**High cryptic diversity of bitterling fish in the southern West Palearctic**

**Running title**: European bitterling in West Palearctic

Veronika Bartáková[1,2], Josef Bryja[1,2], Radek Šanda[3], Yusuf Bektas[4], Tihomir Stefanov[5], Lukáš Choleva[6,7], Carl Smith[1,8,9], Martin Reichard[1,*]

[1] *The Czech Academy of Sciences, Institute of Vertebrate Biology, Květná 8, 603 65 Brno, Czech Republic*

[2] *Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic*

[3] *National Museum, Department of Zoology, Václavské nám. 68, 115 79 Prague 1, Czech Republic*

[4] *Department of Basic Sciences, Faculty of Fisheries, Recep Tayyip Erdogan University, 53100, Rize, Turkey*

[5] *National Museum of Natural History, Bulgarian Academy of Sciences, 1 Tsar Osvoboditel Blv., 1000 Sofia, Bulgaria*

[6] *The Czech Academy of Sciences, Institute of Animal Physiology and Genetics, Laboratory of Fish Genetics, Rumburská 89, 277 21 Liběchov, Czech Republic*

[7] *Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, Ostrava 710 00, Czech Republic*

[8] *Department of Ecology & Vertebrate Zoology, University of Łódź, Banacha 12/16, 90-237 Łódź, Poland*

[9] *School of Biology and Bell-Pettigrew Museum of Natural History, University of St Andrews, St Andrews KY16 8LB, UK*

***Correspondence**: Martin Reichard, Institute of Vertebrate Biology, The Czech Academy of Sciences, Květná 8, 603 65 Brno, Czech Republic; E-mail: reichard@ivb.cz

**ABSTRACT**

South-east Europe, along with the adjacent region of south-west Asia, is an important biodiversity hotspot with high local endemism largely contributed by contemporary continental lineages that retreated to southern refugia during colder Quaternary periods. We investigated the genetic diversity of the European bitterling fish (*Rhodeus amarus*) species complex (Cyprinidae) across its range in the western Palearctic, but with a particular emphasis in the region of Balkan, Pontic and Caspian refugia. We genotyped 12 polymorphic microsatellite loci and a partial sequence of mitochondrial gene cytochrome *b* (*CYTB*) for a set of 1,038 individuals from 60 populations. We used mtDNA sequences to infer phylogenetic relationships and historical demography, and microsatellite markers to describe fine-scale genetic variability and structure. Our mtDNA analysis revealed six well-supported lineages, with limited local co-occurrence. Two lineages are distributed throughout central and western Europe (lineages "A" and "B"), with two zones of secondary contact. Another two lineages were restricted to the Ponto-Aegean region of Greece (lineages "C" and "D") and the final two lineages were restricted south of the Caucasus mountains (lineage "E" from the Black Sea watershed and lineage "F" from the Caspian watershed). A signal of recent expansion was revealed in the two widespread lineages and the Ponto-Aegean lineage "C". The geographic distribution of clusters detected by nuclear microsatellites corresponded well with mitochondrial lineages and demonstrated finely sub-structured populations. A profound population structure suggested a significant role of genetic drift in differentiation among lineages. Lineage divergence in the Ponto-Aegean and Caspian regions are substantial, supporting the validity of two described endemic species (*Rhodeus meridionalis* as lineage "D" and *Rhodeus colchicus* as lineage "E") and invite taxonomic evaluation of the other two southern lineages (Thracean "C" and Caspian "F").

**KEYWORDS**

**Balkan refugium; Cyprinidae; Mediterranean endemism; Ponto-Caspian region; gene flow; genetic drift**

## 1. Introduction

Contemporary species distributions and intraspecific diversity are largely driven by Pleistocene climatic oscillations, with climatic dynamics of the Holocene having major impacts on the species and intraspecific diversity of many Palearctic taxa (Hewitt, 1999). In cold periods, thermophilic species retreated to thermal refugia; in the west Palearctic they were primarily located in the Mediterranean peninsulas (Iberian, Apennine, Balkan) and Caspian-Caucasian region (Stewart et al., 2010). Some species retreated to smaller cryptic refugia at higher latitudes (Stewart and Lister, 2001; Stewart et al., 2010) with sheltered topography and suitable microclimates (Stewart and Lister, 2001). The contribution of disparate refugia to the current diversity of the European biota was uneven, with an overrepresentation of lineages expanding from the Balkan refugium where the geography favoured dispersal (Hewitt, 1999). The Ponto-Caspian refugia from the region along the Black Sea coast and on the north-western slopes of the Caucasus Mountains (Adams and Faure, 1997), also supported re-colonisation of Europe, but to a much lesser extent (Culling et al., 2006). These areas may have been relatively fragmented, supporting genetically diverse populations, termed 'refugia-within-refugia' (Gómez and Lunt, 2007), rather than representing a single continuous refugium with large and interconnected populations. A possible outcome is that these former refugial areas may represent hotspots of biodiversity as a consequence of two different proximate mechanisms (Dufresnes et al., 2016); long-term persistence of large refugial populations or, alternatively, a set of small, geographically isolated populations with limited gene flow. These two scenarios are predicted to bear unique genetic signatures. In the case of large continuous refugial populations with substantial effective population sizes over successive glacial cycles, genetic diversity should be high but homogenous across the refugium. In the case of refugia-within-refugia, in contrast, spatially restricted and highly structured refugial distributions predict well-defined population genetic structure within the greater refugium region, but with signals of population expansion and genetically rich populations in secondary contact zones.

South-east Europe and adjacent West Asia (northwestern Middle East: Asia Minor, Caucasian and Caspian regions) are important hotspots of genetic diversity for a number of taxa (e.g. Kramp et al., 2009; Dufresnes et al., 2016; Jablonski et al., 2016) with many locally endemic lineages (Geiger et al., 2014). A combination of the effects of historical climate, topography and dramatic changes in sea level during the last 5 million years resulted in variable patterns of continental colonization from this refugium (e.g. Durand et al., 1999; Kotlík and Berrebi, 2007), which is reflected in the complex structure of the Balkan biotic assemblage (Economidis and Banarescu, 1991). In particular, the Aegean Sea was approximately 120 m lower than it is today during the Plio-Pleistocene (Perissoratis and Conispoliatis, 2003), redisposing rivers, streams, wetlands, and their connections (Hewitt, 2000). Until the early Holocene (11,500 BP), the Black Sea basin was an extensive freshwater lake (Degens and Ross, 1972), and likely to have supported, rather than limited, dispersal of freshwater taxa. A connection between the Black and Caspian Seas established periodically during the Pleistocene, with termination of the last connection estimated at 17,000 – 16,000 years BP (Reid and Orlova, 2002).

We investigated genetic diversity of bitterling fishes in the West Palearctic (*Rhodeus amarus* species complex), with special attention on the relative contribution of refugial populations to the current bitterling diversity in the region. Bitterling belong to a distinct subfamily of cyprinid fishes (Cyprinidae: Acheilognathinae) (Chang et al., 2014) that parasitize freshwater molluscs by ovipositing in their gill chambers (Smith et al., 2004). The global centre of distribution of bitterling is in East Asia (China, Japan, Korea and adjacent countries) (Chang et al., 2014; Kawamura et al., 2014). One bitterling lineage colonized the western Palearctic (Bohlen et al., 2006; Chang et al., 2014) and three species of the genus *Rhodeus* have been formally named in region (Kottelat and Freyhof, 2007). In addition to widespread *Rhodeus amarus* (Bloch, 1782), *Rhodeus colchicus* Bogutskaya and Komlev, 2001 was described from the foothills of the western Caucasus on the basis of osteological characters and confirmed as a separate lineage by mitochondrial genetic data (Bohlen et al., 2006; Zaki et al., 2008). Populations in the River Vardar were described as *Rhodeus sericeus meridionalis* Karaman 1924 and were proposed as representing a valid species, *R. meridionalis*, by Kottelat and Freyhof

104 (2007) on the basis of its genetic divergence. However, bitterling diversity in other areas of the West
105 Palearctic distribution is poorly explored and their phylogenetic relationships and intraspecific
106 structure are unresolved, perhaps as a result of relatively rapid diversification, incomplete lineage
107 sorting and/or gene flow among populations/species (Bohlen et al., 2006; Bryja et al., 2010; Chang et
108 al., 2014).

109 The natural distribution of bitterling in the West Palearctic covers a large part of continental Europe,
110 excepting the Iberian and Apennine Peninsulas, Fennoscandia, Denmark and the region east and
111 north of the Dnieper basin (Kottelat and Freyhof, 2007). Recent introductions have expanded its
112 range to Great Britain, Denmark, Italy and several basins in European Russia (Kozhara et al., 2006;
113 Bartáková et al., 2018). Bitterling are thermophilic (Smith et al., 2004; Van Damme et al., 2007) and
114 former studies established that their populations survived the Pleistocene glacial periods in refugia
115 located in the Balkans, Black Sea region, lower Danube, and southern Caucasian region (Bohlen et al.,
116 2006; Zaki et al., 2008; Bryja et al., 2010; Bektas et al., 2013). Some of these populations repeatedly
117 expanded during warmer climatic conditions and colonized large parts of Europe (Kozhara et al.,
118 2007; Van Damme et al., 2007; Bryja et al., 2010). Populations from the lower Danube spread
119 throughout the River Danube basin and colonised most of Central and West Europe, while a
120 population from a putative refugium in the northern Black Sea region colonized north-eastern
121 Europe (Bohlen et al., 2006; Bryja et al., 2010), with the area of secondary contact in Central and
122 West Europe (Bryja et al., 2010). Preliminary evidence suggested that other populations remained
123 isolated in their refugia and differentiated in allopatry (Bryja et al., 2010).

124 To investigate the patterns of genetic differentiation of the West Palearctic bitterling species
125 complex, we used dense and fine-scale sampling of bitterling populations in the region, with a
126 particular focus on the Balkan region and other parts of the greater Mediterranean region with the
127 reported presence of bitterling populations. We specifically concentrated on the analysis of (1)
128 genetic differentiation among bitterling lineages to describe their phylogenetic patterns, (2) genetic

diversity within the lineages and populations to understand their demographic history and (3) geographic aspects of the distribution of particular lineages to characterise past and recent connections between the lineages and their contribution to the current expansion of the bitterling in the West Palearctic region.

## 2. Materials and methods

### *2.1. Sampling and DNA extraction*

We analysed a total of 1,038 fish from 60 sampling sites, with particular attention to the bitterling distribution in the southern part of its range in the West Palearctic. Reference to zoogeographical regions follows terminology for freshwater fishes from Economidis and Banarescu (1991). On the basis of preliminary analysis of an individual-based Bayesian clustering procedure, implemented in the STRUCTURE software, we pooled some populations for the final analyses. The pooled samples came from adjacent sites within the same streams, with the exception of the IRAZ population that was composed of pooled samples from several bitterling populations from the Caspian region (Azerbaijan, Iran) that lacked a precise geo-reference (Azerbaijan) and were too small for most population genetic analyses. All individuals in the IRAZ sample belonged to the same mitochondrial haplogroup. Fish were collected between 2004 and 2015. Sampling sites, along with details on the number of analysed individuals and haplotype composition are listed in Appendix A. DNA extraction was performed from small fin clips taken from the caudal fin and stored in 96% ethanol, using the DNeasy Blood and Tissue Kit (Qiagen) following a standard protocol. Extracted DNA was stored at -20°C.

### *2.2. Genotyping*

All bitterling individuals were genotyped at 12 microsatellite loci in three multiplex PCR sets (Table C.1). Primer names and sequences were taken from Dawson et al. (2003) and Reichard et al. (2008). A detailed genotyping protocol is provided in Bartáková et al. (2018). The length of DNA fragments was analysed manually using GeneMapper v. 5.0 (Applied Biosystems).

A partial sequence of mitochondrial gene cytochrome *b* (*CYTB*) was amplified by primers Thr-H (5´-ACCTCCRATCTYCGGATTACA-3´) and Glu-L (5´-GAAGAACCACCGTTGTTATTCAA-3´) in a subset of individuals (Appendix A) following the protocol of Bohlen et al. (2006) with conditions described in Bartáková et al. (2018). PCR products were commercially Sanger-sequenced by Macrogen Europe. Sequence editing was performed in SeqScape V.2.5 (Applied Biosystems) and aligned in BioEdit v.7.0.9.0 (Hall, 1999), producing a final alignment of 914 bp. All sequences have been deposited in GenBank (accession numbers MH041650–MH041876).

### *2.3. Phylogenetic analysis and haplotype distribution based on mitochondrial DNA*

The most appropriate substitution model for the *CYTB* dataset was the Generalised time-reversible model with a gamma-distributed rate variation across sites (GTR + G), which was selected on the basis of BIC in TOPALi v. 2.5 (Milne et al., 2009). Three unique sequences of *Rhodeus sericeus* (the sister lineage of the western Palearctic bitterling complex (Chang et al., 2014)) from Lake Kenon (River Amur basin) were used as outgroups in all phylogenetic analyses. Phylogenetic relationships were inferred by maximum likelihood (ML) and Bayesian (BI) approaches. ML analysis was performed in RAxML 8.2.10 (Stamatakis, 2014), applying the GTR + G model (option -m GTRGAMMA). The robustness of the nodes was assessed by the default rapid bootstrap procedure with 1,000 replications (option -# 1000). Bayesian analysis was performed by Markov Chain Monte Carlo (MCMC) simulation using MrBayes 3.2.6 (Ronquist and Huelsenbeck, 2003). Two independent analyses were initiated from random trees. Three heated and one cold chain were run for 10 million

generations per run, sampling every 1,000 generations. For each run, 25% of trees were discarded as burn-in. Bayesian posterior probabilities were used to evaluate branch support of the consensus tree. All phylogenetic analyses were performed on Cipres Science Gateway webserver (Miller et al., 2010). The final trees were edited in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree).

To analyse the phylogeny in a time-calibrated framework, we estimated the relative time to the most recent common ancestor (TMRCA) for all unique sequences in BEAST v. 2.4.6 (Bouckaert et al., 2014). To avoid unrealistic assumptions of a strict molecular clock, a relaxed molecular clock model (uncorrelated lognormal) was used for the analysis (Drummond et al., 2006). We performed two runs (10 million generations each) under the GTR + G model with parameters sampled every 1000 iterations. We discarded the first 20% as burn-in, based on likelihood stationarity visualized using Tracer 1.5 (Rambaut et al., 2018). The effective sample sizes (*ESS*) of all parameters sampled from MCMC were > 300. The resulting parameter and tree files from the two runs were combined in LogCombiner 2.4.6 (Bouckaert et al., 2014) and a maximum clade credibility tree was calculated in TreeAnnotator 2.4.6 (Bouckaert et al., 2014). Because there are no reliable tools for absolute calibration of the *Rhodeus* molecular clock (no suitable fossils or well-dated biogeographic events), we used this analysis only for relative comparison of divergence events.

The haplotypes and their frequencies were identified using DnaSP v. 5.10.01 (Librado and Rozas, 2009). The relationships among haplotypes were also visualised as a median-joining (MJ) network in Network 5.0.0.1 (Bandelt et al., 1999) using an equal transition/transversion ratio. All sequences were geo-referenced and the distribution of the haplogroups was inspected visually. The matrix of mean p-distances within/between *CYTB* clades (using 101 unique haplotypes of 914 bp) was calculated in MEGA v. 6 (Tamura et al., 2013) and standard errors were estimated with 1000 bootstraps.

200 *2.4. Historical demography based on mtDNA variation*

201 To analyse historical demography we identified six genetic lineages (see Appendix A for assignment
202 to lineages) based on the phylogenetic analyses, the haplotype network of mtDNA and the
203 geographical distribution of haplotypes. We used the reduced dataset of 208 sequences (914 bp)
204 (Appendix A). We excluded three populations (RHI2, STR2, MER1) because they represented
205 secondary contacts of differentiated lineages and contained individuals with haplotypes from two
206 lineages. Diversity estimates for the six lineages; i.e. number of polymorphic sites (*Np*), number of
207 haplotypes (*Nh*), haplotype diversity (*Hd*), nucleotide diversity (*p*), and average number of nucleotide
208 differences (*k*) were computed in DnaSP 5.00.04 (Librado and Rozas, 2009).

209 To infer demographic histories, we estimated the neutrality indices (Tajima's D and Fu's Fs statistic)
210 in DnaSP v. 5.10.01 (Librado and Rozas, 2009). These indices are sensitive to population size change
211 and return significantly negative values in the case of recent population expansion. As an additional
212 test of demographic expansion, the distribution of pairwise nucleotide differences in each lineage
213 (mismatch distribution; MD) was calculated in DnaSP. We used the sum of square deviations (SSD)
214 between the observed and expected mismatch as a test statistic for the validity of the estimated
215 stepwise expansion model (Schneider and Excoffier, 1999). Parameter $\tau$ (the moment estimator of
216 time to the expansion) was estimated with DnaSP using the moment method of Rogers (1995)
217 assuming the infinite sites model (IFM) and, additionally, in ARLEQUIN using the method of Schneider
218 and Excoffier (1999) to relax the IFM assumption. Confidence intervals were obtained by a
219 parametric bootstrap approach based on 1,000 replicates performed in ARLEQUIN. Under the
220 assumption of the (sudden) demographic expansion model the MD also permits estimation of the
221 time of onset of population expansion $\tau$ ($\tau = 2*t*\mu$; $t$ = time in years, $\mu$ = mutation rate per locus of
222 914 bp). We converted the parameter $\tau$, calculated from the mismatch distribution, to estimate the
223 time since the expansion ($t$) using the equation $t = \tau/2\mu$, assuming the commonly accepted
224 substitution rates of 0.76% per million years for *CYTB* in cyprinid fishes (Zardoya and Doadrio, 1999)

and an average generation time of one year for the bitterling (Smith et al., 2004; Konečná and Reichard, 2011).

We reconstructed historical population size dynamics of the main lineages backward in time using the coalescent-based Bayesian skyline plot (BSP) in BEAST 2.4.6. (Drummond and Rambaut, 2007). Analyses were run twice for each lineage and the model of sequence evolution for each lineage was selected with the jModelTest (Posada, 2008) under the BIC criterion. The MCMC simulations were run with 50 million iterations with a sampling increment of 5,000 and 25% burn-in. Results were checked for convergence and stationarity of different runs in Tracer 1.6 and the outputs from two runs were combined in the LogCombiner 2.4.6 module. The BSPs were produced in Tracer 1.6.

*2.5. Analysis of genetic variability and structure based on microsatellite data*

To analyse interpopulation and intrapopulation genetic variability on nuclear markers, the proportion of null alleles (*NA*) at each locus and population was estimated in FreeNA (Chapuis and Estoup 2007). The number of alleles (*A*), and observed ($H_O$) and expected ($H_E$) heterozygosities were calculated in GENETIX. The rarefaction method in FSTAT 2.9.3.2 (Goudet, 2001) was used to calculate allelic richness (*AR*) for each population standardized for a minimum sample size of 7 individuals. Deviations from Hardy-Weinberg equilibrium (*HWE*) were tested for each locus and population using the Markov chain method in the software GENEPOP and the correction for multiple testing was performed using the false discovery rate (FDR) approach (Benjamini and Hochberg, 1995) in QVALUE (Storey, 2002). For these analyses populations with a low number (<10) of sampled individuals were not used (see Table B.2). Analyses of intrapopulation variability were computed only from 10 microsatellite loci; locus *Rser13* (D05) was excluded due to high polymorphism (110 alleles) and *Rser09* (D12) was excluded given a high level of null alleles (mean of 5.64% per population). Genetic

248 differentiation among study populations was quantified by $F_{ST}$ (Weir and Cockerham, 1984) and their
249 significance was tested by 1000 permutations in GENETIX 4.05 (Belkhir et al., 1996-2004).

250 An individual-based Bayesian clustering procedure implemented in STRUCTURE 2.3.4 (Hubisz et al.,
251 2009) was used to detect the best genetic structure among sampled individuals (n = 1002 individuals
252 from 52 populations). The program was run with 20 independent simulations for each of *K* from 1 to
253 30, with 1 million iterations for each simulation, following a burn-in period of 100,000 iterations. The
254 computation was realised using an admixture ancestry model and correlated allele frequencies
255 model (with $\lambda$ = 1). The web-based software STUCTURE HARVESTER (Earl and vonHoldt, 2012) was
256 used for parsing and summarizing output data from STRUCTURE. It reformatted data for downstream
257 programs and produced the likelihood of *K*; i.e. Ln Pr(X|K), for inferring the best number of real
258 populations in the datasets using the method of Evanno et al. (2005). The results of 20 replicate runs
259 for each *K* were combined using the Greedy algorithm of CLUMPP 1.1.1 (Jakobsson and Rosenberg,
260 2007) and summary outputs for each *K* were displayed using DISTRUCT v. 1.1 (Rosenberg, 2004).

261

262 *2.6. Specific testing of isolation/gene flow*

263 Based on the comparison of results from *CYTB* and microsatellite analyses, we specifically tested the
264 origin of the MER1 population with respect to possible gene flow between genetic groups "C" and
265 "D" using Approximate Bayesian Computation (ABC; Beaumont et al., 2002) implemented in DIYABC
266 2.0.4 (Cornuet et al., 2014). This analysis models population histories by combining population
267 divergences, admixtures and population size fluctuations and compares alternative evolutionary
268 scenarios by estimating their relative support and quantifying parameters for particular scenarios
269 (Cornuet et al., 2014). Three "populations" were generated according to the Bayesian assignment of
270 their genetic structure and haplotype distribution. The populations MER1 (35 individuals) and MER3
271 (26 individuals) were formed exclusively by those single population samples. The SSTR (i.e. "South

Struma/Strymon") population was composed of STR4, STR5 and STR6 population samples (total of 57 individuals). Effective population size, timing of events (merging, splitting or change in effective population size) and rates of admixture in the case of merging events, were used to describe seven scenarios (Table B.3, Figure C.1). The range of uniform priors is provided in Table B.4. The generalized stepwise model was used as a mutation model (GSM; Estoup et al., 2002). All microsatellites used have regular motifs (motif lengths of 2 bp, except of a 4 bp motif in Rser11) and all microsatellite mutation parameters were at default settings. An average generation time of one year was used (Smith et al., 2004).

We simulated 1 million data sets per scenario. For each simulation, a set of summary statistics was computed to compare with the observed dataset for the best model selection. The relative posterior probability (95% credible intervals) of each scenario was determined using the 1% of simulated data sets closest to the observed data (Euclidean distances) and logistic regression was used to select the most likely evolutionary scenario. The posterior parameter distributions were estimated from the closest 1% of simulated data sets of the most likely model through a local linear regression procedure (Cornuet et al., 2008). The assessment of goodness-of-fit of the best model was checked by evaluating consistency of the observed data with the posterior predicted distribution of the model for the best scenario. We carried out model checking using all summary statistics, including those that had not been used in the initial ABC analyses for model selection (Cornuet et al., 2010).

**3. Results**

Genotyping success of microsatellites was high (97.14%), resulting in multilocus genotypes of 1,002 individuals from 52 populations. All analyses of genetic structure were based on a complete dataset of all 12 loci. The analyses of intrapopulation genetic variability were performed using a reduced dataset of 10 loci and 924 individuals from 38 populations; excluding populations with <10

296 individuals (Appendix A). Analysis of mitochondrial variability was based on 227 *CYTB* sequences (914
297 bp; 101 haplotypes) from 57 sites (Appendix A).

298

299 *3.1. Analysis of mitochondrial variability*

300 Phylogenetic analysis of 101 unique *CYTB* sequences revealed a highly structured tree of West
301 Palaearctic bitterling (Figure 1, Figure C.2) with six well-supported phylogenetic lineages, referred to
302 as lineages "A"-"F". However, relationships among these lineages were not well resolved, except the
303 sister position of lineages "C" and "D". The topologies of the trees differed between BEAST and
304 MrBayes (Figure 1a vs. 1b), suggesting a rapid radiation that produced the current diversity of
305 mitochondrial lineages. This view is further strengthened by similar estimates for the age of the first
306 divergence within particular lineages (Figure 1a).

307 A haplotype network supported the same pattern (Figure 2, Figure C.3). Two lineages occur
308 throughout Central and West Europe (Figure 3a). Lineage "A" (Danubian, green on Figures 1-3a) was
309 the only lineage distributed in the River Danube basin and was dominant across Central Europe.
310 Surprisingly, lineage "A" was also found in the ancient lakes of Prespa and Skadar in the western
311 Balkans (south Adriatic-Ionian division). Lineage "B" (Eastern, orange on Figures 1-3a) was widely
312 distributed around the Black Sea coast (excluding the eastern shore) and extended to north-eastern
313 Europe (Figure 3a). Its distribution included East Bulgaria and Thracean-East Macedonia subdivisions
314 of the Balkans (the Rivers Mesta/Nestos and Marmaras, Lake Vistonis). It was also present in the
315 Mediterranean region of southern France (the River Rhône), a case of a disjunct distribution. One
316 middle Rhine population (RHI2, River Nida, central Germany) represented a likely secondary contact
317 between the "A" and "B" lineages; two individuals possessed haplotype h8 of lineage "B" while three
318 individuals had haplotypes h1 and h9 of lineage "A". Another case of sympatry between "A" and "B"
319 lineages was located in the upper River Struma/Strymon (STR2, southern Bulgaria, Aegean

watershed); four individuals had the h1 haplotype commonly found in lineage "A" and three individuals possessed haplotype h22 of the "B" lineage.

Two other lineages were restricted to the western part of the Ponto-Aegean division of the Balkans in Greece (Figure 3a). Lineage "C" (Thracean, yellow on Figures 1-3a) occurred in the Thracean-East Macedonian subdivision (the River Struma/Strymon basin, including Lake Volvi, upper River Mesta/Nestos and River Macropotamos). Lineage "D" (meridionalis, black on Figures 1-3a) was found only in Macedonia-Thessaly subdivision of the Ponto-Aegean region (Rivers Vardar/Axios, Aliakmon/Haliacmon and Pinios/Pineios) and represents the putative *R. meridionalis*. Interestingly, the "C" and "D" lineages coexist in the River Aliakmon/Haliacmon (MER1); one individual had haplotype h47 of the lineage "C" and three individuals had haplotypes of the "D" lineage (h46, h77). No sympatric coexistence between "B" and "C" lineages was detected despite their interwoven geographical distribution in the Thracean-East Macedonia subdivision along the northern Aegean coast (Figure 3a).

Finally, two lineages were exclusive to the eastern part of the study area, southern Caucasian region. Lineage "E" (colchicus, grey on Figures 1-3a) occurred in the Transcaucasian part of the Black Sea region (Georgia; described as *R. colchicus*). Lineage "F" (Caspian, blue in Figures 1-3a) included samples from southern Caspian region (Azerbaijan and Iran). Mean p-distances within and between the *CYTB* lineages are provided in Table B.5.

### *3.2. Demographic changes in mtDNA lineages*

Demographic history was analysed separately for all six mitochondrial lineages. The summary of mtDNA variation and outcome of neutrality tests are detailed in Table 1. The two lineages with a continental distribution displayed a lack of neutrality (suggestive of a recent expansion); the Eastern lineage "B" in both estimates (Tajima's and Fu's Fs tests) and Danubian lineage "A" in Fu's Fs test only

344 (Table 1). However, all lineages demonstrated unimodal mismatch distribution curves of population
345 growth (Figure C.4). None of the sums of squared deviations (SSD) of the mismatch distribution was
346 significant, indicating that the curves fitted the sudden expansion model (Table 1). The mismatch
347 distribution was similar among lineages (Figure C.4), with the lowest mean number of differences in
348 the lineages "F" (Caspian) (0.865; age expansion parameter $\tau$ = 0.984), "A" (Danubian) and "E"
349 (colchicus) (1.879; $\tau$ = 1.930) (Table 1), suggesting relatively recent demographic expansions. Mean
350 number of differences, $\tau$ values and estimated timing of the most important demographic expansion
351 for each group indicate relatively older expansions of the lineages "B" (Eastern) and "C" (Thracean)
352 (Table 1). Based on BSP, population growth was detected in the most widespread lineages ("A", "B",
353 "C") but no change in population size was detected in lineages "D", "E" and "F" (Figure C.5), in
354 agreement with the neutrality test results (Table 1).

355

356 *3.3. Intrapopulation variability – microsatellites*

357 Expected ($H_e$) and observed ($H_o$) heterozygosity, allelic richness (*AR*), and tests of HWE for all 38
358 population samples are given in Table B.2. Significant departure from HWE was detected in 13 out of
359 38 populations. After FDR correction, deviations from HWE were mostly limited to a single locus,
360 suggesting locally increased frequency of null alleles. Only the population IRAZ showed deviance
361 from HWE on six loci and population STR2 on four loci. The deficit of heterozygotes in the IRAZ
362 population was likely due to pooling individuals from several localities (Wahlund effect and different
363 allele frequencies in each population), and in the STR2 population it was possibly caused by
364 admixture of two genetic (mitochondrial) lineages and a subpopulation structure. The range of *AR*
365 was 2.23-6 (rarefaction estimate for the lowest sample size N = 7). The populations with the highest
366 genetic diversity were from the lower reaches of large rivers of the Black Sea region – the Danube
367 (DNB8; mitochondrial lineage "A", DNB10; lineage "B") and Dnieper and Dniester (DNP, DNS; lineage
368 "B"), and the population from the River Pinios/Pineios (MER3, lineage "D"). In contrast, the lowest AR

was detected in the populations from small streams in the European part of Turkey (REZ, SUT; western Black Sea watershed; lineage "B"), the River Vistula in Poland (VIS, lineage "B") and in Lake Skadar (SKA, lineage "A").

### *3.4. Genetic structure inferred from nuclear markers*

Populations were highly structured, indicating a significant role for genetic drift. Non-significant estimates of $F_{ST}$ were only found between two pairs of geographically adjacent populations (DNB2 and DNB3 in the River Morava and MAR1 and MAR2 in the River Maritsa) (Table B.6). The best model in STRUCTURE separated genetic variation into 25 clusters; other acceptable models were for 2, 4, 6, 7, 9 and 18 populations (Figure C.6). The distribution of mitochondrial lineages corresponded well with the geographic distribution of clusters detected by microsatellite data. The geographic distribution of individual clusters for $K$ = 9 rather than $K$ = 6 is shown on Figure 3b, because the populations from the most genetically diverse region were further separated into well-supported clusters in that model. Arrangement of individuals into 2–30 clusters, along with their assignment to particular mitochondrial lineages, is visualised in Figure 4, with geographic distribution of individual clusters for $K$ = 7 and 18 in Appendix C (Figures C.7 and C.8, respectively).

The populations from the Ponto-Aegean division (the most genetically diverse region) were separated into four main clusters in the model for $K$ = 9. The first is composed of populations from a Macedonia-Thessaly subdivision (MER1, MER3), where most individuals had mitochondrial haplotypes from lineage "D". The second cluster is composed of populations from the River Struma/Strymon (STR1-STR2, STR4-STR6), in close geographic proximity on the upper River Mesta/Nestos (MES1) and River Macropotamos (MCP), and corresponding closely to mitochondrial lineage "C". In population STR2, a sympatric occurrence of mitochondrial "A" and "C" lineages was detected (compare Figures 3a and 3b).

393 The next two clusters represent a subdivision of populations belonging to the widespread
394 mitochondrial lineage "B" and closely follow the division between Thracean-East Macedonia and East
395 Bulgaria subdivisions. One cluster (orange-blue in Figure 3b) is formed by populations from Thrace-
396 East Macedonia (the River Maritsa; MAR1-MAR5), both sides of the Marmara Sea in western Turkey
397 (REZ, SUT, BIG, BBU), River Marmaras (MRM), the lower River Mesta/Nestos (MES2) and Lake
398 Vistonis (PLS). Another cluster (light green in Figure 3b) consists of populations of the East Bulgaria
399 subdivision – the rivers emptying into the Black Sea (RUS, GOL, IZV, ROP, FAK, VEL, KAM). These two
400 clusters possess a finer substructure at higher *K* (Figures C.6 and C.7). Higher values of *K* (*K* = 18,
401 Figure C.8) also reveal some other populations to be discrete, including from the lakes Skadar (SKA)
402 and Prespa (PRE) in the South Adriatic-Ionian subdivision, and River Yeşilırmak on the southern coast
403 of the Black Sea (SMS).

404 In accord with the mitochondrial dataset, populations from Southern Caucasian region
405 (mitochondrial lineages "E" and "F") formed consistent and unique clusters. The structure of
406 bitterling populations from Central Europe has been addressed elsewhere (Bryja et al., 2010;
407 Bartáková et al., 2018) and current data confirmed former conclusions; two lineages ("A" and "B"
408 colonized the European continent via north-eastern and Danubian routes, with secondary contacts in
409 Central Europe.

410

411 *3.5. Testing reproductive isolation of R. meridionalis*

412 We used ABC to test alternative scenarios for the origin of MER1, in which two distinct mitochondrial
413 lineages ("C" and "D") co-occur. The best model suggested that the MER1 population derived from
414 MER3 (i.e. meridionalis from the Pinios/Pineios drainage) without immigration from the SSTR
415 population, where mitochondrial lineage "C" predominates (Table B.3, Figure C.1). This indicates that
416 the lineage "C" haplotype present in the MER1 population has a different origin to that in the SSTR

417 population or, alternatively, that the level of introgression in MER1 is negligible. Model checking
418 (Figure C.9) demonstrated that the observed dataset falls within the cloud of simulated parameter
419 estimates. Estimates of the posterior parameter distribution are shown in Table B.7. The analysis
420 suggests that the MER1 and SSTR populations arose in the same period but from different sources.
421 Admixture was not supported; the best model had a support of 37.6%, while the model with
422 admixture between MER1 and SSTR was supported by only 6.5%.

423

424 **4. Discussion**

425 *4.1. Continental perspective*

426 The genetic structure of bitterling populations across Europe illustrates how postglacial expansions
427 have led to the loss of genetic variation (Bernatchez and Wilson, 1998). The overall phylogenetic
428 pattern was concordant between mitochondrial and nuclear DNA markers. Deeply divergent
429 lineages, signals of expansion and admixture events suggested persistence in multiple small isolated
430 populations during climatically unfavourable periods. This finding supports predictions from the
431 refugia-within-refugia model (Goméz and Lund, 2007) that has been established in other animal (e.g.
432 Dufresnes et al., 2016; Jablonski et al., 2016) and plant taxa (e.g. Kramp et al., 2009). The range of
433 bitterling in Europe north of the Balkans is dominated by two lineages that expanded from refugia in
434 south-eastern Europe. Lineage "A" colonized Central and West Europe from the lower Danube
435 refugium via the River Danube system (lineage WEST *sensu* Bohlen et al. 2006) and lineage "B" (EAST
436 lineage *sensu* Bohlen et al., 2006) colonized eastern and northern Europe from an area in the
437 northern part of the Black Sea via river systems east of the Carpathians (Bryja et al., 2010). This
438 colonization pattern is congruent with those recorded in many other freshwater fishes (e.g. Durand
439 et al., 1999; Nesbø et al., 1999). Two other European lineages ("C" and "D") are restricted south of
440 the Balkans and have not contributed to the contemporary continental colonization. Two West Asian

441 lineages from regions south of the Caucasus mountain range also remained endemic and have not
442 expanded. Lineage "E" from the Black Sea basin (Georgia) has been formally described as *R. colchicus*
443 (Bogutskaya and Komlev, 2001), while lineage "F" from the Caspian basin (Azerbaijan and northern
444 Iran) has yet to be taxonomically investigated.

445 Interestingly, we have confirmed the presence of lineage "B" in the River Rhône in the
446 Mediterranean region of southern France, an apparent mismatch of a predictable spatial distribution
447 of the lineages "A" and "B" across continental Europe. Similarly, Bohlen et al. (2006) reported the
448 presence of the same mitochondrial lineage from the River Saone from central France, a tributary of
449 the Rhône. Relatively divergent haplotypes in the River Rhône population (Bryja et al., 2010) suggest
450 that this region contains a relict population from continental colonisations in previous interglacials
451 that survived the last glacial maximum in local refugia, similarly to the pattern hypothesised for the
452 European barbel, *Barbus barbus* (Kotlík and Berrebi, 2001). A region of secondary contact between
453 the lineages "A" and "B" in the River Elbe and the River Rhine basins in Central Europe is a pattern
454 repeatedly found in other freshwater (Durand et al., 1999; Bernatchez, 2001; Kotlík and Berrebi,
455 2001) and terrestrial taxa (Hewitt, 1999). The results of the present study, analysed at a finer
456 resolution than previous studies, confirmed previous conclusions on the geography and dynamics of
457 postglacial colonization (Bohlen et al., 2006; Zaki et al., 2008; Bryja et al., 2010; Bektas et al., 2013)
458 and add further support to the emerging evidence that Danubian and Black Sea refugia were the
459 almost exclusive sources of contemporary postglacial colonization of continental Europe by
460 freshwater fishes (Durand et al., 1999; Nesbø et al., 1999; Bernatchez, 2001).

461

462 *4.2. Diversity of southern populations*

463 The current study primarily focused on the genetic diversity in the region of putative bitterling
464 refugia during glacial cycles in the Balkans, around the Black Sea and in adjacent Caspian region.

Bitterling are generally thermophilic, with a predominantly subtropical distribution in the Far East (Chang et al., 2014, Kawamura et al., 2014). Only a single species complex is present in Europe, apparently following a single dispersal event to Europe dated to the late Pliocene (< 3 Mya) with subsequent rapid diversification (Bohlen et al., 2006; Chang et al., 2014).

The Ponto-Aegean region of the Balkans was found to support a diverse geographic mosaic of populations, though the three genetic lineages present in that region were relatively well geographically separated. The lineage "D" of the Macedonia-Thessaly region, has been suggested to represent a valid species, *Rhodeus meridionalis* Karaman 1924, by Kottelat and Freyhof (2007). Bryja et al. (2010) contested the validity of *R. meridionalis* by demonstrating the presence (albeit rare) of a widespread haplotype of the lineage "A" in *R. meridionalis* populations. Indeed, tributaries of the Danube and Vardar/Axios rivers were hypothesised to be connected via river captures by Economidis and Banarescu (1991). Our current analysis does not demonstrate any evidence of the connection between the lineages "A" (Danubian) and "D" (meridionalis). Instead, we found that two endemic mitochondrial lineages "D" and "C" co-occur in the lower River Aliakmon/Haliacmon (MER1). Data from nuclear microsatellites analysed by ABC modelling did not detect any admixture between the nuclear genomes of "C" and "D", indicating a case of ancestral polymorphism. We acknowledge that our limited sample size cannot exclude the role of recent translocation.

The lineage "C" (Thracean) represents a second unique Balkan lineage that did not contribute to the continental expansion of the bitterling. It has a mosaic distribution with lineage "B" in the Thracean-East Macedonia region. There was a connection between the Black and Aegean Seas via the former River Aegeopotamos (Economidis and Banarescu, 1991) that likely supported dispersal of the lineage "B" to the Thracean region. Lineage "C" is prevalent in the western part of the Thracean-East Macedonia region, while lineage "B" populations are more common in the east of the region and replace lineage "C" in the River Maritsa basin and eastwards (Figure 3a). The pattern is repeated on nuclear genetic markers, with an indication of some admixture of both gene pools in one

Mesta/Nestos population (MES2) (Figure 3b). We hypothesise that the mosaic distribution of the "B" and "C" lineages is contingent on paleogeographic patterns during the lowered level of the Aegean Sea when dispersal via the River Aegeopotamos was possible.

The River Struma/Strymon harbours the highest bitterling genetic diversity. The upper Struma/Strymon in the north contains a southern secondary contact of the two continental lineages ("A" and "B") that are also admixed in Central Europe. While it could be a consequence of past connections between tributaries of the River Danube and Struma/Strymon basins (Banarescu, 1990), a recent introduction by anglers that commonly use the bitterling as a baitfish in that region (Bogoev, 1999; Kozhara et al., 2007) is a more plausible explanation. The dominant "B" lineage is replaced by the "C" lineage in the middle and lower Struma/Strymon, with no record of their sympatric occurrence. The same pattern is reported for *Cobitis* loaches in the River Struma/Strymon basin (Choleva et al., 2008), another frequently used commercial baitfish in Bulgaria (Bogoev, 1999). In other regions, the mitochondrial and nuclear genetic structure of the European bitterling was found to mismatch boundaries of watersheds, with lineage "A" (Danubian) present in the River Oder (Baltic watershed) in the north-eastern part of the Czech Republic and a signal of admixture from the Danube basin in the bitterling populations from the River Elbe basin (North Sea watershed) (Bartáková et al., 2018). These departures from otherwise congruent fine-scale genetic structure in Central Europe were also attributed to human-mediated translocations, related to aquaculture trade or game-fish management (Bartáková et al., 2018).

The bitterling range includes two lake systems of ancient origin (> 5 MYA), Lake Prespa (part of the Lake Ohrid system) and Lake Skadar, which harbour diverse freshwater lineages, including many endemic species (Albrecht and Wilke, 2008; Pešić and Glöer, 2013). Unexpectedly, our data suggest that bitterling populations in the lakes are of recent origin, with depauperate genetic diversity. Both lakes support lineage "A" populations that are closely related to geographically proximate

514 populations from the River Danube basin, suggesting that baitfish introduction to those lakes might
515 also explain their presence and low genetic diversity.

516 The region around the Black Sea (except the eastern part) is dominated by lineage "B" populations
517 that are finely sub-structured at nuclear markers. In the south-western part of that region (where our
518 sampling was particularly dense), the populations divide according to suggested regional division
519 between Thracean-East Macedonia and East Bulgaria subdivision of Economidis and Banarescu
520 (1991). Thracean-East Macedonian populations form cluster with the populations from the southern
521 coast of the Marmara Sea, suggesting recent gene flow. Elevated freshwater discharge from the
522 Black Sea across the Bosporus to the Marmara and Aegean Seas led to low salinity conditions in the
523 northern Aegean Sea approximately 16–8,500 BP (Aksu et al., 1999), permitting recent connection
524 across the Marmara Sea. A similar pattern was observed in other taxa, for example in *Triturus* newts
525 (Taberlet et al., 1998). Populations from many coastal streams of southwest Black Sea region form a
526 separate genetic group. Further genetic divergence is apparent in distant populations from the
527 Danube delta, lower River Dnieper and Dniester, as well as from Asia Minor along the northern coast
528 of the Black Sea (Figs 4, C.6 and C.7).

529 The west Transcaucasian (Georgian) region of the Black Sea basin is inhabited by lineage "E" that has
530 been described as a separate species, *Rhodeus colchicus* by Bogutskaya and Komlev (2001) on the
531 basis of osteological characters. Its uniqueness was confirmed at a mitochondrial marker by Bohlen
532 et al. (2006) and we corroborated its distinction using nuclear microsatellite markers. Bitterling
533 populations from the Caspian watershed of the southern Caucasian region (Azerbaijan, Iran) formed
534 a separate, well-characterised lineage ("F"). Distinct Caucasian and Caspian lineages have been
535 reported in amphibians (e.g. Dufresnes et al., 2016), a semi-anadromous fish (*Rutilus frisii*) (Kotlík et
536 al., 2008) and in freshwater fishes of the genus *Alburnoides* (Stierandová et al., 2016). Caspian
537 bitterling populations were not included in previous analyses of the European bitterling complex
538 (Bogutskaya and Komlev, 2001; Bohlen et al., 2006; Zaki et al., 2008; Bryja et al., 2010; Bektas et al.,

2013), though a subfamily level phylogeny suggested its distinctiveness (Chang et al., 2014). A finer-scale analysis of the two Asian bitterling lineages in the West Palearctic would require denser sampling in the area.

## 5. Conclusions

Describing fine-scale genetic variability and structure, we demonstrate the plausibility of the refugia-within-refugia model for European bitterling populations, a freshwater fish species with a limited dispersal capacity. West Palearctic bitterling persisted during periods of less favourable climatic conditions in several isolated populations along the southern margin of their current distribution. During recent interglacials, only two lineages colonized much of continental Europe, with secondary contact in West and Central Europe. A diverse mosaic of populations was detected in the Aegean region, with limited local coexistence. Recent admixture of separate lineages in the south could be a consequence of ancestral polymorphism or recent introductions related to the use of bitterling as baitfish rather than natural secondary contact zones in that region, in contrast to natural secondary contact in Central Europe. Divergence of populations in the Aegean and Caspian regions, and in the region east of the Black Sea, reflect their different origins, genetic distinction and failure to contribute to the recent continental expansion of the West Palearctic bitterling. In the present study, we refrain from raising any taxonomic implications of the findings until morphological samples of divergent lineages are available, but we acknowledge that such a high level of genetic divergence coupled with geographic and genetic isolation requires taxonomic examination of the species complex and demonstrates that the Caspian and Thracean lineages represent unique evolutionary units.

## Acknowledgements

Principal financial support came from the Czech Science Foundation (13-05872S) to MR. RŠ was

supported by the Ministry of Culture of the Czech Republic (DKRVO 2018/14, National Museum, 00023272). LC acknowledges institutional support RVO: 67985904. We thank J. Boehlen, K. Schindehutte, M. Koutrakis, E. Lasne, M. Ondračková, Z. Valová, A. Naseka and T. Eros for providing samples from their collection and to S. Zogaris, S. Shumka and D. Ulqini for their help with fieldwork. Sampling complied with the legal regulations of the respective countries. Samples utilised in the study have been lawfully acquired and were collected prior to The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from Their Utilization to the Convention on Biological Diversity has been in effect. New sequences used in this study are available in GenBank under accession numbers MH041650–MH041876. Further details on used samples are specified in Appendix A.

## Author contributions

M.R. and J.B. conceived the idea; R.Š., Y.B., T.S., L.C., C.S. and M.R. collected the material, V.B. produced genetic data, V.B. and J.B. analysed the data, V.B., J.B. and M.R. wrote the first version of the manuscript that was commented and approved by all authors.

## References

Adams, J.M., Faure, H., 1997. Preliminary vegetation maps of the world since the Last Glacial Maximum: an aid to archaeological understanding. J. Archaeol. Sci. 24, 623-647.

Aksu, A.E., Hiscotta, R.N., Yaşarb, D., 1999. Oscillating Quaternary water levels of the Marmara Sea and vigorous outflow into the Aegean Sea from the Marmara Sea-Black Sea drainage corridor. Mar. Geol. 153, 275-302.

Albrecht, C., Wilke, T., 2008. Ancient Lake Ohrid: biodiversity and evolution. Hydrobiologia 615, 103-140.

Banarescu, P., 1990. Zoogeography of Fresh Waters. General distribution and dispersal of freshwater animals, v. 1. Aula Verlag, Wiesbaden, Germany.

590 Bandelt, H.-J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific
591 phylogenies. Mol. Biol. Evol. 16, 37-48.

592 Bartáková, V., Bryja, J., Reichard, M., 2018. Fine-scale genetic structure of the European bitterling at
593 the intersection of three major European watersheds. BMC Evol. Biol. 18, 105.

594 Beaumont, M.A., Zhang, W.Y., Balding, D.J., 2002. Approximate Bayesian computation in population
595 genetics. Genetics 162, 2025-2035.

596 Bektas, Y., Belduz, A.O., Turan, D., 2013. The phylogenetic position of Turkish populations within the
597 European Bitterling, Rhodeus amarus (Osteichthyes: Cyprinidae). Zool. Middle East 59, 39-50.

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

601 Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate - a practical and powerful
602 approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57, 289-300.

603 Bernatchez, L., 2001. The evolutionary history of brown trout (*Salmo trutta*) inferred from
604 phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. Evolution 55,
605 351-379.

606 Bernatchez, L., Wilson, C.C., 1998. Comparative phylogeography of Neartic and Paleartic fishes. Mol.
607 Ecol. 7, 431-452.

608 Bogoev, M., 1999. Bulgarian Fishing Encyclopedia. Trud Publishing House, Sofia. (In Bulgarian)

609 Bogutskaya, N.G., Komlev, A.M., 2001. Some new data on the morphology of *Rhodeus sericeus* sensu
610 lato (Cyprinidae: Acheilognathinae) and a description of a new species, *Rhodeus colchicus*, from West
611 Transcaucasia. Proc. Zool. Inst. 287, 81-97.

612 Bohlen, J., Šlechtová, V., Bogutskaya, N., Freyhof, J., 2006. Across Siberia and over Europe:
613 phylogenetic relationships of the freshwater fish genus *Rhodeus* in Europe and the phylogenetic
614 position of *R. sericeus* from the River Amur. Mol. Phylogenet. Evol. 40, 856-865.

615 Bouckaert, R., Heled, J., Kühner, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A.,
616 Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS
617 Comput. Biol. 10, e1003537.

618 Bryja, J., Smith, C., Konečný, A., Reichard, M., 2010. Rangewide population genetic structure of the
619 European bitterling (*Rhodeus amarus*) based on microsatellite and mitochondrial DNA analysis. Mol.
620 Ecol. 19, 4708-4722.

621 Chang, C.H., Li, F., Shao, K.T., Lin, Y.S., Morosawa, T., Kim, S., Lee, J.S., He, S., Smith, C., Reichard, M.,
622 Miya, M., Sado, T., Uehara, K., Lavoué, S., Chen, W.J., Mayden, R.L., 2014. Phylogenetic relationships
623 of Acheilognathidae (Cypriniformes: Cyprinoidea) as revealed from evidence of both nuclear and
624 mitochondrial gene sequence variation: Evidence for necessary taxonomic revision in the family and
625 the identification of cryptic species. Mol. Phylogenet. Evol. 81, 182-194.

626 Chapuis, M.-P., Estoup, A., 2007. Microsatellite null alleles and estimation of population
627 differentiation. Mol. Biol. Evol. 24, 621-631.

628 Choleva, L., Apostolou, A., Ráb, P., Janko, K., 2008. Making it on their own: sperm-dependent hybrid
629 fishes (*Cobitis*) switch the sexual hosts and expand beyond the ranges of their original sperm donors.
630 Philos. Trans. R. Soc. Lond., B, Biol. Sci. 363, 2911-2919.

631 Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.M., Estoup,
632 A., 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about
633 population history using single nucleotide polymorphism, DNA sequence and microsatellite data.
634 Bioinformatics 30, 1187-1189.

635 Cornuet, J.-M., Ravigne, V., Estoup, A., 2010. Inference on population history and model checking
636 using DNA sequence and microsatellite data with the software DIYABC (v1.0). Bmc Bioinformatics 11,
637 401.

638 Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T.,
639 Estoup, A., 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate
640 Bayesian computation. Bioinformatics 24, 2713-2719.

641 Culling, M.A., Janko, K., Boron, A., Vasilev, V.P., Cote, I.M., Hewitt, G.M., 2006. European colonization
642 by the spined loach (*Cobitis taenia*) from Ponto-Caspian refugia based on mitochondrial DNA
643 variation. Mol. Ecol. 15, 173-190.

644 Dawson, D.A., Burland, T.M., Douglas, A., Le Comber, S.C., Bradshaw, M., 2003. Isolation of
645 microsatellite loci in the freshwater fish, the bitterling *Rhodeus sericeus* (Teleostei: Cyprinidae). Mol.
646 Ecol. Notes 3, 199-202.

647 Degens, E.T., Ross, D.A., 1972. Chronology of the Black Sea over the last 25,000 years. Chem. Geol.
648 10, 1-16.

649 Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC
650 Evol. Biol. 7, 214.

651 Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with
652 confidence. PLoS Biol. 4, 699-710.

653 Dufresnes, C., Litvinchuk, S.N., Leuenberger, J., Ghali, K., Zinenko, O., Stöck, M., Perrin, N., 2016.
654 Evolutionary melting pots: a biodiversity hotspot shaped by ring diversifications around the Black Sea
655 in the Eastern tree frog (*Hyla orientalis*). Mol. Ecol. 25, 4285-4300.

656 Durand, J.D., Persat, H., Bouvet, Y., 1999. Phylogeography and postglacial dispersion of the chub
657 (*Leuciscus cephalus*) in Europe. Mol. Ecol. 8, 989-997.

658 Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing
659 STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359-361.

660 Economidis, P.S., Banarescu, P.M., 1991. The distribution and origins of freshwater fishes in the
661 Balkan Peninsula, especially in Greece. Int. Revue ges. Hydrobiol. 76, 257-283.

662 Estoup, A., Jarne, P., Cornuet, J.M., 2002. Homoplasy and mutation model at microsatellite loci and
663 their consequences for population genetics analysis. Mol. Ecol. 11, 1591-1604.

664 Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the
665 software structure: A simulation study. Mol. Ecol. 14, 2611-2620.

666 Geiger, M.F., Herder, F., Monaghan, M.T., Almada, V., Barbieri, R., Bariche, M., Berrebi, P., Bohlen, J.,
667 Casal-Lopez, M., Delmastro, G.B., Denys, G.P., Dettai, A., Doadrio, I., Kalogianni, E., Kärst, H., Kottelat,
668 M., Kovačić, M., Laporte, M., Lorenzoni, M., Marčić, Z., Özuluğ, M., Perdices, A., Perea, S., Persat, H.,
669 Porcelotti, S., Puzzi, C., Robalo, J., Šanda, R., Schneider, M., Šlechtová, V., Stoumboudi, M., Walter, S.,
670 Freyhof, J., 2014. Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding
671 accuracy of its freshwater fishes. Mol. Ecol. Resour. 14, 1210-1221.

672 Gómez, A., Lunt, D.H., 2007. Refugia within refugia: patterns of phylogeographic concordance in the
673 Iberian Peninsula, in: Weiss, S., Ferrand, N. (Eds.), Phylogeography of southern European refugia.
674 Springer, Dordrecht, Netherlands. pp. 155-188

675 Goudet, J., 2001. FSTAT, a programme to estimate and test gene diversities and fixation indices
676 (version 2.9.3). Available from https://www2.unil.ch/popgen/softwares/fstat.htm.

677 Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program
678 for Windows 95/98/NT. Nucleic Acids Symp. Ser. No., 41, 95-98.

679 Hewitt, G.M., 1999. Post-glacial recolonization of European biota. Biol. J. Linnean Soc. 68, 87-112.

680 Hewitt, G.M., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907-913.

681 Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J. K., 2009. Inferring weak population structure with
682 the assistance of sample group information. Mol. Ecol. Resour. 9, 1322-1332.

683 Jablonski, D., Jandzik, D., Mikulíček, P., Džukić, G., Ljubisavljević, K., Tzankov, N., Jelić, D., Thanou, E.,
684 Moravec, J., Gvoždík, V., 2016. Contrasting evolutionary histories of the legless lizards slow worms
685 (*Anguis*) shaped by the topography of the Balkan Peninsula. BMC Evol. Biol. 16, 99.

686 Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for
687 dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23,
688 1801-1806.

689 Kawamura, K., Ueda, T., Arai, R., Smith, C., 2014. Phylogenetic relationships of bitterling fishes
690 (Teleostei: Cypriniformes: Acheilognathinae), inferred from mitochondrial cytochrome *b* sequences.
691 Zool. Sci. 31, 321-329.

692 Konečná, M., Reichard, M., 2011. Seasonal dynamics in population characteristics of European
693 bitterling *Rhodeus amarus* in a small lowland river. J. Fish Biol. 78, 227-239.

694 Kotlík, P., Berrebi, P., 2001. Phylogeography of the barbel (*Barbus barbus*) assessed by mitochondrial
695 DNA variation. Mol. Ecol. 10, 2177-2185.

696 Kotlík, P., Berrebi, P., 2007. Nested clade phylogeographical analysis of barbel (*Barbus barbus*)
697 mitochondrial DNA variation, in: Weiss, S., Ferrand, N. (Eds.), Phylogeography of Southern European
698 Refugia. Springer, Dordrecht, Netherlands, pp. 315-325.

699 Kotlík, P., Marková, S., Choleva, L., Bogutskaya, N.G., Ekmekçi, F.G., Ivanova, P.P., 2008. Divergence
700 with gene flow between Ponto-Caspian refugia in an anadromous cyprinid *Rutilus frisii* revealed by
701 multiple gene phylogeography. Mol. Ecol. 17, 1076-1088.

702 Kottelat, M., Freyhof, J., 2007. Handbook of European Freshwater Fishes. Kottelat, Cornol,
703 Switzerland.

704 Kozhara, A.V., Zhulidov, A.V., Gollasch, S., Przybylski, M., Poznyak, V.G., Zhulidov, D.A., Gurtovaya,
705 T.Y.U., 2007. Range extension and conservation status of the bitterling, *Rhodeus sericeus amarus* in
706 Russia and adjacent countries. Folia Zool. 56, 97-108.

707 Kramp, K., Huck, S., Niketić, M., Tomović, G., Schmitt, T., 2009. Multiple glacial refugia and complex
708 postglacial range shifts of the obligatory woodland plant *Polygonatum verticillatum*
709 (Convallariaceae). Plant Biol. 11, 392-404.

710 Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism
711 data. Bioinformatics 25, 1451-1452.

712 Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. "Creating the CIPRES Science Gateway for inference of
713 large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE),
714 14 Nov. 2010, New Orleans, LA pp. 1-8.

715 Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F., Wright, F., 2009. TOPALi
716 v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and
717 multi-core desktops. Bioinformatics 25, 126-127.

718 Nesbø, C.L., Fossheim, T., Wøllestad, A., Jakobsen, K.S., 1999. Genetic divergence and
719 phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial
720 refugia and postglacial colonization. Mol. Ecol. 8, 1387-1404.

721 Perissoratis, C., Conispoliatis, N., 2003. The impacts of sea-level changes during latest Pleistocene
722 and Holocene times on the morphology of the Ionian and Aegean seas (SE Alpine Europe). Mar. Geol.
723 196, 145-156

724 Pešić, V., Glöer, P., 2013. A new freshwater snail genus (Hydrobiidae, Gastropoda) from Montenegro,
725 with a discussion on gastropod diversity and endemism in Skadar Lake, ZooKeys 281, 69-90.

726 Posada, D., 2008. jModeltest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25, 1253-1256.

727 Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Tracer v1.7, Available from
728 http://tree.bio.ed.ac.uk/software/tracer/

729 Reichard, M., Smith, C., Bryja, J., 2008. Seasonal change in the opportunity for sexual selection. Mol.
730 Ecol. 17, 642-651.

731 Reid, D., Orlova, M., 2002. Geological and evolutionary underpinnings for the success of Ponto-
732 Caspian species invasions in the Baltic Sea and North American Great Lakes. Can. J. Fish. Aquat. Sci.
733 59, 1144-1158.

734 Rogers, A., 1995. Genetic evidence for a Pleistocene population explosion. Evolution 49, 608-615.

735 Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed
736 models. Bioinformatics 19, 1572-1574.

737 Rosenberg, N.A., 2004. Distruct: a program for the graphical display of population structure. Mol.
738 Ecol. Notes 4, 137-138.

739 Schneider, S., Excoffier, L., 1999. Estimation of past demographic parameters from the distribution of
740 pairwise differences when the mutation rates vary among sites: application to human mitochondrial
741 DNA. Genetics 152, 1079-1089

742 Smith, C., Reichard, M., Jurajda, P., Przybylski, M., 2004. The reproductive ecology of the European
743 bitterling (*Rhodeus sericeus*). J. Zool. (Lond.) 262, 107-124.

744 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and Ronquistpost-analysis of
745 large phylogenies. Bioinformatics 30, 1312-1313.

746 Stewart, J.R., Lister, A.M., 2001. Cryptic northern refugia and the origins of the modern biota. Trends
747 Ecol. Evol. 16, 608-613.

748 Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of
749 species in space and time. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 277, 661-71.

750 Stierandová, S., Vukić, J., Vasil'eva, E.D., Zogaris, S., Shumka, S., Halačka, K., Vetešník, L., Švátora, M.,
751 Nowak, M., Stefanov, T., Koščo, J., Mendel, J., 2016. A multilocus assessment of nuclear and
752 mitochondrial sequence data elucidates phylogenetic relationships among European spirlins
753 (*Alburnoides*, Cyprinidae). Mol. Phylogenet. Evol. 94, 479-491.

754 Storey, J.D., 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Series B 64, 479-498.

755 Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and
756 postglacial colonization routes in Europe. Mol. Ecol. 7, 453-464.

757 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary
758 Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725-2729.

759 Van Damme, D., Bogutskaya, N., Hoffmann, R.C., Smith, C., 2007. The introduction of the European
760 bitterling (*Rhodeus amarus*) to west and central Europe. Fish Fish. 8, 79-106.

761 Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure.
762 Evolution 38,1358-1370.

763 Zaki, S.A.H., Jordan, W.C., Reichard, M., Przybylski, M., Smith, C.A., 2008. A morphological and
764 genetic analysis of the European bitterling species complex. Biol. J. Linnean Soc. 95, 337-347.

765 Zardoya, R., Doadrio, I., 1999. Molecular evidence on the evolutionary and biogeographical patterns
766 of European cyprinids. J. Mol. Evol. 49, 227-237.

**Figure Captions**

**Figure 1** Mitochondrial phylogeny of the European bittering complex based on 98 ingroup and 3 outgroup haplotypes of the mitochondrial gene *CYTB*. (a) Ultrametric tree with relative dating from BEAST 2.4.6. Posterior probabilities (/bootstrap support for maximum likelihood tree from RAxML in Fig. 1b) are shown above branches (only values higher than 0.70/50 are shown). Putative taxonomic names or a label of geographic distribution are shown for each lineage. (b) Bayesian reconstruction from MrBayes 3.2.6 with Bayesian inference posterior probabilities/bootstrap support from the maximum likelihood analysis for each node.

**Figure 2** Haplotype network of 101 haplotypes (914 bp) from 227 *CYTB* sequences (including *R. sericeus* as the outgroup) of the European bitterling complex. Length of branches in the network is proportional to the number of substitutions along a given branch. Circle size is proportional to haplotype frequency. Further information on haplotype data is provided in Appendix A and Figure C.2.

**Figure 3** Geographic distribution of (a) mitochondrial lineage diversity and (b) genetic diversity from nuclear microsatellites based on assignment to 9 clusters following STRUCTURE analysis across study area. (a) The colours correspond to the lineages as defined in Figure 1 and indicate the relative proportions of lineages at a particular locality. The inset details the distribution in the most diverse part of the Balkan region. Names of localities correspond to those in Appendix A. Borders of freshwater zoogeographical regions (divisions and subdivisions) *sensu* Economidis and Banarescu (1991) are shown by red lines. (b) Pie chart colours represent the proportional membership of individuals to microsatellite-based clusters inferred from the models selected using the approach of Evanno et al. (2005) (for further details and barplots for all models see Appendix C, for geographic

791 distribution of the diversity assigned to K = 7 and K = 18 clusters, see Figures C.6 and C.7,

792 respectively).

793

794 **Figure 4** Bayesian analysis of genetic similarity among *Rhodeus* populations performed in STRUCTURE

795 for 1002 individuals from 52 populations for *K* = 2-25. Assignment to a particular mitochondrial

796 lineage ("A" to "F") is indicated for each population.

## TABLES

### Table 1

Analysis of historical demography within six clades based on 914 bp of CYTB.

| Lineage | *N* | *S* | *H* | *Hd* | *Pi (%)* | *k* | *Tajima´s D* | *Fu´s Fs* | *τ Arl* (95% CI) | *τ DnaSP* | *SSD* | *P(SSD)* | *Obs. mean* | *t* (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| "A" | 42 | 21 | 17 | 0.858 ±0.047 | 0.0026± 0.0004 | 2.334 | -1.721 | -9.496* | 1.156 (0.295-4.936) | 1.533 | 0.0014 | 0.81 | 2.334 | 83 225 (21 238 – 355 364) |
| "B" | 100 | 52 | 46 | 0.965±0.008 | 0.0047±0.0003 | 4.316 | -1.898* | -38.564* | 4.605 (3.123 - 5.424) | 4.316 | 0.0016 | 0.29 | 4.336 | 331 533 (224 838 – 390 497) |
| "C" | 25 | 12 | 12 | 0.9±0.037 | 0.0036±0.0003 | 3.327 | -0.116 | -3.451 | 4.484 (1.818 - 6.721) | 2.983 | 0.0279 | 0.10 | 3.327 | 322 822 (130 886-483 873) |
| "D" | 8 | 7 | 7 | 0.964±0.077 | 0.0023±0.0004 | 2.071 | -1.107 | -4.418 | 2.273 (0.291 - 3.783) | 2.071 | 0.0192 | 0.45 | 2.071 | 163 643 (20 950 – 272 354) |
| "E" | 14 | 6 | 6 | 0.868±0.054 | 0.0021±0.0004 | 1.879 | -0.014 | -1.136 | 1.930 (0.938 - 3.156) | 1.879 | 0.0102 | 0.45 | 1.879 | 138 949 (67 531 – 227 214) |
| "F" | 19 | 4 | 5 | 0.637±0.105 | 0.001±0.0002 | 0.866 | -0.718 | -1.555 | 0.984 (0.424 - 1.885) | 0.865 | 0.0056 | 0.52 | 0.865 | 70 842 (30 526 – 135 709) |
| Total | 208 | | | | | | p<0.05 * | p<0.01 * | | | | | | |

*N* = number of individuals*; S* = number of variable sites*; H* = number of haplotypes; *Hd* = haplotype (gene) diversity ± SD*; Pi* = nucleotide diversity (in %) ± SD; *k* = average number of nucleotide differences; *τ* = onset of population expansion assuming the stepwise growth model (τ = 2*t*μ; t = time in years, μ = mutation rate per locus); *t* = time in years computed from τ Arl; *SSD* = sum of squared deviations; *P(SSD)* is the probability of observing a less good fit between the model and the observed distribution by chance; *Obs. mean* = the mismatch observed mean. Fu's Fs significance p < 0.01 is marked by *. Tajima's D significance p < 0.05 is marked by *.

805 Supplementary material and additional supporting information associated with this article can be
806 found, in the online version, at appendices A-C.

807

808 **Appendix A:** List of analysed populations (XLS)

809

810 **Appendix B:** Additional tables for population genetic analysis (XLS)

811 Table B.1: Microsatellite primers

812 Table B.2: Genetic diversity

813 Table B.3: Summary of scenarios used in the ABC analysis

814 Table B.4: The prior parameter distributions used in the ABC analysis

815 Table B.5: Mean p-distances within and between the *CYTB* clades.

816 Table B.6: Matrix of pairwise $F_{ST}$

817 Table B.7: The posterior parameter distributions inferred from the ABC winning scenario

818

819

820 **Appendix C:** Additional figures for population genetic analysis (PPT)

821 Fig. C.1: Graphical schemes of the ABC scenarios

822 Fig. C.2: DensiTree

823 Fig. C.3: Haplotype network with haplotypes

824 Fig. C.4: Mismatch distribution

825 Fig. C.5: Bayesian coalescent skyline plots

826 Fig. C.6: Evanno´s delta K

827 Fig. C.7: Alternative STRUCTURE pie-charts of Q-values, K = 7

828 Fig. C.8: Alternative STRUCTURE pie-charts of Q-values, K = 18

829 Fig. C.9: PCA plots of the best model from the ABC analysis