⁵, Yahui Zhao⁶, Marine Combe², R.J. Scott McCairns¹, and Rodolphe Gozlan²

¹INRA

²IRD

³Montpellier-SupAgro, UMR CBGP

⁴Aix Marseille Université

⁵CNRS, Station d'Ecologie Théorique et Expérimentale (SETE)

⁶Chinese Academy of Sciences

February 24, 2021

Abstract

Invasive species are significant contributors to global changes and constitute a severe threat to biodiversity. Yet invasions offer an incredible framework to understand how small and low-diverse introduced populations adapt to novel environmental conditions and succeed in colonizing large areas. However, due to the insufficient data on the origin of the first introduced propagule and the first stage of invasion, reconstructing a species' invasion history is challenging. Here, we applied genetic clustering methods and explicit admixture tests combined with ABC models and Machine Learning algorithms to describe the phylogeography of native and invasive populations and infer the most probable demographic invasion scenarios of *Pseudorasbora parva*, a highly invasive freshwater fish and the healthy carrier of a novel lethal fungi-like pathogen (*Sphaerothecum destruens*), which is responsible for the decline of several fish species in Europe. We found that the current genetic structuring of the native *P. parva* range has been shaped by waves of gene flow originating from southern and northern Chinese populations. Furthermore, our results strongly suggest that the invasive genetic diversity is the outcome of past recurrent global invasion pathways of admixed native populations. Our study also illustrates how the combination of admixture tests, ABC, Machine Learning can be used to detect high-resolution demographic signatures and reconstruct an integrative biological invasion history.

1 INTRODUCTION

Invasive species are significant contributors to global changes and are often leading to biotic homogenization and biodiversity losses (Chapin Iii et al., 2000; Clavero & García-Berthou, 2005; Didham et al., 2005; Villéger et al., 2011; Simberloff, 2013). Globalization of trade and growth of worldwide transportation are the main drivers of non-native species introductions (Hulme, 2009). Major ecological impacts from biological invasions include a destabilization of trophic networks (Stiers et al., 2011; Gallardo et al., 2016), a competition for resources and habitats (Graebner et al., 2012; Perdureau et al., 2011), a predation on native species (Salo et al., 2007) and the transmission of novel infectious pathogens (Crowl et al., 2008). Invasions also lead to evolutionary consequences, such as genome introgression or a selection pressure on native species (Crispo et al., 2011; Mooney & Cleland, 2001; Sinama et al., 2013; Philips et al., 2006). Yet invasions offer an incredible framework to study adaptation to novel environmental conditions, as well as to understand how small introduced populations with supposedly low level of genetic diversity succeed in colonizing large areas (Bossdorf et al., 2005; Peischl & Excoffier, 2015; Sax et al., 2007; Roman & Darling, 2007; Facon et al., 2006).

Reconstructing the invasion history of a species is challenging. Typically, few individuals may constitute the initial colonizing group but data on the size of the propagule as well as its origin and time of first introduction are often lacking. Historical records of introduction events, census data or environmental monitoring projects often miss the first stages of invasion (Holsbeek et al., 2008; Mergeay et al., 2005). Molecular ecology allows the reconstruction of events that have not been directly observed but have left a genomic signature (Cristescu, 2015). Moreover, through the use of novel methods of demographic inference, it is now possible to identify source populations or even to model the demo-genetic processes underpinning invasion history (Beichman et al., 2018; Cabrera & Palsbøll, 2017; Shafer et al., 2015; Estoup & Guillemaud, 2010). Recent population genetic studies of invasion pathways have shown that invasion patterns could be more complex than previously thought (Lombaert et al., 2014). A single introduction event is rare and multiple introductions and/or admixture between source populations is often the cornerstone of successful invasions (Estoup et al., 2016; Roman & Darling, 2007).

The topmouth gudgeon *Pseudorasbora parva* (Temminck & Schlegel) has been the fastest invasive fish species in the world. It has a very large native distribution in East Asia (East China including Taiwan, Korea and Japan) with broad environmental tolerance from continental climates to tropical ones (Gozlan, 2012; Zhang & Zhao, 2016). It was initially accidentally released outside its native range via aquaculture partnership exchanges of Chinese carp between China and former USSR countries (Gozlan et al., 2010). During the 1960s, multiple introductions of *P. parva* took place all around the black sea area, followed by further introductions in the 80's in Eurasia and North Africa. After these initial phase of man-made introductions, natural local colonizations of entire river networks occurred across major European rivers to the Middle-East. Its life history traits, which include an early maturity (1 year) coupled to a nest guarding behavior, that ensure maximum probability of survival along with a short longevity has been identified as key traits to explain its rapid establishment and spread (Gozlan et al., 2010; Gozlan, 2012; Gozlan et al., 2020). A major biodiversity concern associated with *P. parva* is the fungi-like Rosette Agent *Sphaerothecum destruens*, for which it acts as a healthy carrier that has subsequently spread into invaded areas and in many naïve native host species. This novel pathogen is a lethal intracellular pathogen responsible for the mortality of a large number of European native freshwater fish species (Andreou & Gozlan, 2016; Combe & Gozlan, 2018).

Genetic structure and phylogeography, based upon classical mitochondrial and nuclear markers (microsatellites), have been extensively described in both native and invasive *P. parva* populations suggesting the existence of two genetic lineages within non-native populations in Europe (Hardouin et al., 2018; Simon et al., 2011, 2015). This general pattern has been recently confirmed by a study using 13,785 single nucleotide polymorphisms in Slovakian and Turkish introduced populations (Baltazar-Soares et al., 2020). Yet, the fine-scale genetic structure of the native populations and the gene flow that has shaped it, as well as the demographic dynamics underpinning the invasion pathways, remain unresolved. Assessing the source populations within the native range by extensive sampling along with high-throughput genotyping is crucial to detect high-resolution signatures of demographic history (Muirhead et al., 2008; Beichman et al., 2018; Shafer et al., 2015). Although such large datasets require advanced statistical methods that are computationally demanding, Approximate Bayesian Computation (ABC) and associated Machine Learning algorithms take advantage of large sets of summary statistics, exploring huge parameter spaces with reduced computational effort (Beaumont et al., 2002; Cabrera & Palsbøll, 2017; Pudlo et al., 2016; Raynal et al., 2017; Rey et al., 2015).

Here we took advantage of genomic signatures, ABC models and Machine Learning algorithms to i) characterize the genetic structure of native populations and test whether they aggregate on homogeneous and coherent demes and ii) retrace *P. parva*'s introduction history from Asia to Europe. To do so, we first applied genetic clustering methods and explicit tests of admixture to describe the phylogeography of both native and invasive ranges. We then identified putative source populations of introduced demes with population assignment tests. We finally used ABC model-based procedures to infer the most probable demographic scenarios of *P. parva* invasion and reconstruct an integrative biological invasion history.

2 MATERIALS AND METHODS

2.1 Sampling material

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

2.2 GBS sequencing

In total, 858 DNA samples (746 individuals in total, with 75 individuals replicated across at least one sequencing lane) were genotyped for single nucleotide polymorphism (SNP) markers by first digesting genomic DNA with PstI, followed by genotyping-by-sequencing (GBS), yielding an average of 2,702,000 raw sequencing reads per sample. SNP calling was performed using programs comprising the Stacks (v1.46) bioinformatics pipeline (Catchen et al., 2013). Replicated individuals were used to estimating genotyping error and only markers with less than 1% errors were retained. A minimum read depth of 20 was required for each marker, ultimately yielding 3,999 validated SNP markers. To prevent biases due to large proportions of missing data, populations with more than 70% missing data, individuals with more than 60% missing data and loci with more than 45% of missing data were removed from the dataset. After trimming, the final dataset contained 300 individuals from eighteen sites in the historic Asian range (including invasive populations in Tibet and South-West China) and 168 individuals from eleven invasive sites across Europe, Turkey and Iran, with on average 16 individuals per sampled site (Figure S1, Table 1).

2.3 Genetic diversity

Statistics of genetic diversity were estimated with the ‘adegenet’ and ‘hierfstat’ R packages (Goudet, 2005; Goudet & Jombart, 2015; Jombart, 2008; Jombart & Ahmed, 2011) for R version 3.5.3 (R Core Team, 2019) (Table S1). Sensitivity analyses on summary statistics showed that biases were minimized without loss of power when estimates were inferred from 2,000 to 3,000 loci with the lowest proportion of missing data (Figure S2). Consequently, the dataset for inferring population assignment consisted of 2,112 SNPs to reduce computational load (maximum 45% missing data per locus); however, up to 3,000 SNPs (maximum ~50% missing data) were retained for other analyses to keep most of the genetic information.

2.3 Genetic clustering

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

Additionally, we used STRUCTURE 2.3 to infer the number of genetic clusters and to estimate admixture between them (Pritchard et al., 2000). The admixture model with correlated allele frequencies was parameterized with a fixed Lambda value (parameter of the allele frequencies distribution) directly estimated from the data. Sampling location was set as a prior to improve inferences on weak genetic structure. STRUCTURE was performed for K=1-21 (number of sampled sites + 3) for Asian sites and K=1-14 for European and Middle-Eastern sites. Twenty replicates were computed for each K value, with 100,000 sampling iterations

after a burn-in of 100,000. The most probable K was assessed after considering the smallest value of K minimizing differences of likelihood (i.e. the plateau method of Pritchard et al., 2000), the highest value of ΔK given using Evanno’s method, computed with STRUCTURE HARVESTER (Earl & vonHoldt, 2012; Evanno et al., 2005) and the recent Puechmaille statistics (Puechmaille, 2016), implemented in the STRUCTURE SELECTOR web interface (Li & Liu, 2018). We then used CLUMPP to aggregate STRUCTURE replicates to produce mean individual admixture proportions with a ‘greedy’ search algorithm over 1,000 repetitions (Jakobsson & Rosenberg, 2007). Convergence among parameter sampling chains was assessed via CLUMPP’s H' statistic of similarity amongst several replicates. Additionally, within-chain convergence for parameter estimates (i.e. ancestry coefficient Alpha and Ln Likelihood) was assessed with diagnostics implemented in the R package ‘coda’ (Plummer et al., 2006).

Finally, based on the observed partitioning of genetic variance, sampled sites were pooled into demes (also referred to as populations) that made sense biologically (i.e. continuous gene flow among sites) and geographically. Admixture proportions inferred with STRUCTURE helped to delineate putative demes and were computed as the mean of the major ancestry coefficients (Q) in a given site. Sampled sites with a mean major ancestry coefficient (Q) under 70% were considered as admixed sites. Sampled sites that did not cluster well based on genetic markers were clustered into putative demes based on geographical and historical data. The consistency of our final clustering of sampled sites in genetic populations was assessed by a hierarchical AMOVA implemented in the ‘poppr’ R package (Kamvar et al., 2015, 2014) and the significance of variance proportions was tested with 1,000 random permutations (Table S3). Maps were drawn from the R package ‘maps’ with the ‘world’ database (Becker et al., 2018). The main river network was drawn from ‘RNaturalEarth’ (South, 2017).

2.4 Assignment to source populations with supervised machine learning

The R package ‘AssignPOP’ was used to assign invasive individuals to candidate source populations (i.e. native demes) with a machine learning classification algorithm (Chen et al., 2018). The supported algorithm was SVM (Support Vector Machine) trained with a Monte Carlo cross-validation method, with complementary validation by a K -fold procedure. One hundred training iterations were computed for different levels of training individuals (i.e. sampled proportions of individuals in each iteration of 0.5, 0.7 and 0.9) and training loci (i.e. sampled proportions of loci in each iteration of 0.25, 0.5 and 1). The training error rate was evaluated with the assignment accuracy test implemented in the package, which attempts to reassign known individuals to their source population. Source populations of invasive individuals were predicted from posterior assignment probabilities. Only individuals for whom the first probability was at least twice that of the second were retained as confident assignments.

2.5 Inference of invasion history with ABC

Approximate Bayesian Computation (ABC) was used to infer past demographic events shaping contemporary genetic diversity (Supporting information). We simulated large datasets under various invasions and admixture demographic scenarios and estimated the probability that data were observed under a given demographic scenario (Estoup et al., 2012). Scenarios were designed in a two-step hierarchical procedure of increasing complexity, with the second step derived from findings at step 1 (Figure S3, S4; see supplementary method for details). In the first step, we tested 3 independent sets of scenarios about the origins in the native range of 3 independent invasive demes (Western Europe, Eastern Europe and Iran; see Results). For each independent invasive deme, we tested if the source population was one of the three candidates or an admixture between the three candidates. To avoid the trap of infinite combinations of exhaustive scenarios, source populations predicted with population assignment served as initial candidate source populations. Choice of candidate source populations was also cross-validated with a Maximum-likelihood phylogenetic tree (Figure S8) inferred by the program ‘TreeMix’ version 1.13 (Pickrell & Pritchard, 2012). As the Italian sample’s origin was ambiguous in previous genetic clustering results, we performed the ABC analysis with and without Italian samples. The comparison of both replicated scenarios allowed us to assess the sensitivity of the selected scenario to Italian individuals (Table 2). The second step was dedicated to resolving a more complex invasion pattern, comparing competing worldwide invasion pathways encompassing all inva-

sive demes. Source populations of invasive demes were proved to be admixed (results of step 1, described below). Therefore, plausible hypotheses to test at step 2 were: (1) three independent introductions from three independent admixed source populations; (2) three independent introductions from a single admixed native population; or (3) a single continental introduction from an admixed native population.

Demo-genetic scenarios were simulated and summary statistics estimated, with DIYABC version 2.1 (Cornuet et al., 2014). Prior probabilities of scenarios were set to uniform. Parameter prior distributions were first set to a biologically reasonable range of values, then confidence in priors checked (*via* DIYABC test of goodness of fit), with distributions refined iteratively until the simulated scenarios fit the data (final parameter spaces given in supplementary methods). For each scenario, 10,000 simulations were conducted, estimating all available summary statistics in DIYABC for each (Cornuet et al., 2014). The best model was selected using two machine-learning algorithms trained on the simulated datasets: a Neural Network algorithm implemented in the R package ‘abc’ (Csilléry et al., 2012) and a Random Forest algorithm in the ‘abcrf’ package (Pudlo et al., 2016). One thousand Neural Networks were trained with 5 to 12 units of hidden neural layers. Parameter sets were weighted by an Epanechnikov kernel. The tolerance rate was 0.2. Other configuration values were set to the default value. Power of ABC model selection with Neural Network was evaluated by leave-one-out cross-validation repeated 100 times. Independently, 1,000 trees were grown in the Random Forest training set, with linear discriminant analysis scores added to summary statistics when it reduced the prior error rate. Power of Random Forest was evaluated by out-of-bag prior misclassification error rate. Lastly, the quality of the selected scenario was checked by comparing the marginal posterior predictive distributions to the observed values of the summary statistics. The marginal posterior predictive distribution was computed from 10,000 simulations under the selected scenario, parameterized with estimated posterior parameter distributions as priors.

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

2.6 Genotype phasing and imputation

Genotypes of the 468 sequenced individuals were phased and missing ones were imputed using Beagle 5.1 (Browning & Browning, 2007; Browning *et al.*, 2018) to perform population migration modeling and admixture tests. Beagle uses Hidden Markov Model (HMM) to infer the most likely haplotype pair for each individual and uses identity by descent (IBD) combined with an HMM to perform genotype imputation.

2.7 Native population migration modeling

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

2.8 Native population admixture test

We selected one deme from each mainland China region (North, North East, Central, South East and South) to perform admixture tests. We used the four-population test, *D-statistic* (Green *et al.*., 2010; Patterson *et al.*., 2012), implemented in Popstats (Skoglund *et al.*., 2015) to test for admixture and gene flow directionality within the native populations. The notation used by Popstats for the *D-statistic* is $D(O, P3; P1, P2)$, where O is the outgroup, P3 the test population and P1, P2 the sister populations. A significant negative D indicates that P3 exchanged genes with P1; conversely, a positive D indicates that P3 exchanged genes with P2 (Durand *et al.*., 2011). The D-statistic was estimated for each combination of demes. *D-statistic* significance was assessed by block jackknife of 5kb and the standard error (SE) was used to estimate the Z-score (Skoglund *et al.*., 2015). We considered a *D-statistic* to be significantly different from 0 when $|D| > 3 \times SE$.

3 RESULTS

3.1 Complex genetic structure of the native range

We searched for the genetic structure of the native range in order to aggregate sampled sites in fewer homogeneous and consistent demes, hence simplifying further demographic inferences based on ABC modelling. The highest level of structure estimated by STRUCTURE was $K=5-6$ (Figure 1a, Table S3, Supplementary information). To select the most plausible number of genetic clusters K, multiple criteria were assessed and two clustering methods cross-validated (Table S3, Supplementary information). The modal value of the distribution of ΔK suggested $K=3$ in the native area, whereas the shape of the plateau of $\ln(\Pr(X|K))$ suggested a value of K between 5 and 7. At $K=3$, Northern China was well separated from Central China/Japan and Southern China. However, at $K=6-7$, a better resolution on Central China (S4, S6, S10, S11) and Tibet was obtained, with signals of a north-south admixture. The overall patterns of clustering and admixture were similar between $K=6$ and $K=7$ with a spurious cluster at $K=7$. Convergence tests assessed that most sampling chains were convergent, yet some of the sampling chains remained non-convergent even after a burn-in of 100,000 and 100,000 sampling iterations. The highest values of H' was for $K=3$ (0.99), though H' reached 0.81 for $K=6$. Moreover, the DAPC approach yielded very similar results with a most plausible $K=5$. Besides, STRUCTURE and DAPC clustering were congruent and revealed the same separation between Northern and Southern China. The selected K was the one meeting both statistical, geographical and historical criteria. Hence $K=6$ was chosen as the parsimony value for the highest level of genetic structure.

3.2 Definition of native putative demes

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

3.3 Two main gene flow directions shaped the native populations

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

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

3.4 Genetic structure of the invasive range

Genetic clustering in the invasive range was used to narrow the putative number of non-native populations' origins to test (i.e. the number of scenarios). STRUCTURE results indicated a clear genetic structure in the European invasive range with $K=3$, despite some uncertainty around the actual highest level of genetic clustering (Table S3). Indeed, for all values of K between 2 and 9, Turkey and Bulgaria clustered together without admixture; Iran formed a similarly distinct cluster (Figures S5a-S5b). Conversely, Western Europe showed a pattern of admixture for all values of K and admixture persisted unless K was equal to the number of sites (Figures S5b-S5c). This indicated that Western Europe is most likely a single deme with a strong sub-structure. Furthermore, hierarchical STRUCTURE analysis performed on the subset of West European sites confirmed a substructure at this scale. Within West European sites (Aus, Bel, Hun, Ita, Pol, Spa, UK), $K=3-4$ was the most probable grouping according to Evanno's method and $K=5$ according to the plateau method (Figure S5c). $K=5$ was displayed for the Western European sub-structure (Figure S5c) because spurious admixture indicated overfitting for higher values (Figure S5b). H' was high (0.98) for $K=3-4-5$ in the STRUCTURE analysis of the Western European deme. Under Geweke's diagnostic and trace plots, sampling chains seemed convergent or close to convergence, yet some seemed to show departure from stationarity. Heidelberg and Welsh's diagnostic supported that most of the Markov Chains were a stationary distribution. Additionally, Gelman and Rubin's diagnostic assessed that replicated chains converged on similar values. Finally, despite the strong sub-structuring observed, the Western European deme was considered as a single consistent genetic population in ABC scenarios. Hence, we retained three independent demes in the invasive range formed by Western Europe, Eastern Europe and Iran.

3.5 Population assignment to source populations

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

3.6 ABC inference of source populations and invasion pathways

ABC simulations were first used to infer the most probable source population in Asia of each one of the three main demes identified in the non-native area (step 1), followed by discrimination between different

competing invasion pathways at a global scale (step 2).

Step 1: Source populations of non-native populations

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

Step 2: Invasion pathways

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

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

3.7 Invasion process and founding populations

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

4 DISCUSSION

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

Paleogeographic influence

Large-scale genetic structure was consistent with the known phylogeography of the species (Hardouin et al., 2018; Simon et al 2011). Likewise, the retained fine-scale structure in the native range was consistent with non-genetic data (historical, geographical and morphological data), suggesting that the estimated value of K is representative of actual genetic structure. Asian phylogeography revealed a complex genetic structure involving locally high gene flow and variable admixture. Results of STRUCTURE, TreeMix and D-statistics taken together suggest that the current genetic structuring of the native *P. parva* range has been shaped by waves of gene flow originating from southern and northern populations (edges of the natural native distribution). Fish phylogeography in China has been greatly influenced by geological events (Chiang et al., 2013; Li, 1981) and geological formations can be powerful barriers to gene flow (Brandley et al., 2010).

The complex geological history of the South China landmass acted as a barrier to gene flows and induced vicariance in common cyprinid species (Yang et al., 2016). Within China, the noticeable divergence was between the North and the South, with the Qinling Mountains acting as a strong biogeographic barrier around 2.6 billion years ago (Yuan et al., 2012; Dong et al., 2011). This North/South divergence was clearly assessed in genetic clustering results, explaining most of the genetic variance and corresponded to the separation between the Northern temperate climate and the Southern subtropical climate.

Yet within both biogeographical regions, gene flow was restricted and populations were mainly differentiated per river basins. In China north to the Qinling Mountains, the two main demes (i.e. North China and North East China; Figure 2a) were influenced by two different major river networks, the Amur River and the Yellow River (S1, S2 and S3 are located on the Haihe and Huaihe Rivers but both rivers were influenced by the Yellow River in history). South China (sites S9, S18; Figure 2a) was structured by the Yangtze River. Indeed, most native demes were congruent with the architecture of the main river networks. Historically, five *Pseudorasbora* species, based on morphological characteristics and their association to major river basins, were described as endemic to China (Nichols, 1928). These five species actually corresponded to the same *P. parva* species (Gozlan, 2012), though genetic clustering was largely in agreement with the distribution of historical morphotypes (Nichols, 1928). Yet admixture was also observed in locations at the boundaries between river basins (e.g. site S16 at the frontier between the Yellow and Amur river basins), indicating that recurrent migration has happened between demes.

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

Anthropogenic influence

Genetic structure with admixture has been reported in South China for other fishes, explained by coastal land and tributaries between river basins, especially for the Yangtze and Pearl Rivers (Yang et al., 2016). Admixed populations may be the result of anthropogenic modifications to the hydrological landscape, especially new dispersal pathways such as canals. Recent secondary contacts between the North and South may have been facilitated by the increased structural connectivity between river basins. The construction of the Lingqu Canal 2,200 years ago connected the Yangtze River (sites S9, S18 in South China and to a lesser extent S6 and S4 in South East China; Figure 2a) to the Pearl River (Fengshu, 1990). Moreover, the Beijing-Hangzhou Grand Canal connected the Yangtze River to the Huai He, Hai He and Yellow rivers in the east of China (AD 581-618), a region at the core of the two admixed regions (Central China, North East China and South East China; Figure 2a). Secondary contact due to human-induced corridors can increase local admixture between populations (Crispo et al., 2011) and canals serving as ecological corridors between two interbreeding cyprinid species have ultimately become hybrid zones (Guivier et al., 2019). However, the most important vector for long-distance dispersal in the native range is aquaculture expansion throughout central China. The two more admixed demes correspond to intensive aquaculture regions in China (Zhao et al., 2015; Gong & Tu, 1991). In Central China deme in particular, many ponds and reservoirs were intensively stocked during the 1950s-60s, with *P. parva* ultimately becoming the dominant species (Zhao et al., 2015; Gong and Tu, 1991). Massive human translocations for aquaculture have often led to biotic homogenization (Olden et al., 2004). Moreover, consecutive introductions represent a type of punctuated gene flow that produces admixed genotypes with high genetic diversity but unpredictable evolutionary effects (Crispo et al., 2011; Hasselman et al., 2014). In native locations, admixture is known for negative effects such as outbreeding depression and loss of local adaptation (Côte et al., 2014; Huff et al., 2011; Hufford et al., 2012), but in

novel environments, admixture may increase adaptive potential for translocated individuals (Verhoeven et al., 2011). For example, admixture in the wild between divergent sculpin populations increased their genetic diversity and facilitated colonization of new habitats (Nolte et al., 2009, 2005; Stenshorn et al., 2011).

4.2 The invasion

Invasion starts in Asia

Invasive populations have been successfully established at the edges of the native distribution. In South-West China, sites S19 and S20 are recently introduced populations (1980s) from South-East China and perfectly clustered with S9, S18 (Figure 2a; see also Hardouin et al., 2018). The Tibetan population was also introduced, commercial exchange being most probably the pathway for this invasion given that many living fishes are sold in Tibetan markets and are traditionally released into rivers (Gozlan, 2012). Although Tibet shows similarities with Central China, a continuum of populations between them were not sampled, making it difficult to determine if Tibet should be clustered with this deme or another.

Out of Asia

Three invasive demes were consistently discriminated as (1) Iran, (2) a small Balkan-Anatolian deme (Eastern Europe) and (3) a large pan-European deme (Western Europe). In addition, clear genetic sub-structuring was observed in the Western European deme. A larger number of sampled individuals from this area and a larger number of markers would be then essential to more precisely depict the West European deme's genetic structure. The Iranian population was particular, forming a consistent deme completely isolated from all European populations, consistent with observations of Hardouin et al. (2018), who revealed that Iran was a different mtDNA lineage. Differentiation between Eastern and Western Europe may be imputed to the former USSR era when on one side, the eastern introduction first opens and on the other commercial exchanges with western Europe were limited (see Britton & Gozlan, 2013). The Danube river ensures structural connectivity which has favoured colonization among a large central European region (Weiss et al., 2002), from Germany to Bulgaria through Austria and Hungary. However, regional biogeography could also partially explain the observed population structure. The Balkan region — represented by Bulgarian locations — is a biodiversity hotspot characterized by highly structured local populations of freshwater fishes (Oikonomou et al., 2014). In addition, rapid range expansion of *P. parva* over long distances in the invasive range relied on human-mediated dispersal (e.g. unintentional releases, aquaculture exchanges) rather than natural dispersal as *P. parva* favours lentic habitats. The observed genetic structure of invasive locations may suggest multiple independent introductions without subsequent gene flow. This assumption is strengthened by an invasive population genetic diversity similar to the native populations (Table S1). In a previous study, high genetic diversity among invasive *P. parva* has been reported, first by Simon et al. (2015) and later attributed by Hardouin et al. (2018) to an unbalanced sampling. Then, Baltazar-Soares et al (2020) highlighted a high genomic diversity attributed to a lack of admixture due to a recovery of the genetic bottleneck associated with the introduction. Here, with a balanced sampling (Invasive/Native) and the observed admixture in the invasive population (Figure 2b, Figure S5), our results suggest an invasive genetic diversity shaped by past invasions of admixed native populations.

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

4.3 Origins of invasion: admixture and multiple introductions

Our models described major demographic events and yielded congruent results between the two independent Machine Learning algorithms used to evaluate the origins of the three invasive demes. Results point towards a southern Chinese origin of the Eastern European deme, with alleles coming from and/or shared between the Central, South East and South China demes (Figure 2c). Similar patterns have been demonstrated via population assignment tests, revealing multiple origins in the same location indicative of multiple introductions or admixture within invasive demes (Guillemaud et al., 2011). Studies have also highlighted that admixture is likely a common characteristic for invasive species (Genton et al., 2005; Kolbe et al., 2004; Rius & Darling, 2014); however, it is often difficult to disentangle multiple introductions from admixture before introduction. The reconstruction of admixture history may depend on the genetic structure observed in the native area. Admixture after introduction may be inferred when source populations are clearly differentiated and admixture is demonstrated in the invasive area, as in the case of the mussel, *Mytella charruana* (Gillis et al., 2009). However, this is not the case for *P. parva*, which shows high admixture rates within the inferred source populations. The Central deme corresponded to the Wuhan region, at the core of Chinese aquaculture, with more than two million fishes produced every year (Zhao et al., 2015). Translocations from surrounding regions are common in order to restock ponds and reservoirs and have occurred over centuries in the Wuhan region (Zhao et al., 2015). Consequently, admixture in the native range before the translocation into Europe may be at the origin of the Eastern European deme. The original introduction probably occurred in Romania, close to Bulgaria, in the Nucet Fisheries Research Centre in 1961 (Gozlan et al., 2010) although several other introductions occurred at the same time all around the black sea region. Introduction was most likely accidental, as *P. parva* is a common contaminant of carp stockings (Wolter & Röhr, 2010). The epicentre of carp aquaculture in the 1960s was historically in Central Europe (e.g. Poland and Hungary) and around the Black Sea (e.g. Romania and Bulgaria), close to the probable original introductions of both European demes (Gozlan et al., 2010). Assuming between one and two generations per year among eastern European populations (Gozlan, 2008), it complies with our estimates of the time of invasion (ca. 18 to 62 generations) and the time of Asian fish exports increase. Likewise, development of aquaculture in this part of Europe may have subsequently facilitated the spread of *P. parva* in Turkey as soon as 1982 (Ekmekçi & Kirankaya, 2006; Britton & Gozlan, 2013).

The Western Europe deme was probably colonized from a first introduction in Hungary, noticed in the Paks Fisheries Farm in 1963 (Gozlan et al., 2010), which then spread throughout Western Europe. The source populations mainly differed from those of the Eastern Europe deme, although both share a connection with Central China. Western Europe populations also include an admixed north Chinese origin to both the North East China and North China demes. The invasion time was approximately the same as for Eastern Europe, coincident with a trading cooperation period between countries of the Eastern Bloc and China (Britton & Gozlan, 2013). On average, genetic divergence amongst source demes was relatively high (mean $F_{ST} = 0.27$; Table S5), suggesting limited gene flow and a lower likelihood of native admixture, compared to that for Eastern Europe source demes (mean $F_{ST} = 0.13$; Table S5). Moreover, individuals of the Hungarian location were systematically assigned to two genetic clusters (Figure S5c), making it the most admixed location of Western Europe and suggested further that admixture most likely happened from multiple introductions in the same invasive area. Temporal dynamics of invasive mosquitoes in the US have shown the same admixture-like clustering pattern (Fonseca et al., 2010). Range expansion across European countries most probably began in the Hungarian admixed region and its likely chronology has been reconstructed from census data (Gozlan et al., 2010). *P. parva* colonized Austria as soon as 1982 (Weber, 1984), then Germany from 1984 to 1987 (Arnold, 1985; Lelek & Köhler, 1989) and Belgium by 1992 (Vandelannoote & Yseboodt, 1998). Austrian, Belgian and Polish sites clustered together, suggesting a common ancestry and their connectivity along the Danube and Rhine river systems may have facilitated natural dispersal (Hegediš et al., 2007; Leuven et al., 2009). However, natural dispersal cannot explain the strong genetic structure observed, nor expansion over geographical barriers (e.g. the English Channel, the Alps). Thus, Hungary may be the invasive bridgehead that initiated the Western Europe colonization, followed by secondary bottlenecks associated with a series of founding events during range expansion to the West. Patterns of invasive bridgehead can be recognized by hierarchically structured populations within a deme (van Boheemen et al., 2017). In the UK, a small population was founded in the mid-1980s by artificial

release in an aquaculture in the south of England from a German source population. It has since spread to multiple large populations (Gozlan et al., 2002). Additionally, *P. parva* appeared in Spain in 2001 (Aparicio et al., 2012), colonizing the Iberian Peninsula following accidental releases in local fish farms (Caiola & Sostoa, 2002). Surprisingly, Italy showed different origins than other Western European locations in population assignment tests (Central China instead of North-East China; Figure 2b), but was nonetheless considered in the Western European deme based on genetic clustering results. The origin of this small population was difficult to explain, as it may be part of an unsampled fourth invasive population linked with Slovakia that successfully spread across the Italian peninsula (Carosi et al., 2016; Záhorská & Kováč, 2009). This uncertainty in the Italian origin underlined the necessity of the same extensive sampling in Europe as in Asia, especially in the Southeastern Europe that was believed to be the core of past introductions (e.g. Slovenia, Hungary, Romania, Bulgaria). This punctual sampling represents an inherent limitation to the definition of putative demes and subsequently to the number of invasion pathways inferred.

The Iranian population origin was complex to infer because of an ancient admixed origin between South East China and Japan. Iran showed a pattern of assignment similar to that of the Japanese admixture in nuclear markers between endemic and Chinese populations, potentially explaining why it was not possible to disentangle China and Japan as putative sources. In line with our results, a recent study (Ganjali et al., 2020), based on two mtDNA markers, identified three Iranian matrilineal haplotypes belonging to two distinct lineages. Two haplotypes from an older Japanese lineage traced back to the Iranian introduction of *P. parva* in the late 1980s and one haplotype from a Chinese lineage corresponding to a recent natural dispersal from Azerbaijan. They also highlighted an admixture of highly divergent Japanese and Chinese lineages. The mtDNA lineage of Iranian individuals defined by Hardouin et al. (2018) was, however, different from all lineages found in Chinese populations. Many questions remain, mainly because the history of Japanese populations has not been fully resolved. Despite importation of Chinese carp from the Amur river basin (Coad & Abdoli, 1993), we uncovered no evidence of a North China origin in Iran. The high number of private alleles in Iran (Table S1) and the long drift separating the Iranian population from the most recent common ancestor with Japan (Figure S8) might suggest that the true source population had not been sampled. Thus, additional sampling would be required to effectively determine the origins of Iranian *P. parva*.

4.4 A successful invasion

The surprisingly large introduced effective population sizes in Western Europe likely enhanced the probability of introduced populations survival and may explain its invasiveness. The gene flow resulting from the multiple introductions highlighted in Western Europe largely increased the effective population size as well as genetic diversity. However, introduced effective population sizes were smaller in Iran and Eastern Europe. Nevertheless, the bottleneck they experienced was not severe, with 100 to 200 individuals in the founder population. *P. parva* probably overcame the loss of diversity induced by bottlenecks due to admixture between native populations within its new environment and prior to range expansion, as evidenced by first colonizations reported twenty years after the first introduction. Our results demonstrated multiple source populations coming from a wide range of climatic conditions, the most flagrant being the admixture in the Wuhan region between subtropical and temperate populations, which represent different morphotypes (Nichols, 1928; Gozlan et al., 2020). Bridging sampling gaps in both native and invasive areas combined with wide-genome diversity capturing will be critical to bringing further insights into the invasive species' adaptive potential.

5 CONCLUSION

Our results shed light on the complex demo-genetic history of an invasive species' source populations and its spread across recurrent global invasion pathways. Understanding these aspects is of primary importance to disentangle one of the most critical ecological disturbances: biological invasion processes. Reconstruction of invasion pathways is not merely an evolutionary question, though. It could also be a powerful tool for conservation biology and management to prevent further non-native species introductions and potential associated pathogens.

Acknowledgements

We would like to thank all co-authors of Gozlan et al. 2020 for their help in collecting *P. parva* samples across the native and invasive range as well as B. Haenfling. This study is part of the project GENESIS ANR-AF 13-ADAP-0005-001 supported by the Agence Nationale de la Recherche. S.B. was co-funded by the project PROBIS (Biodiversa) and EC was funded by an IRD postdoctoral fellowship.

Data Accessibility

All sequencing data (GBS reads) will be submitted and publicly-available in the SRA repository.

The original genetic data yielded by genotyping (in vcf format), additional metadata (e.g. geographical coordinates) and all scripts necessary to reproduce the results presented in the paper will be available in a publicly-available Github repository: <https://github.com/ThomasBrazier/PseudorasboraGENESIS>.

Author Contributions

TB, EC performed research, analyzed the data. TB, EC and REG wrote the manuscript; R.E.G., SMcC, JFM, AG and SB designed research and contributed with data; Y.Z., REG, MC contributed with data; SMcC assisted with analyses. All authors commented on the manuscript. REG, AG, JFM and SB received funding for this research from ANR Bioadapt.

References

- Andreou, D., & Gozlan, R. E. (2016). Associated disease risk from the introduced generalist pathogen *Sphaerothecum destruens* : management and policy implications. *Parasitology*, 143(9), 1204-1210.
- ApAricio, E., Peris, B., Torrijos, L., Prenda, J., Nieva, A., & Perea, S. (2012). Expansion of the invasive *Pseudorasbora parva* (Cyprinidae) in the Iberian Peninsula: first record in the Guadiana River basin. *Cybium*, 36 (4), 585-586.
- Arnold, A. (1985). *Pseudorasbora parva* (Schlegel 1842) nun auch in der DDR. *Z. Binnenfisch. DDR*, 32, 182-183.
- Baltazar-Soares, M., Blanchet, S., Cote, J., Tarkan, A. S., Záhorská, E., Gozlan, R. E., & Eizaguirre, C. (2020). Genomic footprints of a biological invasion: Introduction from Asia and dispersal in Europe of the topmouth gudgeon (*Pseudorasbora parva*). *Molecular Ecology*, 29 (1), 71-85.
- Barnes, G. L. (2003). Origins of the Japanese Islands: The New "Big Picture". *Nichibunken Japan Review*, 3-50.
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162 (4), 2025-2035.
- Becker, O.S. code by R.A., Minka, A.R.W.R. version by R.B.E. by T.P., Deckmyn, A. (2018). Maps: draw geographical maps. R package version 3.3.0. <https://CRAN.R-project.org/package=maps>
- Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*.
- Blum, M. G., & François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and computing*, 20 (1), 63-73.
- Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., & Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, 144 (1), 1-11.
- Brandley, M. C., Guiher, T. J., Pyron, R. A., Winne, C. T., & Burbrink, F. T. (2010). Does dispersal across an aquatic geographic barrier obscure phylogeographic structure in the diamond-backed watersnake (*Nerodia rhombifer*)? *Molecular Phylogenetics and Evolution*, 57(2), 552-560.

- Britton, J. R., & Gozlan, R. E. (2013). Geo-politics and freshwater fish introductions: How the Cold War shaped Europe's fish allodiversity. *Global environmental change* , 23 (6), 1566-1574.
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics* , 103 (3), 338-348.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* , 81 (5), 1084-1097.
- Cabrera, A. A., & Palsbøll, P. J. (2017). Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIYABC. *Molecular Ecology Resources* , 17 (6), e94-e110.
- Caiola, N., & De Sostoa, A. (2002). First record of the Asiatic cyprinid *Pseudorasbora parva* in the Iberian Peninsula. *Journal of fish biology* , 4 (61), 1058-1060.
- Carosi, A., Ghetti, L., & Lorenzoni, M. (2016). Status of *Pseudorasbora parva* in the Tiber river basin (Umbria, central Italy) 20 years after its introduction. *Knowledge and Management of Aquatic Ecosystems* , (417), 22.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology* , 22 (11), 3124-3140.
- Chapin Iii, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., ... & Díaz, S. (2000). Consequences of changing biodiversity. *Nature* , 405 (6783), 234-242.
- Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsing, S. A. (2018). assign POP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine learning framework. *Methods in Ecology and Evolution* , 9 (2), 439-446.
- Chiang, T. Y., Lin, H. D., Zhao, J., Kuo, P. H., Lee, T. W., & Hsu, K. C. (2013). Diverse processes shape deep phylogeographical divergence in *Cobitis sinensis* (Teleostei : Cobitidae) in East Asia. *Journal of Zoological Systematics and Evolutionary Research* , 51(4), 316-326.
- Clavero, M., & Garcia-Berthou, E. (2005). Invasive species are a leading cause of animal extinctions. *Trends in ecology & evolution* , 20 (3), 110.
- Coad, B. W., & Abdoli, A. (1993). Exotic fish species in the fresh waters of Iran. *Zoology in the Middle East* , 9 (1), 65-80.
- Combe, M., & Gozlan, R. E. (2018). The rise of the rosette agent in Europe: An epidemiological enigma. *Transboundary and emerging diseases* , 65(6), 1474-1481.
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... & Estoup, A. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* , 30 (8), 1187-1189.
- Cote, J., Roussel, J. M., Le Cam, S., & Evanno, G. (2014). Outbreeding depression in Atlantic salmon revealed by hypoxic stress during embryonic development. *Evolutionary Biology* , 41 (4), 561-571.
- Crispo, E., Moore, J. S., Lee Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays* , 33 (7), 508-518.
- Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular ecology* , 24 (9), 2212-2225.

- Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., & Lugo, A. E. (2008). The spread of invasive species and infectious disease as drivers of ecosystem change. *Frontiers in Ecology and the Environment* , 6 (5), 238-246.
- Csillery, K., Francois, O., & Blum, M. G. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution* , 3 (3), 475-479.
- Didham, R. K., Tylianakis, J. M., Hutchison, M. A., Ewers, R. M., & Gemmell, N. J. (2005). Are invasive species the drivers of ecological change? *Trends in ecology & evolution* , 20 (9), 470-474.
- Dong, Y., Zhang, G., Neubauer, F., Liu, X., Genser, J., & Hauzenberger, C. (2011). Tectonic evolution of the Qinling orogen, China: review and synthesis. *Journal of Asian Earth Sciences* , 41 (3), 213-237.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution* , 28 (8), 2239-2252.
- Earl, D. A., vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* , 4 (2), 359-361.
- Ekmekci, F. G., & Kirankaya, Ş. G. (2006). Distribution of an invasive fish species, *Pseudorasbora parva* (Temminck & Schlegel, 1846) in Turkey. *Turkish Journal of Zoology* , 30 (3), 329-334.
- Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data : Why, how and so what?. *Molecular Ecology* , 19(19), 4113-4130.
- Estoup, A., Lombaert, E., Marin, J. M., Guillemaud, T., Pudlo, P., Robert, C. P., & Cornuet, J. M. (2012). Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular ecology resources* , 12 (5), 846-855.
- Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion?. *Annual Review of Ecology, Evolution, and Systematics* , 47 , 51-72.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* , 14 (8), 2611-2620.
- Facon, B., Genton, B. J., Shykoff, J., Jarne, P., Estoup, A., & David, P. (2006). A general eco-evolutionary framework for understanding bioinvasions. *Trends in ecology & evolution* , 21(3), 130-135.
- Fengshu, Z. (1990). Preliminary study of geological geomorphological conditions of ling canal. *Carsologica Sinica* , 1 .
- Fonseca, D. M., Widdell, A. K., Hutchinson, M., Spichiger, S. E., & Kramer, L. D. (2010). Fine-scale spatial and temporal population genetics of *Aedes japonicus* , a new US mosquito, reveal multiple introductions. *Molecular Ecology* , 19 (8), 1559-1572.
- Gallardo, B., Clavero, M., Sanchez, M. I., & Vila, M. (2016). Global ecological impacts of invasive species in aquatic ecosystems. *Global change biology* , 22 (1), 151-163.
- Ganjali, Z., Esmaeili, H. R., Zarei, F., Sayyadzadeh, G., Eagderi, S., & Gozlan, R. E. (2020). West Asian colonisation of topmouth gudgeon, *Pseudorasbora parva* (Teleostei : Gobionidae) : Genetic admixture at the crossroad of Europe and east Asia. *Freshwater Biology* .
- Genton, B. J., Shykoff, J. A., & Giraud, T. (2005). High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia* , as a result of multiple sources of introduction. *Molecular ecology* , 14 (14), 4275-4285.
- Gillis, N. K., Walters, L. J., Fernandes, F. C., & Hoffman, E. A. (2009). Higher genetic diversity in introduced than in native populations of the mussel *Mytella charruana* : evidence of population admixture at introduction sites. *Diversity and Distributions* , 15 (5), 784-795.

- Gong, M., & Tu, F. (1991). Fishery in contemporary China. *Contemporary China Press* , Beijing.
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* , 5 (1), 184-186.
- Goudet, J., & Jombart, T. (2015). hierfstat: estimation and tests of hierarchical F-statistics. *R package version 0.04-22* , 10 .
- Gozlan, R. E., Andreou, D., Asaeda, T., Beyer, K., Bouhadad, R., Burnard, D., ... & Robert Britton, J. (2010). Pan-continental invasion of *Pseudorasbora parva* : towards a better understanding of freshwater fish invasions. *Fish and Fisheries* , 11 (4), 315-340.
- Gozlan, R. E., Pinder, A. C., & Shelley, J. (2002). Occurrence of the Asiatic cyprinid *Pseudorasbora parva* in England. *Journal of Fish Biology* , 61 (1), 298-300.
- Gozlan, R. E., Zahorska, E., Cherif, E., Asaeda, T., Britton, J. R., Chang, C. H., ... & Combe, M. (2020). Native drivers of fish life history traits are lost during the invasion process. *Ecology and evolution* , 10(16), 8623-8633.
- Gozlan, R. E. (2008). Introduction of non-native freshwater fish: is it all bad? *Fish and fisheries* , 9 (1), 106-115.
- Gozlan, R. E. (2012). *Pseudorasbora parva* temminck and schlegel (topmouth gudgeon). Handb. Glob. Freshw. Invasive Species Abingdon: Earthscan.
- Graebner, R. C., Callaway, R. M., & Montesinos, D. (2012). Invasive species grows faster, competes better, and shows greater evolution toward increased seed size and growth than exotic non-invasive congeners. *Plant Ecology* , 213 (4), 545-553.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Paabo, S. (2010). A draft sequence of the Neandertal genome. *Science* , 328 (5979), 710-722.
- Guillemaud, T., Ciosi, M., Lombaert, E., & Estoup, A. (2011). Biological invasions in agricultural settings: insights from evolutionary biology and population genetics. *Comptes Rendus Biologies* , 334 (3), 237-246.
- Guivier, E., Gilles, A., Pech, N., DufLOT, N., Tissot, L., & Chappaz, R. (2019). Canals as ecological corridors and hybridization zones for two cyprinid species. *Hydrobiologia* , 830 (1), 1-16.
- Hardouin, E. A., Andreou, D., Zhao, Y., Chevret, P., Fletcher, D. H., Britton, J. R., & Gozlan, R. E. (2018). Reconciling the biogeography of an invader through recent and historic genetic patterns: the case of topmouth gudgeon *Pseudorasbora parva* . *Biological Invasions* , 20 (8), 2157-2171.
- Hasselman, D. J., Argo, E. E., McBride, M. C., Bentzen, P., Schultz, T. F., Perez-Umphrey, A. A., & Palkovacs, E. P. (2014). Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology* , 23 (5), 1137-1152.
- Hegediš, A., Lenhardt, M., Mićković, B., Cvijanović, G., Jarić, I., & Gačić, Z. (2007). Amur sleeper (*Perccottus glenii* Dubowski, 1877) spreading in the Danube River basin. *Journal of Applied Ichthyology* , 23 (6), 705-706.
- Holsbeek, G., Mergeay, J., Hotz, H., Plötner, J., Volckaert, F. A. M., & De Meester, L. (2008). A cryptic invasion within an invasion and widespread introgression in the European water frog complex: consequences of uncontrolled commercial trade and weak international legislation. *Molecular Ecology* , 17 (23), 5023-5035.
- Huff, D. D., Miller, L. M., Chizinski, C. J., & Vondracek, B. (2011). Mixed-source reintroductions lead to outbreeding depression in second-generation descendants of a native North American fish. *Molecular Ecology* , 20 (20), 4246-4258.

- Hufford, K. M., Krauss, S. L., & Veneklaas, E. J. (2012). Inbreeding and outbreeding depression in *Stylidium hispidum* : implications for mixing seed sources for ecological restoration. *Ecology and Evolution* , 2(9), 2262-2273.
- Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of applied ecology* , 46 (1), 10-18.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* , 23 (14), 1801-1806.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , 27 (21), 3070-3071.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* , 24 (11), 1403-1405.
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in genetics* , 6 , 208.
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* , 2 , e281.
- Karsten, M., Jansen van Vuuren, B., Addison, P., & Terblanche, J. S. (2015). Deconstructing intercontinental invasion pathway hypotheses of the Mediterranean fruit fly (*Ceratitidis capitata*) using a Bayesian inference approach: are port interceptions and quarantine protocols successfully preventing new invasions?. *Diversity and Distributions* , 21 (7), 813-825.
- Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature* , 431 (7005), 177-181.
- Konishi, M., Hosoya, K., & Takata, K. (2003). Natural hybridization between endangered and introduced species of *Pseudorasbora* , with their genetic relationships and characteristics inferred from allozyme analyses. *Journal of fish biology* , 63 (1), 213-231.
- Konishi, M., Sakano, H., & Iguchi, K. I. (2009). Identifying conservation priority ponds of an endangered minnow, *Pseudorasbora pumila* , in the area invaded by *Pseudorasbora parva* . *Ichthyological research* , 56 (4), 346.
- Lelek, A., & Köhler, C. (1989). Zustandsanalyse der Fischartengemeinschaften im Rhein (1987–1988). *Fischökologie* , 1 (1), 47-64.
- Leuven, R. S., van der Velde, G., Baijens, I., Snijders, J., van der Zwart, C., Lenders, H. R., & bij de Vaate, A. (2009). The river Rhine: a global highway for dispersal of aquatic invasive species. *Biological invasions* , 11(9), 1989.
- Li, S. (1981). *Studies on zoogeographical divisions for fresh water fishes of China* . Science Press.
- Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* , 18 (1), 176-177.
- Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., ... & Estoup, A. (2014). Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis* . *Molecular ecology* , 23 (24), 5979-5997.
- Mergeay, J., Verschuren, D., & De Meester, L. (2005). Cryptic invasion and dispersal of an American *Daphnia* in East Africa. *Limnology and Oceanography* , 50 (4), 1278-1283.
- Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J. F., Ajmone-Marsan, P., & Colli, L. (2017). BITE: an R package for biodiversity analyses. *BioRxiv* , 181610.

- Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences* , 98 (10), 5446-5451.
- Muirhead, J. R., Gray, D. K., Kelly, D. W., Ellis, S. M., Heath, D. D., & Macisaac, H. J. (2008). Identifying the source of species invasions: sampling intensity vs. genetic diversity. *Molecular Ecology* ,17 (4), 1020-1035.
- Nichols, J. T. (1928). Chinese fresh-water fishes in the American Museum of Natural History's collections: a provisional check-list of the fresh-water fishes of China. Bulletin of the AMNH; v. 58, article 1.
- Nolte, A. W., Freyhof, J., Stemshorn, K. C., & Tautz, D. (2005). An invasive lineage of sculpins, *Cottus* sp.(*Pisces* ,*Teleostei*) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups.*Proceedings of the Royal Society B: Biological Sciences* ,272 (1579), 2379-2387.
- Nolte, A. W., Gompert, Z., & Buerkle, C. A. (2009). Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology* ,18 (12), 2615-2627.
- Oikonomou, A., Leprieur, F., & Leonardos, I. D. (2014). Biogeography of freshwater fishes of the Balkan Peninsula. *Hydrobiologia* ,738 (1), 205-220.
- Olden, J. D., Poff, N. L., Douglas, M. R., Douglas, M. E., & Fausch, K. D. (2004). Ecological and evolutionary consequences of biotic homogenization. *Trends in ecology & evolution* , 19 (1), 18-24.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., ... & Reich, D. (2012). Ancient admixture in human history.*Genetics* , 192 (3), 1065-1093.
- Peischl, S., & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Molecular ecology* , 24 (9), 2084-2094.
- Perdereau, E., Dedeine, F., Christidès, J. P., Dupont, S., & Bagnères, A. G. (2011). Competition between invasive and indigenous species: an insular case study of subterranean termites. *Biological Invasions* , 13 (6), 1457-1470.
- Phillips, B. L., Brown, G. P., Webb, J. K., & Shine, R. (2006). Invasion and the evolution of speed in toads. *Nature* , 439(7078), 803-803.
- Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings* , 1-1.
- Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., ... & Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature communications* , 3 (1), 1-6.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news* ,6 (1), 7-11.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* ,155 (2), 945-959.
- Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests.*Bioinformatics* , 32 (6), 859-866.
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular ecology resources* , 16 (3), 608-627.
- Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference.*Bioinformatics* , 35 (10), 1720-1728.
- Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations?. *Trends in ecology & evolution* , 29 (4), 233-242.

- Roman, J., & Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in ecology & evolution* , 22 (9), 454-464.
- Salo, P., Korpimäki, E., Banks, P. B., Nordström, M., & Dickman, C. R. (2007). Alien predators are more dangerous than native predators to prey populations. *Proceedings of the Royal Society B: Biological Sciences* , 274 (1615), 1237-1243.
- Sanz, N., Araguas, R. M., Vidal, O., Díez-del-Molino, D., Fernández-Cebrián, R., & García-Marín, J. L. (2013). Genetic characterization of the invasive mosquitofish (*Gambusia* spp.) introduced to Europe: population structure and colonization routes. *Biological invasions* , 15 (10), 2333-2346.
- Sax, D. F., Stachowicz, J. J., Brown, J. H., Bruno, J. F., Dawson, M. N., Gaines, S. D., ... & Rice, W. R. (2007). Ecological and evolutionary insights from species invasions. *Trends in ecology & evolution* , 22 (9), 465-471.
- Shafer, A. B., Gattepaille, L. M., Stewart, R. E., & Wolf, J. B. (2015). Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: In silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology* , 24 (2), 328-345.
- Simberloff, D. (2013). Biological invasions: much progress plus several controversies. *Contributions to Science* , 7-16.
- Simon, A., Britton, R., Gozlan, R., Van Oosterhout, C., Volckaert, F. A., & Hanfling, B. (2011). Invasive cyprinid fish in Europe originate from the single introduction of an admixed source population followed by a complex pattern of spread. *PLOS one* , 6 (6), e18560.
- Simon, A., Gozlan, R. E., Britton, J. R., Van Oosterhout, C., & Hanfling, B. (2015). Human induced stepping-stone colonisation of an admixed founder population: the spread of topmouth gudgeon (*Pseudorasbora parva*) in Europe. *Aquatic sciences* , 77(1), 17-25.
- Sinama, M., Gilles, A., Costedoat, C., Corse, E., Olivier, J. M., Chappaz, R., & Pech, N. (2013). Non-homogeneous combination of two porous genomes induces complex body shape trajectories in cyprinid hybrids. *Frontiers in zoology* , 10(1), 1-16.
- Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hunemeier, T., Petzl-Erler, M. L., ... & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature* , 525 (7567), 104-108.
- South, A. (2017). Rnatural: world map data from natural earth. *R package version 0.1.0* . <https://CRAN.R-project.org/package=rnatural>.
- Stemshorn, K. C., Reed, F. A., Nolte, A. W., & Tautz, D. (2011). Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Molecular Ecology* , 20 (7), 1475-1491.
- Stiers, I., Crohain, N., Josens, G., & Triest, L. (2011). Impact of three aquatic invasive species on native plants and macroinvertebrates in temperate ponds. *Biological Invasions* , 13 (12), 2715-2726.
- Tange, O. (2011). Gnu parallel-the command-line power tool. *The USENIX Magazine* , 36 (1), 42-47.
- Team, R. C. (2019). R: A language and environment for statistical computing (version 3.1.2). Vienna, Austria. R Foundation for Statistical Computing; 2014.
- van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology* , 26 (20), 5421-5434.
- Vandelannoote, A., & Yseboodt, R. (1998). *Atlas van de Vlaamse beek-en riviervissen* . Water-Energik-vLario.

- Verhoeven, K. J., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B: Biological Sciences* , 278(1702), 2-8.
- Villegger, S., Blanchet, S., Beauchard, O., Oberdorff, T., & Brosse, S. (2011). Homogenization patterns of the world's freshwater fish faunas. *Proceedings of the National Academy of Sciences* , 108(44), 18003-18008.
- Watanabe, K., Iguchi, K. I., Hosoya, K., & Nishida, M. (2000). Phylogenetic relationships of the Japanese minnows, *Pseudorasbora* (*Cyprinidae*), as inferred from mitochondrial 16S rRNA gene sequences. *Ichthyological research* , 47 (1), 43-50.
- Weber, E. (1984). Die ausbreitung der pseudokeilfleckbarben im donauraum. *Osterreichs Fischerei* , 37 , 63-65.
- Weiss, S., Persat, H., Eppe, R., Schlotterer, C., & Uiblein, F. (2002). Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus* , based on complete sequencing of the mitochondrial DNA control region. *Molecular Ecology* , 11 (8), 1393-1407.
- Wolter, C., & Rohr, F. (2010). Distribution history of non-native freshwater fish species in Germany: how invasive are they?. *Journal of Applied Ichthyology* , 26 , 19-27.
- Yang, J. Q., Hsu, K. C., Liu, Z. Z., Su, L. W., Kuo, P. H., Tang, W. Q., ... & Lin, H. D. (2016). The population history of *Garra orientalis* (*Teleostei* : *Cyprinidae*) using mitochondrial DNA and microsatellite data with approximate Bayesian computation. *BMC Evolutionary Biology* , 16 (1), 73.
- Yuan, J. H., Cheng, F. Y., & Zhou, S. L. (2012). Genetic structure of the tree peony (*Paeonia rockii*) and the Qinling Mountains as a geographic barrier driving the fragmentation of a large population. *PLoS One* , 7 (4), e34955.
- Zahorska, E., & Kovač, V. (2009). Reproductive parameters of invasive topmouth gudgeon *Pseudorasbora parva* (Temminck and Schlegel, 1846) from Slovakia. *Journal of Applied Ichthyology* , 25 (4), 466-469.
- Zhang, C., & Zhao, Y. (2016). Species diversity and distribution of inland fishes in China. Beijing: Science Press.
- Zhao, Y., Gozlan, R. E., & Zhang, C. (2015). *Current state of freshwater fisheries in China* . Oxford: Wiley-Blackwell.

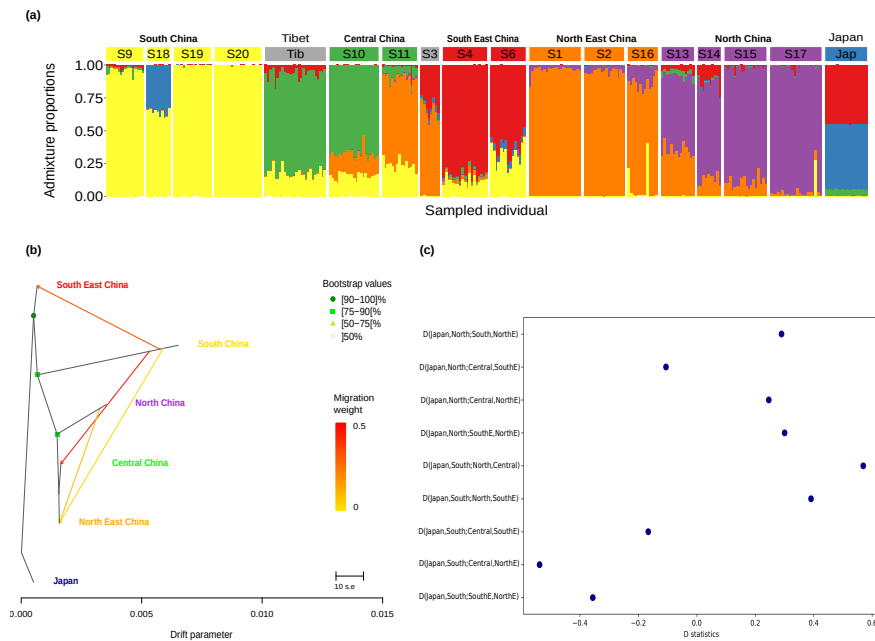
List of Tables and Figures

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

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

Figure 1. Genetic structure and population modeling in the native range. **(a)** Posterior admixture proportions of individuals estimated with STRUCTURE for 18 sampled sites (n=300) in the native range (Asia) and K=6 genetic clusters (proportions aggregated with CLUMPP from 20 independent replicates). **(b)** ML tree based on TreeMix with block size of 500 SNPs. The migration edges pointing from South China to Central China, South Central China, and North East China have migration weights of 45%, 33%, and 2,7%. The migration edge pointing from North China to North East China has a migration weight of 9%. **(c)** D statistics. Demes abbreviations are North=North China, South=South China, Central=Central China, NorthE=North East China, SouthE=South East China.

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



Hosted file

BrazieretalFigure_2.pdf available at <https://authorea.com/users/397838/articles/510606-a-tale-of-an-invader-reconstructing-the-genomic-history-of-invasive-topmouth-gudgeon-pseudorasbora-parva-populations>

Hosted file

Table1.pdf available at <https://authorea.com/users/397838/articles/510606-a-tale-of-an-invader-reconstructing-the-genomic-history-of-invasive-topmouth-gudgeon-pseudorasbora-parva-populations>

Hosted file

Table2.pdf available at <https://authorea.com/users/397838/articles/510606-a-tale-of-an-invader-reconstructing-the-genomic-history-of-invasive-topmouth-gudgeon-pseudorasbora-parva-populations>