Reconciling the biogeography of an invader through recent and historic genetic patterns: the case of topmouth gudgeon Pseudorasbora parva

Emilie A Hardouin, Demetra Andreou, Yahui Zhao, Pascale C Chevret, David Fletcher, J. Robert Britton, Rodolphe E Gozlan

CrossMark

ORIGINAL PAPER

Reconciling the biogeography of an invader through recent and historic genetic patterns: the case of topmouth gudgeon *Pseudorasbora parva*

Emilie A. Hardouin · Demetra Andreou · Yahui Zhao · Pascale Chevret ·
David H. Fletcher · J. Robert Britton · Rodolphe E. Gozlan

Abstract The genetic variability and population structure of introduced species in their native range are potentially important determinants of their invasion success, yet data on native populations are often poorly represented in relevant studies. Consequently, to determine the contribution of genetic structuring in the native range of topmouth gudgeon *Pseudorasbora parva* to their high invasion success in Europe, we used a dataset comprising of 19 native and 11 non-native populations. A total of 666 samples were analysed at 9 polymorphic microsatellite loci and sequenced for 597 bp of mitochondrial DNA. The analysis revealed three distinct lineages in the native range, of which two haplogroups were prevalent in China (100%), with a general split around the Qinling Mountains. Dating of both haplogroups closely matched past geological events. More recently, its distribution has been influenced by fish movements in aquaculture, resulting in gene flow between previously separated populations in Northern and Southern China. Their phylogeography in Europe indicate as few as two introductions events and two dispersal routes. Microsatellite data revealed native populations had higher genetic diversity than those in the invasive range, a contrast to previous studies on *P. parva*. This study confirms the importance of extensive sampling in both the native and non-native range of invasive species in evaluating the influence of genetic variability on invasion success.

E. A. Hardouin (✉) · D. Andreou · D. H. Fletcher ·
J. R. Britton · R. E. Gozlan
Department of Life and Environmental Sciences, Faculty of Sciences and Technology, Bournemouth University, Dorset BH12 5BB, UK
e-mail: ehardouin@bournemouth.ac.uk

Y. Zhao
Institute of Zoology, Chinese Academy of Sciences, Chaoyang District, Beijing 100101, China

P. Chevret
Laboratoire de Biométrie et Biologie Evolutive (UMR CNRS 5558), Université Claude Bernard Lyon, 1 Bat. Mendel, 43 bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

R. E. Gozlan
UMR BOREA IRD-MNHN-Université Pierre et Marie Curie, Muséum National d'Histoire Naturelle, 47 rue Cuvier, 75231 Paris Cedex 5, France

Introduction

A central theme in evolutionary biology is to understand the processes of how species adapt to changing environmental conditions (Bell and Gonzalez 2011; Ghalambor et al. 2007). Although empirical studies have confirmed some important theoretical predictions regarding how population size influences neutral genetic variation, it remains poorly understood how, for example, initial small population sizes affect the long-term survival of populations (Hedge et al. 2014; Hufbauer et al. 2013). Biological invasions are well suited to addressing such fundamental questions since they represent a natural translocation experiment over multiple generations (Blanchet 2012; Vandepitte et al. 2014). Abiotic and biotic conditions often differ between native and invaded environments, with the introduced species not sharing recent evolutionary history with species in the native community (Gallien et al. 2012). The colonisation of the new environment by the introduced species might be impeded by factors including competition for food resources and the unpredictable impacts of native parasites on new hosts (Kumschick et al. 2015), which can constrain its ability to establish self-sustaining populations. Thus, if they are to be successful, invasive species must respond quickly and efficiently to the selective regime imposed by the colonised ecosystem (Broennimann and Guisan 2008).

Introductions of non-native species are often based on the release of a low number of founding propagules containing only a fraction of the genetic variation of the source populations (Hanfling 2007; Searle 2008). During this process, these populations often undergo a succession of genetic bottlenecks, resulting in their invasive populations having lower genetic diversity compared with populations in the native range (Besnard et al. 2014; Facon et al. 2011; Hardouin et al. 2010, 2015). Such reduced genetic diversity theoretically limits a species' ability to establish invasive populations, invoking a genetic paradox over how introduced species can successfully develop invasive populations despite this low diversity (Hanfling 2007).

This paradox of genetic diversity is often counteracted by admixture among genetically divergent source populations upon their introduction (Roman and Darling 2007; Rosenthal et al. 2008). For example, multiple introductions have resulted in high genetic diversity of invasive crustaceans, fish and lizards

(Bermond et al. 2012; Biebach and Keller 2012; Chapple et al. 2013; Keller et al. 2014; Schulte et al. 2013). This can increase the level of heterozygosity in newly established populations and lead to increased fitness through heterosis (Keller et al. 2014). Sampling biases, however, such as extensive sampling in the invasive range and under sampling in the native range, often bias true estimates of genetic variability, phenotypic traits and behaviours, impeding accurate comparisons across the two ranges (Diez-del-Molino et al. 2013; Sanz et al. 2013). Such issues are particularly apparent in studies on freshwater fishes due to the inherent difficulties and expense of extensive sampling programmes (Smith et al. 2014). Since each invaded river catchment potentially represents a biogeographical island where populations may have locally adapted following separation and isolation through, for example, past geological events, it is important to have an extensive representation of variability in the native range (DeFaveri et al. 2014; Wiens et al. 2014).

Small freshwater fishes are often accidentally introduced into new environments via contaminated batches of other fishes used in angling or aquaculture (Britton and Gozlan 2013; Davies et al. 2013) and so invasions of small-bodied freshwater fish are often strongly associated with aquaculture trade pathways (Copp et al. 2010). The topmouth gudgeon *Pseudorasbora parva,* a small cyprinid fish, is naturally distributed in eastern Asia. It was introduced into Europe from China in the 1960s through a succession of accidental introductions into the area around the Black Sea when they contaminated batches of Chinese carp species being moved in aquaculture (Gozlan et al. 2010a; Gozlan 2011). It has now invaded at least 32 countries, including most of Europe, plus Turkey, Iran and North Africa, with their long-distance dispersal also occurring via aquaculture trade routes (Gozlan 2011). *P. parva*'s expansion is not, however, just limited to a non-Asian range, with their translocation to numerous locations within China. The native range of *P. parva* in China encompasses the Yangtzee River around the Qinling Mountains. Human translocations have led to non-native *P. parva* populations in western China, Tibet and Mongolia (Gozlan et al. 2010a, b). The wide geographic range invaded by *P. parva* thus makes this a strong model species for understanding invasion genetics. To date, the two studies that address the genetic diversity of *P. parva* (Simon et al.

2011, 2015) only included a very restricted sampling of their native range. In those studies, higher levels of genetic variability in the introduced range were detected than in the native range, highlighting the potential role of admixture in the invasive range in benefitting invasive vigour (Simon et al. 2011). However, the samples from the native range used potentially biased these conclusions, constraining understanding of their invasion history. Moreover, placing greater emphasis on sampling in the native range, can reveal a more complete picture of the long-term native biogeographic history of the species, including the role of geological events in shaping genetic topology (i.e. lineages), which could have influenced recent invasions (Diez-del-Molino et al. 2013).

Given the potential utility of invasive and native *P. parva* to assist in the understanding of the role of genetic variability in driving invasions, here we use a new and extensive collection of *P. parva* population samples from both their native and invasive range to: (1) characterise the present genetic diversity of *P. parva* by determining its spatial population structure across its major native river catchments; (2) characterise the genetic variability in the invasive range in order to better understand how it has been influenced by the invasion process; and (3) determine the influence of past geological and anthropogenic events on the current genetic structure of *P. parva*.

Materials and methods

Sampling

Samples of *P. parva* from across their native range in China were collected in 17 river catchments, from the Amur River basin in the North to the Pearl River basin in the South (Table 1). Additional samples were obtained from Taiwan and Japan (one population per country). Concomitantly, *P. parva* samples from their invasive range were collected from 18 countries across Eurasia (Table 1). In all cases, the sampling method was appropriate for the water body in question, utilising fish traps, electric fishing and/or micro-mesh seine netting. In the field, the fish were euthanized with an overdose of anaesthetic, preserved in 95% ethanol and then stored at − 70 °C. In the laboratory, each fish

was identified as *P. parva*, measured and a tissue biopsy (fin-clip) taken and stored in 95% ethanol.

Cytochrome b gene sequencing and analysis

DNA was extracted from the *P. parva* fin-clips using DNeasy extraction Kit (Qiagen). A total of 563 Cytochrome B gene sequences were amplified using L15267 and H15891Ph primers (Briolay et al. 1998). All reactions were carried out in a final volume of 10 μL. PCR cycling parameters were as follow: 95 °C for 15 min followed by 35 cycles of 95 °C for 30 s, 60 °C for 1.30 min, 72 °C for 1 and 15 min at 70 °C for the final elongation time. Exo-Sap purification (USB Corp.) was performed with the following incubation: 37 °C for 30 min and 80 °C for 20 min. The cycle sequencing reaction parameters were 96 °C for 1 min followed by 29 cycles of 96 °C for 10 s, 55 °C for 15 s and 60 °C for 4 min using an ABI Applied Biosystem, 3730 DNA Analyzer. The sequences generated were visualized and analysed using CodonCodeAligner (CodonCode Corp.). All the sequences were submitted to Genbank: accession number KR074432 to KR074994.

A phylogenetic tree was generated using MrBayes (Ronquist et al. 2012) using the sequences generated in the present study as well as from Simon et al. (2011) and Liu et al. (2010). *Gnathopogon elongatus* (AB218687) and *Gobiocypris rarus* (AF309083) were used as outgroups. Substitution model HKY + I + G was chosen using jModeltest (Darriba et al. 2012). The generation number was set at 1000,000 with 250,000 of burnin. The tree was visualized using Figtree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). The numbers of haplotypes and haplotype diversity were calculated using DNAsp (Librado and Rozas 2009). Fst, Tajima's as well as Fu's Fs were calculated using Arlequin ver. 3.5.2.2 (Excoffier and Lischer 2010).

Haplogroup dating

Sequences representing two other *Pseudorasbora* species and other Gobioninae lineages were retrieved from Genbank, as well as genera of Xenocyprinae, Tincinae and Leuscininae used as outgroups. The non *Pseudorasbora* genera were sampled to represent Gobioninae lineages and to provide calibration intervals (see below). The final alignment comprised of 18 sequences and 597 positions. Accession numbers of

Table 1 Population genetic parameters for the mitochondrial haplotype and the microsatellite data

Country	Code	Coordinates		Mitochondria			Microsatellites			
		X	Y	N	number of haplotypes	haplotype diversity	N	Hexp	Hobs	Average number of allele per locus
China	S1	115.56	37.55	21	7	0.833	23	0.71+/- 0.27	0.70+/-0.28	13.11
	S2	117.12	34.81	21	8	0.805	23	0.73+/-0.22	0.69+/-0.22	14.78
	S3	118.59	33.19	17	13	0.963	21	0.77+/-0.21	0.76+/-0.19	14.11
	S4	118.57	31.40	17	2	0.118	23	0.77+/-0.24	0.72+/-0.23	14.89
	S5	118.97	30.63	17	6	0.706	19	0.77+/-0.22	0.71+/-0.22	12.33
	S6	119.57	28.12	21	4	0.738	21	0.79+/-0.18	0.75+/-0.20	13.56
	S7	110.32	25.27	22	4	0.26	22	0.64+/-0.27	0.58+/-0.30	8.78
	S9	113.11	29.15	19	4	0.45	22	0.71+/-0.26	0.64+/-0.28	13.33
	S10	111.55	32.56	23	10	0.862	23	0.76+/- 0.14	0.70+/-0.14	10.11
	S11	110.99	34.62	22	11	0.896	23	0.80+/-0.17	0.77+/-0.15	15.67
	S13	122.52	40.10	23	7	0.573	23	0.63+/-0.22	0.60+/-0.23	9.56
	S14	124.99	45.03	17	7	0.662	22	0.59+/-0.30	0.57+/-0.29	11.67
	S15	122.93	42.64	20	7	0.768	23	0.55+/-0.34	0.48+/-0.32	9.89
	S16	118.27	40.90	20	5	0.726	23	0.69+/-0.27	0.69+/-0.28	10.33
	S17	116.89	43.30	16	7	0.842	23	0.47+/-0.29	0.48+/-0.35	8.67
	S19	99.94	23.56	23	1	0	23	0.73+/- 0.23	0.65+/-0.29	10.11
	S20	99.53	23.35	14	2	0.143	22	0.63+/-0.29	0.54+/-0.30	7.44
	CG	113.27	23.13	6	1	0				
	CH	116.61	40.31	15	3	0.448				
	CK	118.36	24.44	6	1	0				
	CRH	117.22	39.12	15	11	0.952				
	CY	113.89	29.97	10	2	0.467				
Japan	JP	139.43	35.67	20	7	0.679	23	0.83+/-0.11	0.76+/-0.15	14.22
	JB	136.07	35.25	3	2	0.667				
Taiwan	TW	121.75	24.81	22	2	0.173	23	0.40+/-0.28	0.40+/-0.31	4.44
	TI	121.75	24.75	15	3	0.59				
	TT	120.64	23.15	15	1	0				
Armenia	A	45.54	40.21	15	1	0				
Iran	IR	54.78	37.05	6	2	0.533	23	0.58+/- 0.12	0.54+/-0.24	4.00
Turkey	TR	30.04	40.91	8	1	0	23	0.70+/-0.08	0.63+/-0.12	5.44
Austria	AT	14.72	48.19	10	3	0.689	21	0.70+/- 0.19	0.64+/-0.22	7.33
Belgium	BE	4.80	50.94	18	3	0.569	20	0.61+/-0.24	0.62+/-0.25	6.33
	BS	5.25	50.93	16	5	0.683				
Bulgaria	BG1	26.85	44.06	10	1	0	10	0.64+/- 0.15	0.59+/-0.18	4.89
	BG2	26.70	43.99	10	1	0	10	0.56+/-0.17	0.53+/-0.21	4.56
Czech Republic	T	14.18	49.15	15	4	0.733				
France	FG	-1.73	47.10	15	4	0.752				
Germany	G	11.15	47.87	16	7	0.867				
Great Britain	GB	-1.45	51.00	17	2	0.309	22	0.63+/-0.25	0.58+/-0.26	6.11
	EB	-1.16	54.20	15	1	0				
	SWS	-4.16	51.68	15	2	0.248				
Hungary	HU	18.87	46.63	22	4	0.403	23	0.59+/-0.25	0.52+/-0.23	6.44
	HA	21.10	47.60	15	5	0.629				
	HE	17.39	46.80	15	3	0.59				
	HG	17.75	47.43	15	5	0.629				
	HS	20.67	47.94	15	3	0.257				
Italy	IN	12.24	43.35	16	3	0.342				
	IT	10.52	44.77	21	2	0.467	23	0.66+/-0.19	0.53+/-0.20	6.33
Poland	PL	17.19	51.19	22	1	0	22	0.64+/-0.27	0.63+/-0.28	6.67
	PU	18.16	50.60	14	3	0.538				
Slovakia	SC	18.82	47.82	15	3	0.676				
Slovenia	SL	15.33	46.26	20	1	0	21	0.57+/- 0.25	0.43+/-0.28	4.67
Spain	SE	0.86	40.72	15	3	0.676				
	ES	2.53	41.57	23	1	0	23	0.34+/-0.29	0.31+/-0.28	2.33

The abbreviation used in the current article for all population can be found in column 2. In grey are the populations published in Simon et al. (2011)

N number of individuals analysed, *Hexp* expected heterozygosity, *Hobs* observed heterzygosity

the sequences included in the phylogenetic analysis were: *Pseudorasbora* (*P. pumila*: AB239599; *P. elongata*: KF05193), other Gobioninae (*Gnathopogon elongates*: AB218687; *Gobiocypris rarus*: AF309083; *Coreius guichenoti* AY953001; *Coreius heterodon* AY953000; *Gobio gobio*: AY953007; *Pseudogobio esocinus*: AP009310; *Hemibarbus barbus*: NC_008644; *Hemibarbus mylodon*: AP011414), and outgroups (*Tinca tinca*: AB218686; *Alburnus alburnus*: AB239593; *Opsariichthys uncirostris*: AB218897; *Aphyocypris chinensis*: AB218688). The best model fitting our data was investigated with jModeltest v2.1.7 (Darriba et al. 2012) using the Akaike criterion (AIC; Akaike 1973).

Divergence dates were estimated using the relaxed molecular clock approach implemented in BEAST v2.4.3 (Bouckaert et al. 2014). MCMC analyses were run for 20 million generations with tree sampled every 1000 generations with the model selected by jModeltest (TN + I + G). Convergence was checked and burn-in was estimated with Tracer v1.6 (Rambaut et al. 2014). The trees were combined into a maximum clade credibility tree using TreeAnnotator using a burn-in of 10%, with the estimation of the mean node height and highest posterior density intervals fixed at 95%. As no reliable calibration interval was available for Gobioninae, we used secondary calibration divergence times (mean and the 95% confidence interval) obtained in Gobioninae and closely related taxa by Saitoh et al. (2011). We retained 4 calibration intervals with lognormal prior: Leuciscinae (*Alburnus alburnus*)/*Tinca:* 77.9 Mya (95%: 64.1–93.1), *Gobio/Pseudogobio:* 61.4 Mya (95%: 49.1–75.3), Gobioninae/(Leuciscinae + *Tinca*): 83.5 Mya (95%: 69.4–99.4) and (*Gobioninae/Leuciscinae* + *Tinca*)/*Xenocyprininae*: 93.6 (95%: 78.7–110.1).

Microsatellites genotyping and analysis

Microsatellites from Konishi and Takata (2004) and Zeng et al. (2011) were tested and 666 *P. parva* samples were successfully genotyped for 9 polymorphic microsatellites (MS07, PA2, PA6, PA14, PA15, PA18_1, PA18_2, PA20 and PA22). The 5′ ends of the forward primers were labelled with either HEX or FAM dyes. All reactions were carried out in a 5 μl volumes using DNA template at 10 ng/μl. Primers pairs were pooled and reactions were performed using the Qiagen multiplex PCR kit (Qiagen). PCR

conditions were as follows: 95 °C for 15 min followed by 28 cycles at 95 °C for 30 s, 50–60 °C (depending on the pool) for 1.30 min, 72 °C for 1.30 min with a final extension at 72 °C for 10 min. PCR products were diluted 1:20 in water. 1 μL of this dilution was added to 10 μL of HiDi formamide and 0.1 μL of 500 ROX size standard (Applied Biosystems) using an ABI Applied Biosystem, 3730 DNA Analyzer. Amplicons were denatured by heating at 90 °C for 2 min followed by cooling at 20 °C for 5 min. Alleles were scored using GeneMapper (Applied Bioscience) by two independent scorers. Allelic richness, as a standardized measure of the number of alleles per locus independent of the sample size was calculated using Fstat (Goudet 1995). The allelic richness between populations was calculated with 1000 permutations using Fstat (Goudet 1995). Calculation of the observed and expected heterozygosity as well as the average number of allele per locus was obtained using GENETIX (Belkhir et al. 2004) for each population.

To determine the genetic structure of the tested populations we used a combination of methods. Firstly, discriminant analysis of principal components (DAPC) (Jombart et al. 2010), was used to assess the population structure using the complete dataset (i.e. native and invasive populations were combined). This was performed in R using the ADEGENET package (Jombart 2008). This multivariate analysis derived the probability of individual membership in each different group. The program covered a range of possible clusters from 1 to 31, 31 representing the total number of populations in the dataset. The interaction number used was 1,000,000,000. 50 principal components were retained as predictors for discriminant analysis in the 666 individuals studied.

In addition, we performed a structure analysis on the dataset. This analysis was performed separately for the Asian and European *P. parva* populations using STRUCTURE software (Pritchard et al. 2000), with a burn-in period of 500,000 simulations followed by a run length of 1000,000 MCMC simulations and ten iterations for each K (number of clusters) with the admixture model. To infer population structure on a finer scale, the analysis was performed from K = 1–18 for the 18 Asian populations and from K = 1–11 for the 11 European populations. The results were analysed using STRUCTURE HARVESTER (Earl and vonHoldt 2012) and the K was chosen using a combination of Evanno's criteria

(Evanno et al. 2005) and investigating the convergence of the runs using CLUMPP (Jakobsson and Rosenberg 2007). Results were summarized using CLUMPP (Jakobsson and Rosenberg 2007) and displayed using Distruct (Rosenberg 2004). We tested the scenario that *P. parva* populations geographically closer to one another would also be more genetically similar by generating matrices with Fst values using Arlequin (Excoffier and Lischer 2010), and correlating these to geographic distances between these populations. The two matrices were correlated using the Mantel test (R software, Vegan).

Results

Mitochondrial DNA

A total of 874 analysed sequences (597-bp) of the Cytochrome B gene revealed four distinct lineages and 102 haplotypes across the samples (Figure 2—Supplementary material 1). As expected more synonymous mutations (91) than non-synonymous mutations (22) were found in our dataset. In its native range of China and Japan, *P. parva* displayed three distinct lineages: (Lineage 1, 2 and 4; Figs. 1, 2). Lineage 1 and 2 were found across China, with a north–south distribution in relation to the Yangtze River (Fig. 1). Lineage 1 was present in relatively high frequency in the North (93% Lineage 1, 7% Lineage 2), whereas lineage 2 was most prevalent in the South (4% Lineage 1, 96% Lineage 2). Average haplotype diversity was higher in the northern China (i.e. north of the Yangtze River S1, S2, S3, S10, S11, S13, S14, S15, S16 and S17—0.78) compared to the south (i.e. south of the Yangtze River: S4, S5, S6, S7 and S9—0.45—Table 1). Haplotype diversity in the invasive range of China (S19 and S20) was low (0.0715—Table 1).

In Japan, the most prevalent lineage was Lineage 4 (Figs. 1, 2). This lineage had 4.7% divergence with the main lineages (1 and 2) with 28 mutations difference to the closest sequences. Lineage 2 was present in Japan but at a lower prevalence (Lineage 2: 30% and Lineage 7: 70%; Fig. 1). Lineage 3 (green) was only been found in Iran and diverged by 5.2% from the other lineages (31 mutations difference with the closest sequence).

In the invasive populations of Europe, only lineages 1 and 2 were identified (Lineage 1: 69% and Lineage

Fig. 1 Summary of mitochondrial haplogroup and the population structure derived from microsatellites across *Pseudorasbora parva*'s geographical distribution: **a** native range, China, Taiwan and Japan and **b** Eurasia. Abbreviations of locations can be found in Table 1. Lineage 1 and 2 are represented in pink and blue respectively

2: 31%). All individuals sampled from Bulgaria, Turkey and Armenia shared the same haplotype. Overall, the average haplotype diversity was lower in Europe (0.4) compared to China (0.68). The European individuals were present throughout the phylogeographical tree (Fig. 2), suggesting several introductions had occurred. Pairwise Fst calculations between Europe, China, Taiwan, Iran and Japan revealed that China and Europe were genetically similar (Fst = 0.09), indicating a direct relationship between those populations (Table 2).

Historical demography and haplogroup dating of native *Pseudorasbora parva*

The historical demography of China, Europe, Iran, Japan and Taiwan was investigated using Tajima's D and Fu's Fs values. The values were positive and not significant for all populations except for China. Tajima's D is -1.57 ($p = 0.02$) and Fu's Fs is -24.47 ($p < 0.01$) for China, indicating that the population is expanding. Regarding haplogroup dating, the chronogram obtained with BEAST revealed the two main lineages, 1 and 2, separated in China 2.5 Mya [0.9, 4.5] with 95% confidence intervals (Fig. 3). Lineages 3 (Iran) and 4 (Japan) separated 3.9 Mya [1.2, 6.7] and diverged from lineages 1 and 2 12.1 Mya [6.5, 18.1] (Fig. 3).

Microsatellite data

A total of 666 individual *P. parva* were successfully genotyped at 9 microsatellite loci. The allelic richness between Europe and China was calculated to be 2.46 and 2.81 respectively, with this difference significant ($p = 0.02$). The allelic richness of the population located to the north of the Yantze River (2.75) was compared to the populations located to the south of the Yantze (2.91) was not significant ($p = 0.39$). Compared with the allelic richness of populations to the north and south of the Yantze river, the European populations were significantly different from the
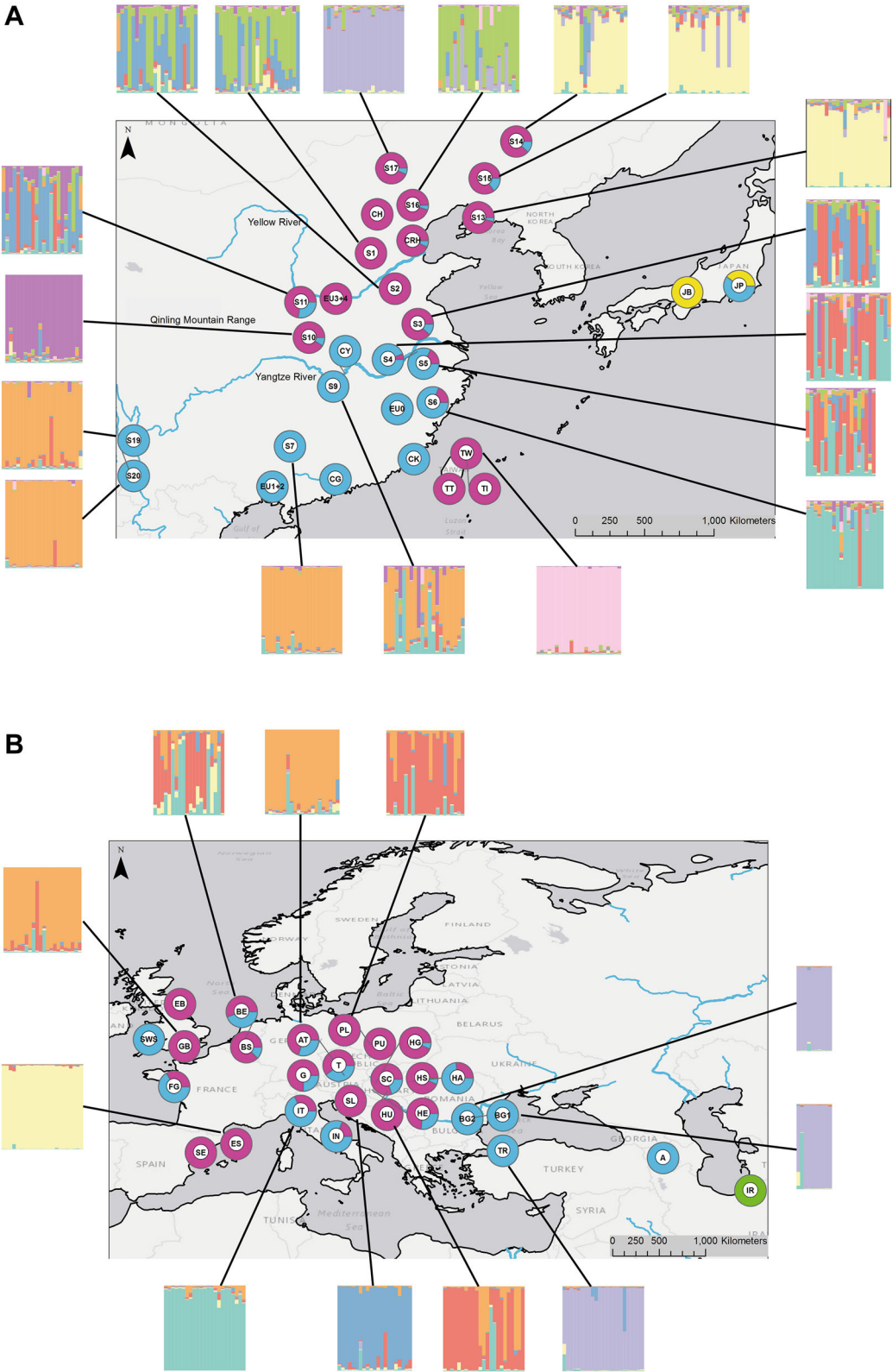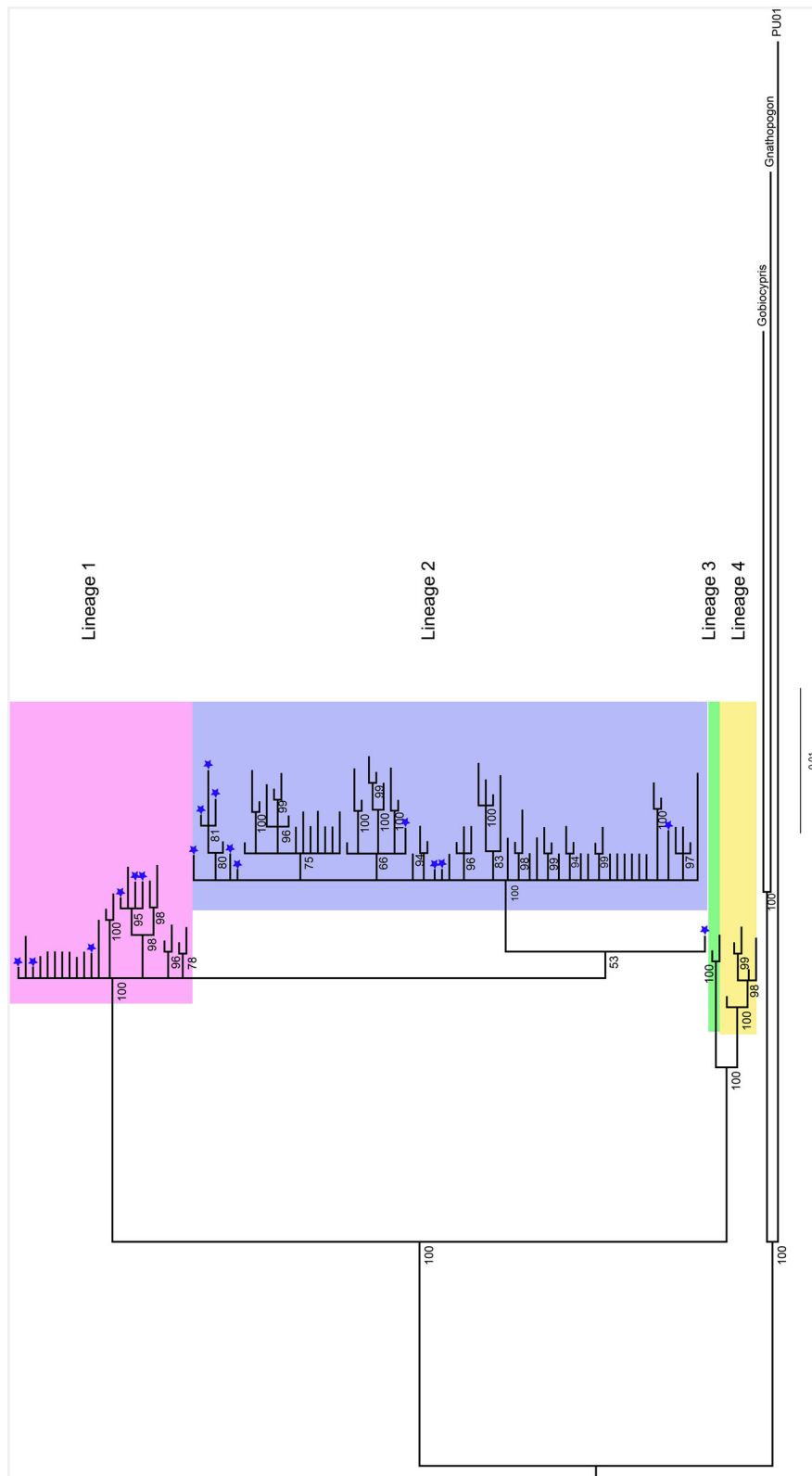
Fig. 2 Bayesian phylogeny reconstructed with cytochrome b dataset. The support, posterior probabilities determined with MrBayes, is indicated for each node. Only the nodes with a support above 50 were retained. Blue stars indicate European samples

southern populations ($p = 0.01$) but not from those to the north ($p = 0.09$). Chinese populations showed a higher genetic diversity, with a Hobs average with and without the two invasive Chinese populations (S19 and S20) of 0.66 and 0.65 respectively. The average number of allele per locus with and without S19 and S20 was 12.05 and 11.67 respectively. European populations had a lower genetic diversity with a mean Hobs average of 0.55 and the average number of allele per locus of 5.56 (Table 1). Populations S19 and S20 are invasive *P. parva* populations in China and they have a genetic diversity that is lower than native

Table 2 Pairwise Fst obtained for the mtDNA sequences

	Europe	China	Iran	Taiwan	Japan
Europe	–				
China	**0.09**	–			
Iran	**0.86**	**0.83**	–		
Taiwan	**0.44**	**0.41**	**0.91**	–	
Japan	**0.60**	**0.52**	**0.48**	**0.62**	–

Numbers in bold indicate significant pairwise comparisons

Chinese populations but higher than European populations (average Hobs = 0.60, average number of allele per locus 8.78). The Fst values for microsatellites were calculated using Arlequin (Excoffier and Lischer 2010) and can be seen in Table 3. Interestingly China and Europe have a very low Fst value (0.04–see Table 3) which expected as *P. parva* populations from China was introduced in Europe. Iran has very high Fst value in agreement with the mitochondrial data. The concordance between the mitochondrial and microsatellites makers was found to be positive with a R^2 coefficient of 0.69.

In the DAPC analysis that assigned the individuals to their genetic population, the lowest BIC value (851.07) was found for K = 13. The first 50 PCs of the PCA, explaining 71.2% of the total variance, were kept. Out of the first six axes, several clusters were found, with Iran showing the highest divergence from the other populations (Fig. 4a). Notably, Spain's genetic distinctness compared to other invasive populations is explained by Axis 3. This result might be due to the low genetic diversity of the Spanish population (Table 1). The two populations from Bulgaria and Turkey (Fig. 4b) clustered together in

Fig. 3 Chronogram based on BEAST analysis of the cytochrome b dataset. The four calibration points are indicated by a star (see text for details); for each node, the mean and the 95% credibility intervals are indicated
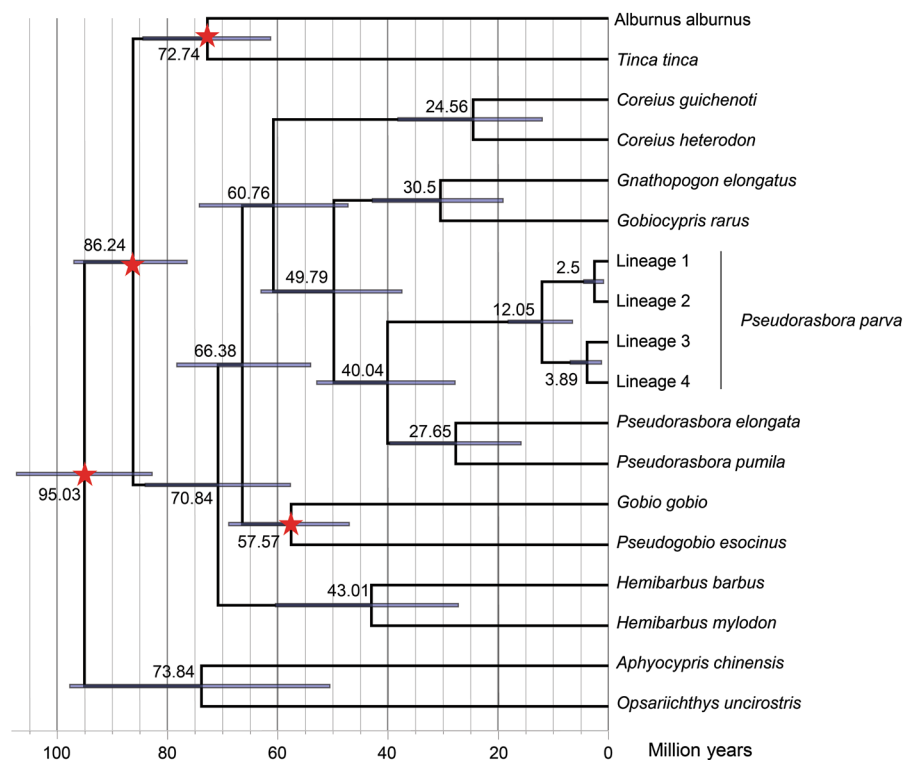
Axis 4 and, finally, Taiwan separated from the other populations on axis 5 (Fig. 4c).

A STRUCTURE analysis was performed, using the Evanno et al. (2005) method and investigating the convergence of the runs using CLUMPP (Jakobsson and Rosenberg 2007). Using the pairwise G′ parameter from CLUMPP (Jakobsson and Rosenberg 2007), we found that only 2 K converged (K = 9 and K = 10) in the Chinese population analysis. K = 9 was selected as the best model for the ancestral populations in China as it had the highest ΔK (Fig. 5). The clustering of the different genetic populations was in agreement with the geographical distribution of the sampled populations (Fig. 1a). The Taiwan populations formed an independent genetic population. S7, S9, S19 and S20 (South-West China) clustered together, as per the DAPC results (Fig. 3). The populations from North-East China, S13, S14 and S15 formed another cluster. Populations from western China, S3, S4 and S5, clustered together, and populations S1 and S16 formed the last cluster. Populations S6, S10, S11 and S17

represented 4 different genetic populations. In Europe, the Evanno method and the study of the convergence using CLUMPP revealed K = 6 (Fig. 5b), where independent clusters were (1) Bulgaria and Turkey, (2) Austria and UK, (3) Hungary, (4) Belgium and Poland, (5) Italy, Spain and final (6) Slovenia. Interestingly K = 2 was also found to be convergent with two clustered being formed (1) Bulgaria and Turkey and (2) the rest of the European populations. Bulgaria and Turkey have the same mitochondrial haplotype and this result suggests a single introduction of *P. parva* in those two countries. The Mantel test indicated that populations that were geographically furthest apart (geographically) were also more genetically distant (Mantel statistic r = 0.5472; p = 0.002).

Discussion

These genetic results for native and invasive *P. parva* populations confirm the importance of extensive sampling in both ranges for invasive species in evaluating the influence of genetic variability on invasion success. Analyses revealed that *P. parva* had higher genetic variability in the native range than the invasive range, contrasting with the previous findings of Simon et al. (2011, 2015). The successful invasion of 32 countries by *P. parva* has been previously attributed to admixture of different source populations in its invasive range (Simon et al. 2011, 2015). However, our data demonstrated that admixture had already occurred in the species' native range, through
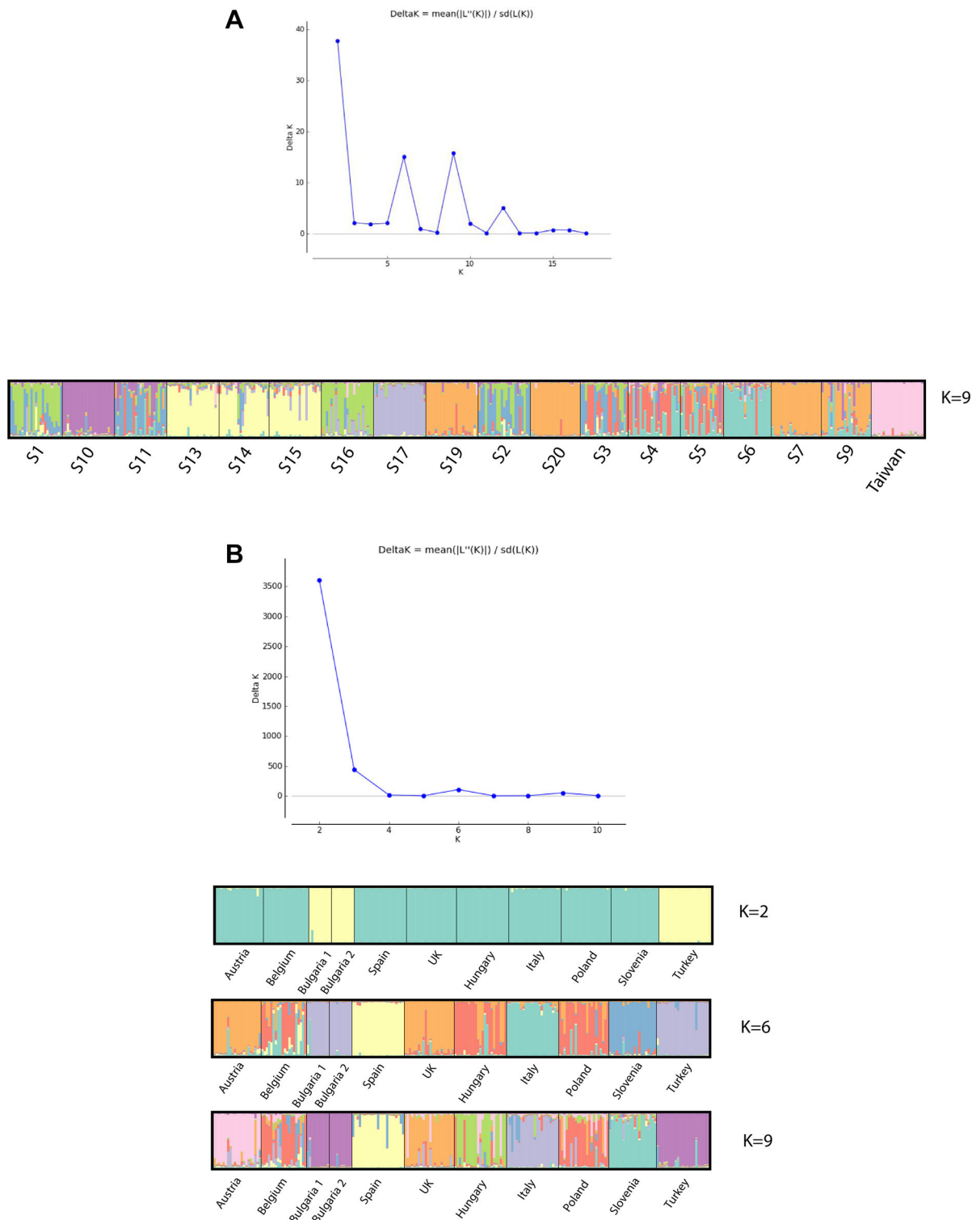
Table 3 Fst values obtained using 9 microsatellites

	Europe	China	Iran	Taiwan	Japan
Europe	–				
China	**0.04**	–			
Iran	**0.30**	**0.22**	–		
Taiwan	**0.19**	**0.12**	**0.34**	–	
Japan	**0.11**	**0.06**	**0.18**	**0.18**	–

Numbers in bold indicate significant pairwise comparisons
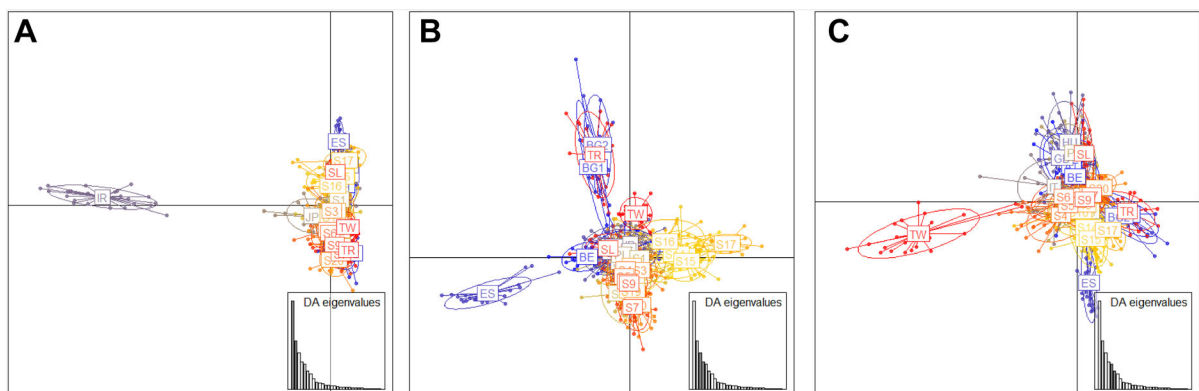
Fig. 4 Representation of the microsatellite variation on the first six axes of a DAPC. The percentage of the total eigenvalue explained by: **a** Axis 1 is 34.07% and Axis 2 is 17.36%; **b** Axis 3 is 12.92% and Axis 4 is 9.27%; **c** Axis 5 is 7.81 and Axis 6 is 5.09%

Fig. 5 Structure results for *Pseudorasbora parva* populations obtained using 9 microsatellites. **a** Structure of the fish population in China; **b** structure of the fish population in Europe

its movement along with aquaculture-valuable species (such as Chinese carps), resulting in highly variable, admixed populations in Chinese reservoirs and aquaculture facilities that were likely to be the source populations for the European introductions.

Genetic variability across its invasive range

The invasive range of *P. parva* not only includes Europe, but also regions within China (i.e. populations S19 and S20). Our results indicated that the *P. parva* introduction to Europe originated from China, not Japan, providing strong support that introductions were ultimately associated with commercial links between China and Eastern European countries in the mid to late Twentieth Century (Britton and Gozlan 2013). The first introduction of *P. parva* has been attributed to Romania (Nucet Fisheries Research Centre, in 1961), with imports of eggs of silver carp *Hypophthalmichthys molitrix* and grass carp *Ctenopharyngodon idella* from Wuhan, China (Gozlan 2011). However, similar cooperation programmes took place at the same time within countries of the Eastern Bloc and resulted in further introductions of *P. parva* to Hungary (Paks Fisheries Farm, in 1963), Lithuania (Dunojus Lake, in 1963) and the Ukraine (Kuchurganskoye reservoir, in 1962) (see Gozlan 2011 for full details).

There was lower genetic diversity in the European and Chinese *P. parva* invasive ranges compared to the native range, as would be expected due to population bottleneck events during the introduction and colonisation process (example: Schimd-Haempel et al. 2007; Hanfling 2007; Hardouin et al. 2010). Introductions from China were also identified in Taiwan and Japan. Japan has endemic lineages of *P. parva* (Lineage 4) and more recent introductions from China (Lineage 2). The genetic analyses of the Iranian *P. parva* indicated that their source was genetically different to all the sampled populations in the Chinese native range. This population had a single mtDNA lineage (Lineage 3) that was highly divergent, with 5.2% difference in comparison to the lineage found in China (31 mutations difference with the closest sequence). The source population is thus divergent to the ones sampled in the Chinese native range but more closely related with the endemic population of *P. parva* in Japan (Lineage 4). These two lineages have been separated 3.9 Mya, thus determining the potential population source of this introduction would require a more intensive sampling of the native range of *P. parva*, including the Korean peninsula, which is currently missing.

Phylogeography in the native range: China, Taiwan and Japan

Understanding the phylogeography and the factors affecting the genetic structuring of invasive species in their native range reveal historic and recent selection events that could have influenced invasion success (Zenni et al. 2014). Recent evidence indicates that genotypes that perform well in their introduced ranges can result from both past evolutionary events in their native range or rapid adaptation following introduction to new environments (Zenni et al. 2014). Here, the Qinling Mountains have played an important role in influencing the phylogeography of *P. parva*. These mountains represent a natural boundary between the North and the South of the country, dividing the Chinese temperate and sub-tropical climatic zones (Ding et al. 2013), resulting in differentiated terrestrial and freshwater fauna (Li 1981; Zhang 1999). The split between lineages 1 and 2, estimated around 2.5 Mya, might have been a consequence of the fast uplift (500–1000 m) of these mountains that was influenced by the Qinghai-Xizang Movement, initiated approximately 3.4 Mya (Teng and Wang 1996; Zhang and Fang 2012).

This distribution of the *P. parva* has also been influenced by past human activities that resulted in gene flow between populations to the north and south of the Qinling Mountains. For example, the construction of the Lingqu Canal 2200 years ago (Zhong 1990) connected the Yangtze River and the Xijiang (Peal River basin), leading to a similar *P. parva* population structure in these areas (S7 and S9). The Beijing-Hangzhou Grand Canal (built from AD 581–618), flows across several river systems including the Yangtze, Huahe, Yellow and Haihe Rivers in the East of China and could explain the presence of the southern lineage 2 in some northern populations (S16, S17, CRH, and S3). China's long history of carp aquaculture and aquaculture trading is also likely to have been an underlying process in formation of similar genetic structures observed in some other populations (e.g. S4, S5, S6) (Yue and Liang 1995; Zhao et al. 2015).

To the north of the Yangtze River, two population clusters were detected, corresponding to the Yellow River (S1, S2, S16) and the Amur River Basin (S13, S14, S15), which includes populations from inner Mongolia (S17). Again, although there was clear genetic structuring of these populations, they mostly formed part of the same northern Lineage 1, indicating recent colonisation of these catchments, probably corresponding to the development of pond aquaculture in China that can be traced back approximately 3300 years (Yue and Liang 1995) and which significantly increased during the Han dynasty (Shi 1999).

Finally, the specific population structure indentified in the two highly admixed populations (S10, S11), which displayed high genetic diversity, was likely to have resulted from the construction of many small and mid-sized dams during the late 1950s–1960s (Gong and Tu 1991) which were restocked (e.g. lakes and reservoirs). In effect, the presence of numerous large reservoirs in this region, such as Danjiangkou, Hongshuihe, Liuyanji, Tankou, Wujiaying and Xipaizihe, made it a hotspot for fish farming in the late 1960's. This contributed to the highly admixed populations, in a period that also corresponds to the first introductions of *P. parva* in Europe.

Conclusion

Our study highlights the importance of understanding both old and new processes shaping the genetic structure in the native range of invasive species. These can reveal processes, which could have enhanced the invasive potential of the species allowing them to quickly respond to new challenges (small founding populations, exposure to new parasitic communities, etc.) and therefore resolve the potential invasive paradox. Thus, future studies on invasive species should employ holistic approaches encompassing the role of both genetics and plasticity in biological invasions, with extensive sampling and studies across all ranges.

Authors' contributions EAH, JRB and REG, conceived the ideas; YZ, JRB and REG conducted the fieldwork. EAH and DA collected the data. EAH, PC and DHF analysed the data; and EAH and DA wrote the manuscript with assistance from YZ, DHF, PC, JRB and REG

References

Akaike H (1973) Maximum likelyhood identification of gaussian autoregressive moving average models. Biometrika 60:255–265

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier

Bell G, Gonzalez A (2011) Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. Science 332:1327–1330

Bermond G, Ciosi M, Lombaert E et al (2012) Secondary contact and admixture between independently invading populations of the western corn rootworm, *Diabrotica virgifera virgifera* in Europe. Plos ONE 7:e50129

Besnard G, Dupuy J, Larter M et al (2014) History of the invasive African olive tree in Australia and Hawaii: evidence for sequential bottlenecks and hybridization with the Mediterranean olive. Evol Appl 7:195–211

Biebach I, Keller LF (2012) Genetic variation depends more on admixture than number of founders in reintroduced Alpine ibex populations. Biol Conserv 147:197–203

Blanchet S (2012) The use of molecular tools in invasion biology: an emphasis on freshwater ecosystems. Fish Manag Ecol 19:120–132

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Comput Biol 10(4):e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Briolay J, Galtier N, Brito RM, Bouvet Y (1998) Molecular phylogeny of cyprinidae inferred from cytochrome bDNA sequences. Mol Phylogenet Evol 9:100–108

Britton JR, Gozlan RE (2013) Geo-politics and freshwater fish introductions: How the cold war shaped Europe's fish allodiversity. Glob Environ Change Hum Policy Dimens 23:1566–1574

Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. Biol Lett 4:585–589

Chapple DG, Miller KA, Kraus F, Thompson MB (2013) Divergent introduction histories among invasive populations of the delicate skink (*Lampropholis delicata*): Has the importance of genetic admixture in the success of biological invasions been overemphasized? Divers Distrib 19:134–146

Copp GH, Vilizzi L, Gozlan RE (2010) Fish movements: the introduction pathway for topmouth gudgeon *Pseudorasbora parva* and other non-native fishes in the UK. Aquat Conserv Mar Freshw Ecosyst 20:269–273

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

Davies GD, Gozlan RE, Britton JR (2013) Can accidental introductions of non-native species be prevented by fish stocking audits? Aquat Conserv Mar Freshw Ecosyst 23:366–373

DeFaveri J, Shikano T, Merilä J (2014) Geographic variation in age structure and longevity in the nine-spined stickleback (*Pungitius pungitius*). PLoS ONE 9(7):e102660. https://doi.org/10.1371/journal.pone.0102660

Diez-del-Molino D, Carmona-Catot G, Araguas R-M et al (2013) Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). PLoS ONE 8:e82501.

Ding Y, Wang S, Zheng J, Wang H, Yang X (2013) China climate. Science Press, Beijing, pp 392–418

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Facon B, Hufbauer RA, Tayeh A et al (2011) Inbreeding depression is purged in the invasive insect harmonia axyridis. Curr Biol 21:424–427

Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W (2012) Invasive species distribution models—How violating the equilibrium assumption can create new insights. Glob Ecol Biogeogr 21:1126–1136

Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Funct Ecol 21:394–407

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

Goudet J (1995) Fstat version 1.2: a computer program to calculate F-statistics. J Hered 86(6):485–486

Gozlan RE (2011) *Pseudorasbora parva* Temminck & Schelegel (topmouth gudgeon). In: Francis R (ed) A handbook of global freshwater invasive species. Earthscan, London, pp 275–285

Gozlan RE, Andreou D, Asaeda T et al (2010a) Pan-continental invasion of *Pseudorasbora parva*: towards a better understanding of freshwater fish invasions. Fish Fish 11:315–340

Gozlan RE, Britton JR, Cowx I, Copp GH (2010b) Current knowledge on non-native freshwater fish introductions. J Fish Biol 76:751–786

Hanfling B (2007) Understanding the establishment success of non-indigenous fishes: lessons from population genetics. J Fish Biol 71:115–135

Hardouin EA, Chapuis J-L, Stevens MI et al (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. BMC Evol Biol 10:325

Hardouin EA, Orth A, Teschke M, Tautz D, Bonhomme F (2015) Worldwide mouse differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot. BMC Evol Biol 15:26

Hedge LH, Leung B, O'Connor WA, Johnston EL (2014) The interacting effects of diversity and propagule pressure on early colonization and population size. J Anim Ecol 83:168–175

Hufbauer RA, Rutschmann A, Serrate B, de Conchard HV, Facon B (2013) Role of propagule pressure in colonization success: disentangling the relative importance of demographic, genetic and habitat effects. J Evol Biol 26:1691–1699

Jakobsson M, Rosenberg NA (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. https://doi.org/10.1093/bioinformatics/btm233

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94 **(PubMed: 20950446 Bioinformatics 23:1801–1806)**

Keller SR, Fields PD, Berardi AE, Taylor DR (2014) Recent admixture generates heterozygosity-fitness correlations during the range expansion of an invading species. J Evol Biol 27:616–627

Konishi M, Takata K (2004) Isolation and characterization of polymorphic microsatellite DNA markers in topmouth gudgeon, *Pseudorasbora* (Teleostei: Cyprinidae). Mol Ecol Notes 4:64–66

Kumschick S, Gaertner M, Vila M et al (2015) Ecological impacts of alien species: quantification, scope, caveats, and recommendations. Bioscience 65:55–63

Li SZ (1981) Studies on zoogeographical divisions for fresh water fishes of China. Science Press, Beijing

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Liu HZ, Yang JQ, Tang QY (2010) Estimated evolutionary tempo of East Asian gobionid fishes (Teleostei Cyprinidae) from mitochondrial DNA sequence data. Chin Sci Bull 55:1501–1510

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rambaut A, Suchard M, Drummond A (2014) Tracer v1.6. http://tree.bio.ed.ac.uk/software/tracer/

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. Trends Ecol Evol 22:454–464

Ronquist F, Teslenko M, Mark PVD, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rosenthal DM, Ramakrishnan AP, Cruzan MB (2008) Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. Mol Ecol 17:4657–4669

Saitoh K, Sado T, Doosey MH, Bart HL Jr, Inoue JG, Nishida M, Mayden RL, Miya MM (2011) Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). Zool J Linn Soc 161:633–662

Sanz N, Araguas RM, Vidal O et al (2013) Genetic characterization of the invasive mosquitofish (*Gambusia* spp.) introduced to Europe: population structure and colonization routes. Biol Invasions 15:2333–2346

Schimd-Haempel P, Schimd-Haempel R, Brunner PC, Seeman OD, Allen GR (2007) Invasion success of the bumbleble, *Bombus terrestris*, despite a drastic genetic bottleneck. Heredity 99:414–422

Schulte U, Veith M, Mingo V, Modica C, Hochkirch A (2013) Strong genetic differentiation due to multiple founder events during a recent range expansion of an introduced wall lizard population. Biol Invasions 15:2639–2649

Searle JB (2008) The genetics of mammalian invasions: a review. Wildl Res 35:185–192

Shi D (1999) A glorious history of Chinese fishery. J Beijing Fish. 9:39–40

Simon A, Britton R, Gozlan R et al (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. https://doi.org/10.1371/journal.pone.0018560

Simon A, Gozlan RE, Britton JR, van Oosterhout C, Haenfling B (2015) Human induced stepping-stone colonisation of an admixed founder population: the spread of topmouth gudgeon (*Pseudorasbora parva*) in Europe. Aquat Sci 77:17–25

Smith JM, Wells SP, Mather ME, Muth RM (2014) Fish biodiversity sampling in stream ecosystems: a process for evaluating the appropriate types and amount of gear. Aquat Conserv Mar Freshw Ecosyst 24:338–350

Teng Z, Wang X (1996) Studies of the tectonic uplift at the Cenozoic era and the regionally environmental effects in the Qinling orogenic belt. Geol Shaanxi 14:33–42

Vandepitte K, De Meyer T, Helsen K et al (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. Mol Ecol 23:2157–2164

Wiens KE, Crispo E, Chapman LJ (2014) Phenotypic plasticity is maintained despite geographical isolation in an African cichlid fish, *Pseudocrenilabrus* multicolor. Integr Zool 9:85–96

Yue P, Liang Z (1995) A sketch on the fishery history and its development in ancient China. Chin J Zool 30:54–58

Zeng C, Gul Y, Yang K, Cui L, Wang WM, Gao ZX (2011) Isolation and characterization of 19 polymorphic microsatellite loci from the topmouth gudgeon *Pseudorasbora parva*. Genet Mol Res 10:1696–1700

Zenni RD, Bailey JK, Simberloff D (2014) Rapid evolution and range expansion of an invasive plant are driven by provenance-environment interactions. Ecol Lett 17:727–735

Zhang R (1999) Zoogeography of China. Science Press, Beijing

Zhang L, Fang X (2012) Palaeogeography of China: formation of China's natural environment. Science Press, Beijing, pp 237–247

Zhao Y, Gozlan RE, Zhang C (2015) Current state of freshwater fisheries in China. In: Craig J (ed) Freshwater fisheries ecology. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9781118394380.ch19

Zhong F (1990) Preliminary study of geological geomorphological conditions of Ling Canal. Carsologica Sin 9:88–93