Hydrobiologia (2021) 848:345–361
https://doi.org/10.1007/s10750-020-04442-4

PRIMARY RESEARCH PAPER

Population genomics of the southern Caspian Sea Vobla *Rutilus lacustris*

Iraj Hashemzadeh Segherloo · Fariborz Ghojoghi · Seyedeh Narjes Tabatabaei · Eric Normandeau · Cecilia Hernandez · Eric Hallerman · Brian Boyle · Louis Bernatchez

Received: 16 April 2020 / Revised: 26 September 2020 / Accepted: 14 October 2020 / Published online: 25 October 2020

Abstract The genus *Rutilus* is widespread in the western and central Palearctic region. In the Caspian Sea, the taxonomic status of different populations of *Rutilus lacustris* has been debated due to different sub-specific names attributed to each population. We genotyped 7,984 single nucleotide polymorphisms and sequenced the mitochondrial cytochrome C oxidase subunit I gene of 37 *R. lacustris* and *Rutilus frisii* from the southeast and southwest Caspian Sea and the Aras River in the Kura River drainage. We analysed data using clustering, Bayes factor delimitation, introgression, assessment of migration rate, and phylogenetic analyses. The results showed that the southeast and southwest Caspian Sea populations of *R. lacustris* were closely related, but highly differentiated from *R. lacustris* in the Aras River. The Bayes factor delimitation test supported the existence of three populations of *R. lacustris* in the studied area. Three hybrid individuals with mtDNA from *Abramis brama* or *R. frisii* and nuclear DNA from *R. lacustris* were detected. To protect *R. lacustris* in the southern Caspian Sea, we propose that the Aras River and sea-run *R. lacustris* be treated as two separate conservation units and the southern Caspian Sea *R. lacustris* populations should be viewed as two potentially separate management units.

Handling editor: Christian Sturmbauer

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10750-020-04442-4) contains supplementary material, which is available to authorized users.

I. Hashemzadeh Segherloo (✉)
Department of Fisheries Sciences, Faculty of Natural Resources and Earth Sciences, Shahr-e-Kord University, Shahr-e-Kord, Iran
e-mail: ihashem@sku.ac.ir

I. Hashemzadeh Segherloo · S. N. Tabatabaei · E. Normandeau · C. Hernandez · L. Bernatchez
Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec G1V 0A6, Canada

F. Ghojoghi (✉)
Department of Fisheries, Faculty of Agriculture and Natural Resources, Islamic Azad University, Azadshahr Branch, Azadshahr, Iran
e-mail: fariborzghojoghi@yahoo.com

E. Hallerman
Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

B. Boyle
Plateforme d'analyses génomiques et Institut de biologie intégrative et des systèmes, Université Laval, Quebec, Quebec, Canada

Keywords Next-generation sequencing · mtDNA · Population genetics · Hybridization · Conservation units

Introduction

Identification of biological units at the population and species levels is highly important for the design and implementation of rigorous conservation and management plans. Fishes are among the animal groups most subject to management actions aimed at both sustainable commercial use and conservation. Fishes, especially freshwater and anadromous species, face a variety of threats, including climate change, impoundment of rivers, overfishing, degradation of spawning grounds, and water pollution (Kiabi et al., 1999; Abdoli, 2000; Reyhani et al., 2010). In the Iranian part of the Caspian Sea, a number of migratory fish species, including sturgeons (Acipenseridae), Caspian salmon *Salmo caspius* Kessler, 1877, vyrezub *Rutilus frisii* (Nordmann, 1840), and Vobla *Rutilus lacustris* (Pallas, 1814), have been artificially propagated and stocked to address these threats to their survival (Abdoli, 2000). The roaches (genus *Rutilus*), distributed in Europe and West Asia, include 13 named species, four of which exist in the Caspian Sea basin (Kottelat & Freyhof, 2007; Levin et al., 2017): common roach *Rutilus rutilus* (Linnaeus, 1758), vyrezub *R. frisii*, Mezamor roach *R. schelkovnikovi* (Derzhavin, 1926) and Vobla *R. lacustris*. *Rutilus lacustris* is one of the most commercially important fish species but comprising declining populations in the south Caspian Sea (Kiabi et al., 1999; Rezvani et al., 2006). Furthermore, the species is an important prey item for the Caspian Sea beluga *Huso huso* (Linnaeus, 1758) (Keyvanshokooh & Kalbassi, 2006). Due to the commercial and ecological importance of declining populations of *R. lacustris* to sustaining southern Caspian Sea sturgeon stocks, the Iranian Fisheries Organization has performed an important restocking program over the past three decades. Conservation and management of *R. lacustris* are faced with critical questions about the definition of the species and critical intraspecific units. Some authors have considered it a subspecies of *R. caspicus* with three races, including Turkmen roach *R. caspicus natio knipowitschi* Pravdin, 1927 in the southeastern Caspian Sea, *R. caspicus natio Tscharchalensis* Berg, 1932, and Kura roach *R. caspicus natio kurensis* Berg, 1932 in the southwest Caspian Sea in the Kura River drainage (Abdoli, 2000; Keyvanshokooh & Kalbassi, 2006; Rezvani et al., 2006).

Population genetic analyses of populations from the southeastern and southwestern Caspian Sea using restriction fragment length polymorphisms (RFLPs) of the mitochondrial cytochrome-*b* gene (Rezvani et al., 2006) and random amplified polymorphic DNA (RAPD) (Keyvanshokooh & Kalbassi, 2006) markers did not reveal significant levels of inter-population differentiation. However, results from microsatellite DNA studies (Keyvanshokooh et al., 2007; Reyhani et al., 2010) returned contradictory results, complicating management of *R. lacustris* in the southern Caspian Sea basin. For example, Keyvanshokooh et al. (2007) concluded that *R. lacustris* populations from the southeast and the southwest Caspian Sea were significantly differentiated, while Reyhani et al. (2010) considered them a single panmictic population. Based on osteological analysis, Eagderi et al. (2017) concluded that the southern Caspian Vobla populations are probably of the same systematic unit with phenotypic plasticity driven by environmental conditions, but also proposed that the Turkmen *R. lacustris* may be a distinct taxon. Also, it has been shown that the southwestern and southeastern *R. lacustris* populations were morphologically distinct (Abdoli, 2000; Keyvanshokooh & Kalbassi, 2006; Rezvani et al., 2006; Keyvanshokooh et al., 2007). All of the aforementioned contradictory conclusions made based on data from morphological and genetic approaches have left the identification of management/conservation units of the species in the southern Caspian Sea unresolved. Therefore, robust conclusions have yet to be reached pertaining to the relationships of these populations or forms of *R. lacustris*.

The development of next-generation sequencing (NGS) that can produce data for hundreds to thousands of loci distributed across the genome in both model and non-model species has enabled conservation biologists to gain much finer and exact understanding of biodiversity and to define defensible conservation units (Alda et al., 2018; MacGuigan & Near, 2018). To support application of NGS technology, bioinformatic approaches to handle large amounts of genomic data also have been developed (Catchen et al., 2013). Such developments have increased our ability to assess

population structure and relationships at fine geographic scales (Benestan et al., 2015; Hashemzadeh Segherloo et al., 2018; Tabatabaei et al., 2020a). This provides the opportunity for conservationists and biodiversity managers to adapt management measures based on a clearer knowledge of population structure, systematics and taxonomy.

To address the methodological drawbacks of previous studies and to provide a clearer picture of the management/conservation units of *R. lacustris* in the southern Caspian Sea basin, our goals were to assess the genetic relationships among three populations representing the southeastern and southwestern Caspian Sea and the Aras River (in the Kura River drainage, southwestern Caspian Sea) populations of *R. lacustris* along with the closely related species *R. frisii*. We used a combination of genomic (genotyping-by-sequencing, GBS) and mitochondrial cytochrome c oxidase subunit I (*COI*) partial sequence data. Our aims were to: a) infer the relationships among the respective populations, and b) provide guidelines for conservation and management of *R. lacustris* in the southern Caspian Sea.

Materials and methods

Collections

The genetic relationships of 27 specimens belonging to three *R. lacustris* populations in the southeastern (ten specimens) and southwestern Caspian Sea (seven specimens) and the Aras River (ten specimens), plus ten *R. frisii* from the southeastern Caspian Sea, which were caught with seine nets by local fishermen in the southeastern and southwestern Caspian Sea and in the Aras Reservoir were analysed (Fig. 1; Table 1).

DNA extraction

DNA was extracted from muscle samples preserved in alcohol with the salt extraction method of Aljanabi & Martinez (1997) and an additional RNAse treatment (Benestan et al., 2015). The quality of extracted total DNA was assessed by electrophoresis through a 1% agarose gel. Degraded DNA samples—when detected—were excluded. For the GBS analysis, the exact concentration of extracted DNA specimens was determined using Picogreen (Invitrogen, Carlsbad, CA) reads and adjusted to 20 ng/µl.

Mitochondrial DNA sequencing

To amplify the mitochondrial *COI* region, the *FishCOI*-F: 5′-AAYCAYAAAGAYATYGGYACCCT-3′ and *FishCOI*-R: 5′-TANACTTCNGGRTGNCCRAAGAAYCA-3′ primers (Ivanova et al., 2007) were used. The 12.5-µl PCR reaction contained 6.25 µl of Accustart II PCR mix (www.quantabio.com), 0.5 µl of a 10-µm solution of each primer, 3.25 µl deionized water, and 2 µl of total DNA (10 ng/µl). Amplification cycles included a preliminary denaturation for 60 s at 94°C; 30 cycles at 94°C for 30 s, 55°C for 30 s, 72°C for 30 s; and a final extension for 45 s at 72°C. Sequencing with the *FishCOI*-F primer was performed using an ABI Prism 3130 sequencer (http://www.thermofisher.com) at the IBIS sequencing platform (Laval University, Quebec City, Canada; http://www.ibis.ulaval.ca).

Genotyping-by-sequencing

The GBS libraries were prepared following the methods of Mascher et al. (2013) (for more step-by-step details, see Abed et al., 2019). In brief, genomic DNA was digested with the *Pst*I and *Msp*I restriction enzymes. After digestion, the DNA samples were ligated to individual-specific adapters containing both barcodes and IonTorrent-specific sequences (Supplementary Tables I and II). After heat inactivation of the ligase, the samples were pooled and the pooled library was size-selected on a BluePippin electrophoretic unit (https://sagescience.com/) with elution times set between 46 and 60 min on a 2% dye-free agarose cassette. After 10 cycles of PCR amplifications, the final library was checked for quality on a DNA high-sensitivity BioAnalyzer chip (Agilent) and quantified using a Qbit (ThermoFisher). The final amplified library contained fragments from 150 to 350 nucleotides in length. Template preparation for sequencing and loading on an IonTorrent P1 chip was performed using an Ion CHEF instrument, and sequencing was performed on an Ion Proton at the Plateforme d'Analyses Génomique of the Institut de Biologie Intégrative et des Systèmes (IBIS, Université Laval, Quebec City, Canada; http://www.ibis.ulaval.ca).

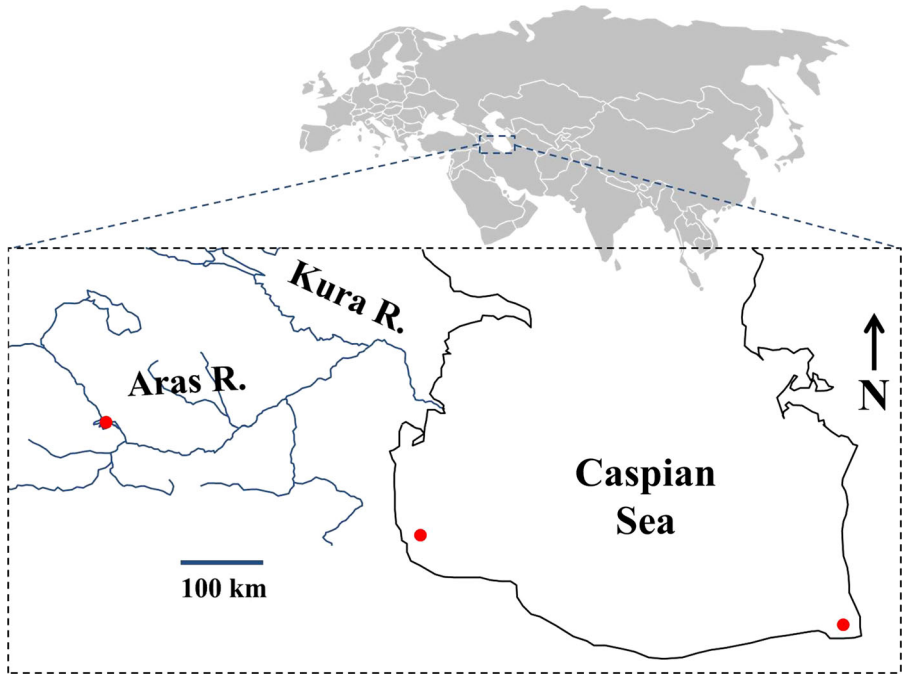


Fig. 1 Sampling localities in the southeast and southwest Caspian Sea and in the Aras River. Red-filled circles denote sampling localities

Table 1 Sampling details including collected sample size, samples with mtDNA sequences (n. mtDNA), samples with SNP data (n. Genome), and geographic details

Species	N_i	N_{COI}	Acces. no.	N_{SNP}	Locality	Geographic coordinates	
						Lat	Long
R. frisii	10	10	MT756344-MT756353	10	Southeast Caspian Sea, Turkmen Port, Iran	36.879978	54.022211
R. lacustris	7	7	MT756354-MT756360	7	Southwest Caspian Sea, Talesh, Iran	37.814623	49.012462
R. lacustris	10	10	MT756361-MT756370	9	Aras River reservoir, border of Iran and Nakhchivan, Azerbaijan	39.136942	45.348522
R. lacustris	10	10	MT756371-MT756380	8	Southeast Caspian Sea, Turkmen Port, Iran	36.879978	54.022211

N_i initial sample size, N_{COI} samples with COI sequence data, N_{SNP} samples with SNP data, *Acces. No.* GenBank Accession number

Bioinformatics

Cytochrome C Oxidase subunit I

The 674-bp partial *COI* sequences were edited using BioEdit v. 7.2.5 (Hall, 2001). The raw *COI* sequences were aligned using Muscle (Edgar, 2004) with the default options in MEGA7 (Kumar et al., 2016). Nucleotide and haplotype diversity indices for non-hybrid individuals were calculated with DnaSP V.6.10.03 (Rozas et al., 2017). To visualize the number of haplotypes, their relative frequencies, their mutational relationships, and their temporal order (Templeton, 2004), a TCS haplotype network

(Clement et al., 2000) was produced using PopART-1.7 (http://popart.otago.ac.nz). To provide context for the haplotype network and to root the phylogenetic tree, *COI* sequences for European Chub *Squalius cephalus* (Linnaeus, 1758) and common nase *Chondrostoma nasus* (Linnaeus, 1758) were used as outgroups (Levin et al., 2017). As some sequences from GenBank did not have the same lengths as our sequences, we used a common 595-bp sequence length for haplotype network and phylogenetic tree reconstructions. In addition, other *Rutilus* spp. sequences from GenBank were used for reconstruction of the haplotype network and phylogenetic tree (Supplementary Table III). In order to clarify phylogenetic relationships, both maximum likelihood (ML) and Bayesian phylogenetic analyses were performed using RaxMLGUI 1.5 beta (Silvestro & Michalak, 2012) and MrBayes-3.2.7 (Ronquist et al., 2012). For both phylogenetic inferences performed using the codon partitioning approach, the best codon partitioning schemes were determined using PartitionFinder-2.1.1 (Lanfear et al., 2017). The options used for ML tree reconstruction were: bootstrap replicates = 1000, phylogenetic model = GTR + Gamma + I, ML + rapid bootstrap, and multiple out-groups = *S. cephalus* and *C. nasus*. For the Bayesian inference of phylogeny, a 2,000,000-generation MCMC search was run with sampling the Markov chain every 100 generations. The first 25% of the threes produced were discarded as burn-in. For both analyses, three codon partitions were used. The phylogenetic trees were visualised using FigTree v. 1.4. As we did not have *Rutilus* samples from other parts of the Caspian Sea basin, Cytochrome b (*Cyt*-b) sequences from the southern Caspian Sea and from the western and northern Caspian Sea basin from GenBank (Supplementary Table IV), sources of some of which overlapping the distribution of the populations we analysed here, were used to construct a TCS haplotype network with PopART-1.7 (Supplementary Figure SI).

Genomic data

To remove the amplification adapter sequences, raw sequence reads were trimmed with Cutadapt (Martin, 2011), and sequence quality was assessed with FastQC (Andrews, 2010). Process_radtags included in STACKS V.1.48 (Catchen et al., 2013) was used to extract and trim (trimmed length: 80 bp) sequences. The loci were built de novo with ustacks (minimum depth of coverage required to create a stack (m) = 4, maximum distance (in nucleotides) allowed between stacks (M) = 3, and maximum distance allowed to align secondary reads to primary stacks (N) = 5). Cstacks, with a maximum between-loci mismatch parameter of 1, was used to group the loci together across individuals and to catalogue them, and then to clarify the alleles at each locus, and loci within each individual were matched against the catalog with sstacks. The rxstacks error-correction pipeline (lnl_filter = − 1, conf_filter, conf_lim = 0.75, prune_haplo) and then cstacks and stacks with the same parameters as above were run. Then populations was used (options: minimum percentage of individuals in a population required to process a locus for that population ($-r$) = 0.5, minimum number of populations a locus must be present into process a locus ($-p$) = 1, minimum genotype coverage ($-m$) = 4, minimum genotype log likelihood (lnl_lim) = − 10), and SNPs were exported as a VCF (variant call format) file. The VCF file created by populations was further filtered with the 05_filter_vcf.py script from stacks_workflow (https://github.com/enormandeau/stacks_workflow) with the following options: maximum allelic imbalance ($-I$) = 4, minimum genotype coverage ($-m$) = 4, minimum proportion of genotyped samples at one SNP ($-p$) = 70, and maximum proportion of heterozygous samples ($-H$) = 0.6, minimum accepted F_{is} ($-f$) = − 0.5, maximum accepted F_{is} ($-F$) = 0.5, maximum number of SNPs in one locus ($-s$) = 10. To avoid problems related to linkage between loci, only unlinked SNPs were kept in the final VCF file using the 11_extract_unlinked_snps.py from stacks_workflow. A Ruby script (vcf_to_nexus.rb available at: https://github.com/mmatschiner/tutorials) was used to convert the SNP data to nexus sequence files. In cases where all individuals of each population showed missing data for a locus, the locus was kept as missing data; otherwise, the missing data were edited according to the genotype found in other individuals of the same population. For heterologous loci, IUPAC codes were used (Emerson et al., 2010).

Phylogenomic tree reconstruction

A Maximum Likelihood (ML) gene tree was reconstructed for SNP sequence data with RaxML v. 1.5 (Silvestro & Michalak, 2012). The options for ML

gene tree reconstruction were: ML + rapid bootstrap, 1,000 bootstrap replicates, and GTR-gamma sequence divergence model (as determined by PartitionFinder-2.1.1).

Inference of dataset incongruence

To provide statistical inference of the congruence of mitochondrial DNA (mtDNA)- and nuclear DNA (nDNA)-based tree topologies, an incongruence length difference (ILD) test (Farris et al., 1994) implemented in PAUP* V.4.0a (Swofford, 2002) was performed. To perform the ILD test, a nexus sequence file with two partitions pertaining to nDNA and mtDNA sequences was used (Supporting data I). The options for this test were: number of replicates = 500, optimality criterion = parsimony, gaps = missing data, number of trees held at each step = 1, branch-swapping algorithm = tree-bisection-reconnection (TBR) with a reconnection limit of 8, mulTrees option, no topological constraints, and trees = unrooted, and maxtrees = 1000. In addition, to have a visual comparison of the topology of nDNA and mtDNA maximum likelihood trees, both trees were reconstructed using data belonging to the subset of individuals for which we had both mtDNA and nDNA data.

Analyses of admixture and introgression

To determine the genomic cluster specific to each species/population, a Bayesian clustering analysis was performed on the SNP data using ADMIXTURE V.1.23 (Alexander et al., 2009). The options for admixture analysis were: bootstrap = 1,000 and K = 1–5. The K value (number of admixture clusters) was selected based on the 10-fold cross-validation error (CV), *i.e.*, we selected the K corresponding with the lowest cross-validation errors as the best-supported number of clusters. In addition, a Principal Components Analysis (PCA) was performed on the SNP data with the adegenet R package (Jombart & Collins, 2015). To reduce the effect of missing data on PCA analysis, the *randomForest* R package (Breiman, 2006) was used to impute missing data based on the existing population/species groups. To further assess introgression among *R. lacustris populations*, *D*-statistics were calculated using the program Dsuite (Malinsky, 2019), with 200-bp blocks of SNPs. *Rutilus frisii* was used as the out-group to identify ancestral SNPs. *D*-statistics for trios of populations/species (P1, P2, and P3) vary between 0 and 1. If $D = 0$, the test provides evidence of no introgression, but if $D = \pm 1$, it signals introgression between P2 and P3 (if positive); otherwise, if negative, it indicates introgression between P1 and P3. Further, we calculated between-population/species F_{ST} distances using the STAMPP R package (Pembleton et al., 2013).

Molecular species/population delimitation

To test our inferences on groups clarified in PCA and admixture analyses and based on phylogenetic tree reconstruction, a Bayes Factor Delimitation analysis (BFD) was performed on SNP data (Leaché et al., 2014). In this approach, candidate species/population delimitation scenarios are compared and ranked according to the marginal likelihood estimates (MLE). The data were analysed using the SNAPP plug-in implemented in BEAST 2 (Bouckaert et al., 2014). To estimate the marginal likelihoods, path sampling was performed for 56 steps, with a Markov Chain Mont Carlo (MCMC) chain of 100,000 with pre-burnin of 10,000 steps for path sampling. XML files were set up following the instructions in the BFD tutorial (Leaché & Bouckaert, 2018). In addition to using sampling localities (for populations of *R. lacustris*) and morphological differences between *R. lacustris* and *R. frisii*, we also considered clustering patterns resolved in PCA and phylogenetic trees to set up different species/population delimitation scenarios. Species/population delimitation models were compared based on the Bayes factor (BF) (Kass & Raftery, 1995). To calculate BF statistics, the equation following equation was used:

$$BF = 2 \text{ X } (MLE_{\text{Mc}} - MLE_{\text{Ma}})$$

where MLE_{Mc} and MLE_{Ma} are the marginal likelihood estimates ($\log_e$) of the species/population delimitation scenario that we considered based on geographic distribution and clustering patterns of the genetic data resolved in clustering analyses and the alternative models, respectively. A positive BF statistic supports the first population/species delimitation model (*Mc* here), and a negative BF statistic supports the alternative population/species delimitation model (*Ma* here). The tested models are presented in Table 2.

Table 2 Models tested using the Bayes factor delimitation (BFD) approach

Model	Species/population combinations				MLE[a]	BF[b]	Rank
M4	*R. frisii* (I)	*R. lacustris* Aras (II)	*R. lacustris* southwest Cas. (III)	*R. lacustris* southeast Cas. (IV)	− 59179.38	–	1
M3_1	*R. frisii* (I)	II + III	–	*R. lacustris* southeast Cas. (IV)	− 60310.12	2261.48	4
M3_2	*R. frisii* (I)	II + IV	*R. lacustris* southwest Cas. (III)	–	− 60278.74	2198.72	3
M3_3	*R. frisii* (I)	*R. lacustris* Aras (II)	III + IV	–	− 59300.48	242.2	2
M2	*R. frisii* (I)	II + III + IV	–	–	− 60884.57	3410.38	5

[a]Marginal likelihood estimate

[b]Bayes Factor; Cas.: Caspian Sea

Migration rates

To determine contemporary migration rates among the studied populations, a Bayesian likelihood method using BayesAss3-SNPs (Wilson & Rannala, 2003; Mussmann et al., 2019) was performed on all SNP data (5,510 SNPs) for 7,000,000 generations with a burnin value of 700,000 The optimal mixing parameters were $-a = 0.7750$, $-f = 0.0875$, and $-m = 0.3250$, which were determined using BA3-Autotune (Mussmann et al., 2019).

Results

Mitochondrial DNA

The haplotype diversity (H_d) and nucleotide diversity (P_i) in *R. frisii* (number of individuals (*N*): 10; number of haplotypes (nH): 2; H_d: 0.533; P_i: 0.001), were higher than those for *R. lacustris* from the southwestern Caspian Sea (*N*: 6; nH: 2; H_d: 0.333; P_i: 0.000), *R. lacustris* in the Aras River (Kura drainage) (*N*: 10; nH: 2; H_d: 0.2; P_i: 0.000), and *R. lacustris* from the southeast Caspian Sea (*N*: 8; nH: 2; H_d: 0.25; P_i: 0.000). Among *R. lacustris* populations and *R. frisii* specimens that were analysed, eight *COI* haplotypes were defined. One of the haplotypes in *R. lacustris* from the southwest Caspian Sea was actually from *A. brama*. There were also *R. frisii* haplotypes detected in two *R. lacustris* individuals from the southeast Caspian Sea (Fig. 2). *Rutilus lacustris* from the southeast and southwest Caspian Sea possessed shared haplotypes, but *R. lacustris* from the Aras River (Kura drainage) had a different frequent haplotype differing by 2–3 mutations from *R. lacustris* haplotypes in the southeast and southwest Caspian Sea. In addition, *R. lacustris* from the south Caspian Sea had a haplotype also reported in the lower Volga River (GenBank: MK791226.1).

Within the haplotype network reconstructed for *Cyt*-b sequences from GenBank, a few *R. lacustris* haplotypes observed in the south Caspian Sea also were observed in localities up to the lower Volga River (Supplementary Figure SI). Haplotypes from the Aras River drainage close to our sampling locality were placed among haplotypes for *R. lacustris* from the south, west, and a few from the north Caspian Sea. The Aras River *Cyt*-b haplotype was closely related to that for *R. schelkovnikovi* from a tributary of the Aras River, the Mezamor River. On the other hand, haplotypes from the Kura River were closely related to haplotypes from the west and north Caspian Sea, but not from the south.

Genomic analyses

Three of the 37 individuals analysed did not have enough reads (less than 2,000,000 reads) for reliable sequence coverage, and those data were excluded from further analyses. A total of 96,970 SNPs distributed across 50,712 loci existed in the input file. After filtering-out 88,423 SNPs (91.2%), 8,547 SNPs were retained. To avoid problems related to linkage in the

final VCF file which we used for further analyses, we kept only unlinked SNPs (7,984 SNPs).

Phylogeny

The ML and Bayesian phylogenetic trees reconstructed for *COI* sequences were similar in topology; hence, only the Bayesian phylogenetic tree with bootstrap support values for the ML tree is presented (Fig. 3). Sequences for all *Rutilus* spp. collected in this

Fig. 3 Bayesian phylogenetic tree reconstructed for a 595-bp ▶ *COI* 5′-end sequence shared among species from GenBank and our own sequences. The numbers beside tree branches are the Bayesian posterior probabilities (before slash) and the Maximum likelihood (ML) bootstrap support values (after slash) calculated using 1000 permutations. The colored boxes denote the respective populations/species considered here. A color key is presented on the upper left side of the phylogram. Sequences from GenBank are presented with their scientific names

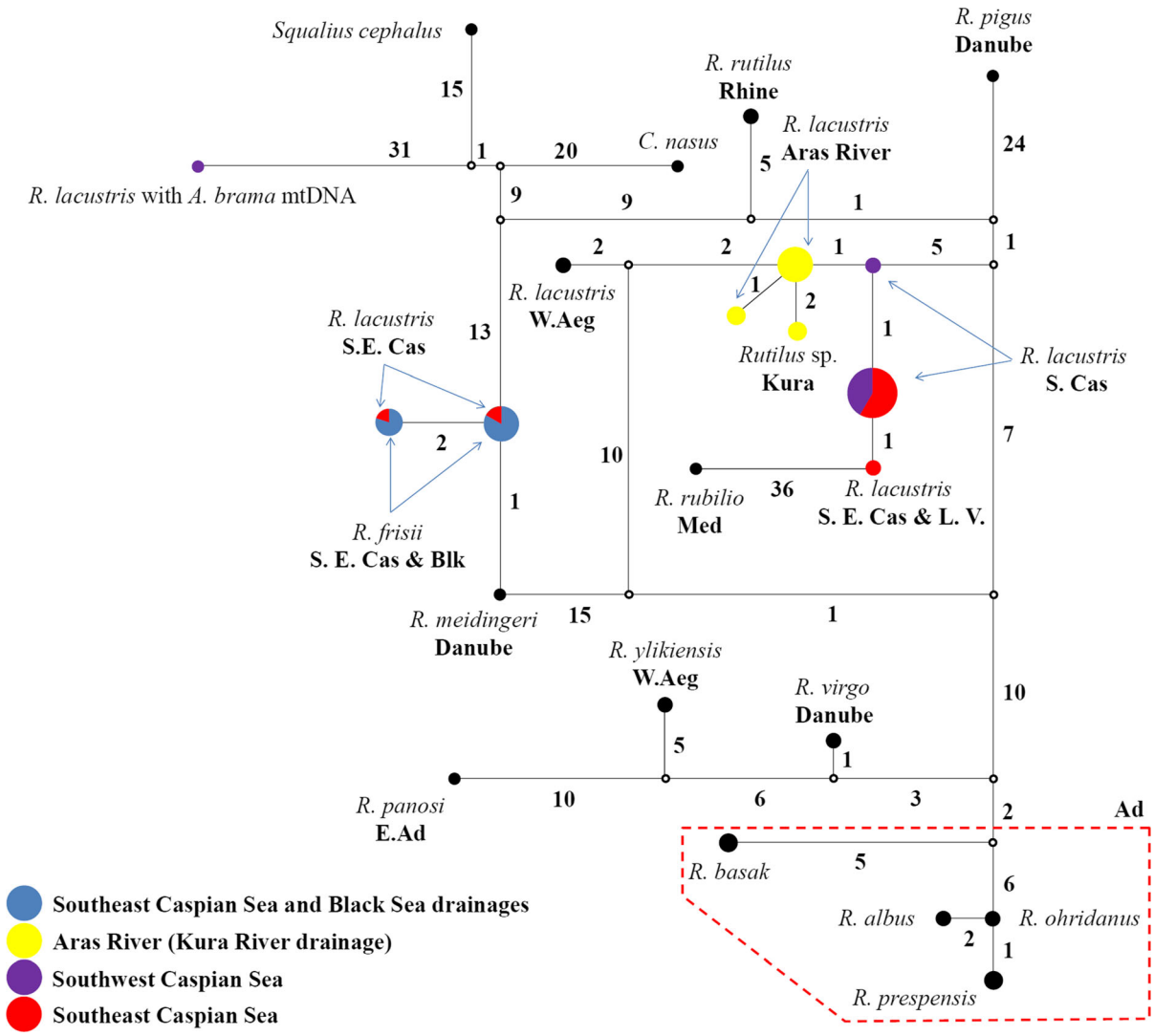
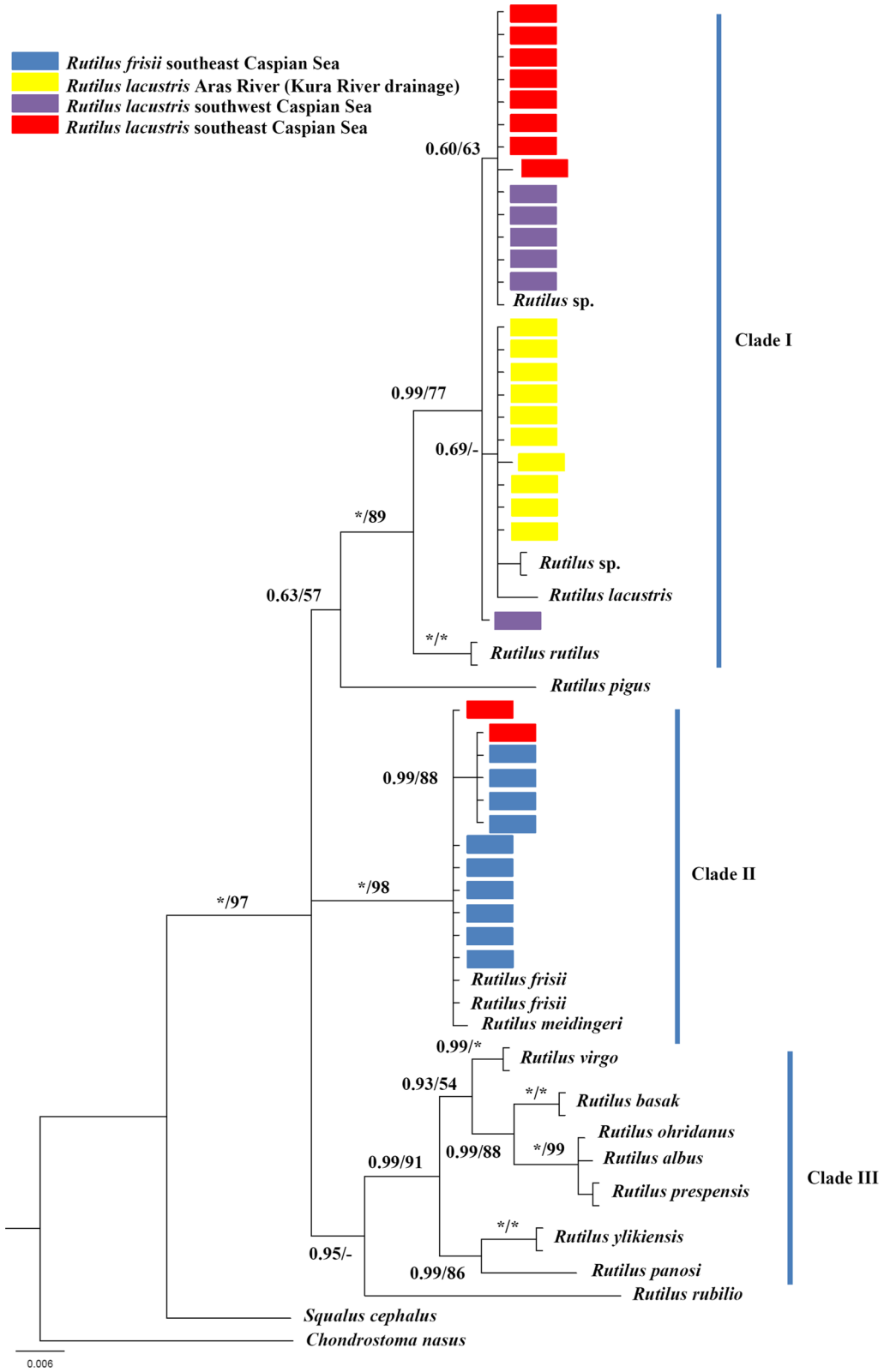


Fig. 2 Haplotype network reconstructed for a 595-bp 5′-end sequence of *COI* shared among species from GenBank and our own sequences. The sizes of the pie diagrams (for haplotypes observed in this study) reflect the relative frequencies of the respective haplotypes. The colors of the slices in the pie diagrams denote the relative haplotype frequencies for each species or geographic population. Species or populations to which the colors correspond are denoted by a scientific name and the sampling locality. The hatched lines denote the number of mutational differences between each pair of haplotypes. The black-filled circles with no names beside them are inferred haplotypes not observed in this study


Rutilus frisii southeast Caspian Sea
Rutilus lacustris Aras River (Kura River drainage)
Rutilus lacustris southwest Caspian Sea
Rutilus lacustris southeast Caspian Sea
0.60/63
Rutilus sp.
Clade I
0.99/77
0.69/-
*/89
Rutilus sp.
Rutilus lacustris
0.63/57
/
Rutilus rutilus
Rutilus pigus
0.99/88
Clade II
*/98
*/97
Rutilus frisii
Rutilus frisii
Rutilus meidingeri
0.99/*
Rutilus virgo
0.93/54
/
Rutilus basak
Rutilus ohridanus
0.99/88
*/99
0.99/91
Rutilus albus
Rutilus prespensis
Clade III
/
Rutilus ylikiensis
0.95/-
0.99/86
Rutilus panosi
Rutilus rubilio
Squalus cephalus
Chondrostoma nasus
0.006

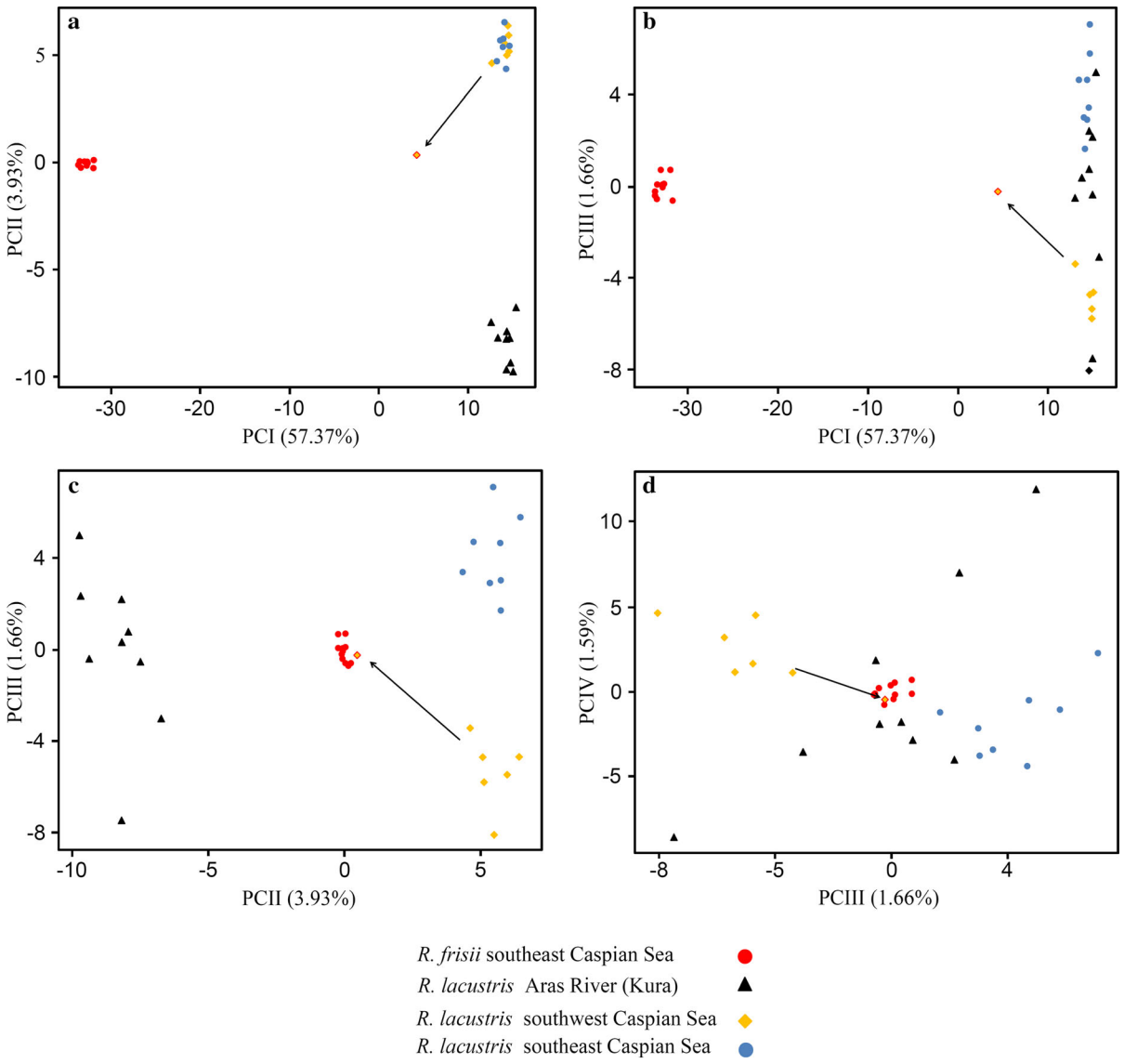


Fig. 4 Distribution of different populations/species in PCA graphs along the first four PC axes. The arrow points to the *R. lacustris* individual with *COI* haplotype belonging to *A. brama*

study or retrieved from GenBank were monophyletic (bootstrapping value (*BS*) = 100%; Bayesian Posterior probability (*BP*) = 0.98; Fig. 3). Among *Rutilus* spp. considered here, three clades—including clade I (*R. lacustris*, *R. rutilus*, and *R.* sp.; *BS* = 90%; *BP* = 1.0), clade II (*R. frisii* and *R. meidingeri*; *BS* = 99%; *BP*: 1.0), and clade III (*R. ylikiensis*, *R. panosi*, *R. virgo*, *R. basak*, *R. prespensis*, *R. ohridanus*, *R. albus*; *BS* = 90%; *BP* = 0.99)—were identified. In addition to the noted clades, there were a few species, including *R. rubilio* and *R. pigus* that did not nest within any of these three clades with acceptable support values; we considered ML bootstrap support values over 70% as strong.

In both genomic and mtDNA phylogenetic trees reconstructed for individuals used in this study, populations/species nested in separate groups (Supplementary Figure SII). In both trees, *R. lacustris* from the southeast and southwest Caspian Sea and the Aras River (Kura drainage) nested within a monophyletic clade with absolute bootstrap support (*BS* = 100). *Rutilus lacustris* from the sowtheast and southwest

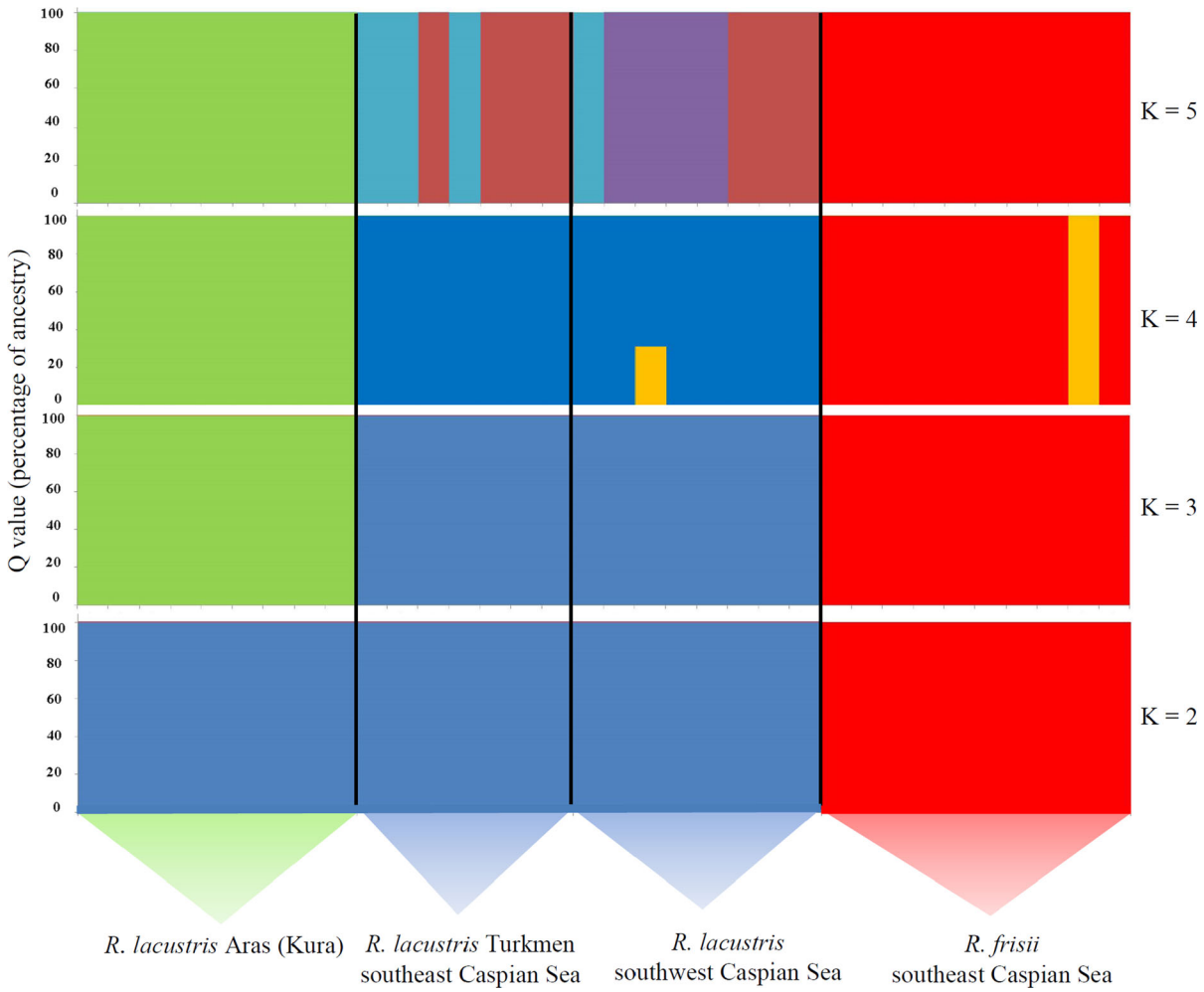


Fig. 5 Admixture graphs showing genomic clusters and percentage of ancestry (Q values) for different values of K (numbers of genomic clusters) from 2 to 5. The respective populations/species are indicated below the figure

Caspian Sea nested within a monophyletic sub-clade in both trees, but the bootstrap support for this sub-clade in the genomic tree (65%) was not as high as that in the mtDNA tree (96%), although this met the 50% majority-rule criterion. In both trees, no clear intra-clade relationship was resolved for *R. lacustris* from the Aras River relative to *R. lacustris* from the southeast and southwest Caspian Sea. A reciprocal incongruency was observed in the positions of three individuals within the genomic and mtDNA phylogenetic trees: all these individuals nested within the *R. lacustris* genomic clade, but nested in *R. frisii* or *A. brama* mtDNA clades (Fig. 2 and Supplementary Figure SII). This results suggests that these individuals may be backcrosses resulting from past interspecific hybridization events. The incongruence length difference (ILD) test result also was significant ($P < 0.005$), implying discordance between the nDNA and mtDNA tree topologies.

Population genomics

The first four principal components (PCs) explained 64.55% of the total variability in the genomic data. In the graphs produced for the first three PCs, three clearly separate clusters including: a) *R. frisii*, b) *R. lacustris* (from the west and east Caspian Sea in a shared cluster), and c) *R. lacustris* from the Aras River, were resolved (Fig. 4). The clusters of *R. frisii* and *R. lacustris* were clearly separated along the first

(57.37% of total variation) and second (3.93% of variation) PC axes (Fig. 4). The clusters belonging to *R. lacustris* were clearly separated along the second PC axis. Although the east and west Caspian Sea populations of *R. lacustris* were not discriminated along the first two PCs, they were separated along the third PC axis (1.66% of variation). The *R. lacustris* individual with the mtDNA haplotype belonging to *A. brama* was not included in any of the resolved clusters along the first two PCs, but its position was closer to *R. lacustris* clusters.

Based on the admixture program cross-validation errors (CV), the most probable number of genomic clusters among the studied populations/species was $K = 2$ (CV = 0.33). The next most probable number of clusters was $K = 3$ (CV = 0.34). With $K = 2$, all *R. lacustris* individuals appeared to be members of a common admixture cluster different from *R. frisii,* but with $K = 3$, a third admixture cluster belonging to *R. lacustris* from the Aras River was resolved. Further increasing the K value over 3, the hybrid status for the individual with mtDNA from *A. brama* was clarified (Fig. 5). Also, the *D*-statistic was not significant ($D = 0.006$; $P > 0.05$), indicating the occurrence of gene flow between *R. lacustris* populations in the southeast and southwest Caspian Sea, but not between the Caspian Sea populations and the Aras River population. The pairwise genetic differentiation (F_{ST}) estimates between *R. lacustris* populations and *R. frisii* were very high and varied between 0.70 and 0.78. The same distance between the southeast and southwest Caspian Sea *R. lacustris* populations was 0.03, while it was around 0.12 between *R. lacustris* from the Aras River and each of its conspecific southeast and southwest Caspian Sea populations.

Bayes factor delimitation test

Among the five models tested using the Bayes factor delimitation test, the model in which three populations of *R. lacustris* were assumed—respectively in the southeast and southwest Caspian Sea and the Aras River—was supported robustly based on *BF* values calculated in path sampling (Table 2). The second best-supported model (rank 2) of populations of *R. lacustris* was the model in which the southeast and southwest Caspian Sea populations were considered as one population.

Migration rates

Migration rates estimated between each pair of populations were mostly negligible (Supplementary Table V). Among the individuals analysed in the southwest Caspian Sea 0.2421 ± 0.0422 individuals were identified to be recent migrants from southeast Caspian Sea but there was no migrants migrating from southwest to southeast Caspian Sea population of *R. lacustris*.

Discussion

Population genetics

Based on our results, *R. lacustris* populations from the southeastern and southwestern Caspian Sea are closely related and show low levels of divergence. This finding is in agreement with those of Reyhani et al. (2010) using microsatellite markers and Rezvani et al. (2006) using mitochondrial DNA markers, but contradictory to those of Keyvanshokooh et al. (2007) using microsatellite markers. Reyhani et al. (2010) found that genotype frequencies at a majority of the microsatellite loci they had used to compare the southeastern and southwestern Caspian Sea populations of *R. lacustris* were not in Hardy–Weinberg equilibrium, which could be a signature of migration and population mixing between the southwestern and southeastern Caspian Sea populations or segregation of null alleles. Observation of haplotypes shared among both localities, as well as most of our data analyses—including *D*-statistic, admixture and PCA clustering patterns, and F_{ST} distances—indicate that the two *R. lacustris* sample sets from the southeastern and southwestern Caspian Sea belong to populations that are connected by fairly high levels of gene flow. Also, a non-significant level of divergence was previously reported between *R. frisii* populations from the southern and northern Caspian Sea (Kotlik et al., 2008), which also suggests pronounced gene flow between populations in both of these *Rutilus* spp. The similarity between the populations from the southeastern and southwestern Caspian Sea suggests that imprinting behavior (adults returning to natal rivers for spawning) in *R. lacustris* is not as strong as that reported in other fishes such as salmonids. However, results from the BFD test support the view that the

southeastern and southwestern *R. lacustris* are two distinct populations, which may therefore be considered as distinct management units (sensu Moritz, 1994).

All our analyses support *R. lacustris* in the Aras River as a separate population, a finding which is contradictory to results of an osteological comparison of *R. lacustris* from the same region. Eagderi et al. (2017) concluded that the southwestern Caspian and the Aras River populations of *R. lacustris* were similar and that the southeastern Caspian population was probably a distinct lineage. In our results, however, the Aras River population differed markedly from the populations collected from the southeastern and southwestern Caspian Sea in having a separate haplotype and comprising a clearly distinct genomic PCA cluster. We note that the specimens that we collected from the Aras River all came from the reach upstream of the Aras Dam (at the border of Iran and Azerbaijan (Nakhchivan)) that has since 1971 interrupted connectivity between the downstream reaches of the river and the mainstem Kura River that drains into the southwestern Caspian Sea. The genomic differentiation of *R. lacustris* in the Aras River may in part be attributable to the Aras Dam, but as the haplotype found in the Aras *R. lacustris* is different by 2–3 bp from those of the southeastern and southwestern Caspian Sea populations considered here, the results may indicate the existence of a distinct river-resident population, a life history which has been documented for roaches (Levin et al., 2017). We could not obtain and genotype specimens from unblocked sections of the Kura River drainage, so any definitive inference about their phylogenetic status based on our data must be made cautiously. However, the data from GenBank also showed that *Cyt*-b haplotypes in the Kura River are different from the haplotypes reported from the Aras River (Levin et al., 2017). Also, the haplotypes in the Aras River are closely related to haplotypes mostly from the southern and western Caspian Sea, and haplotypes from the Kura River are close to haplotypes from the western and northern Caspian Sea.

Hybridization

Our results showed signatures of introgressive hybridization between *R. lacustris* and *R. frisii,* and in one case between *R. lacustris* and *A. brama*. Hybridization among the members of cyprinid subfamily Leuciscinae, especially between *A. brama* and *R. rutilus*, has been known since the late eighteenth century and is an intergeneric model system for studying hybridization in fishes (Nazu Matondo et al., 2010; Konopiński & Amirowicz, 2018). Hybridization between *R. lacustris* and *R. frisii* also has been reported in previous studies (Larmuseau et al., 2009; Levin et al., 2017). In all cases of hybridization, mtDNA from *R. frisii* and *A. brama* was observed in individuals with genomic clusters and morphological attributes of *R. lacustris*, suggesting backcrossing of hybrids into *R. lacustris*. Similar observations have been reported in previous studies of roaches in other parts of their natural distribution (Wyatt et al., 2006; Hayden et al., 2010; Toscano et al., 2010; Kuparinen et al., 2014). There, hybridization was shown to occur between female *A. brama* and male *R. rutilus*, with low levels of backcrossing (Hayden et al., 2010; Toscano et al., 2010). Our results appear to be in agreement with the reported direction of hybridization (Kuparinen et al., 2014), except that all hybrid individuals detected in our study were probably old backcrosses, since none contained detectable nDNA clusters from the mtDNA donor species. High levels of mito-nuclear discordance also have been reported in other fish species, which probably is caused by loss of species-specific mating signals and non-assortative mating (Krück et al., 2013; Geiger et al., 2016; Sousa-Santos et al., 2018; Feng et al., 2019). The persistence of advanced-generation hybrid individuals suggests some level of reproductive compatibility between *A. brama* and *R. lacustris*, as well as survival of their hybrid progeny. However, a better inference on the success of hybrid individuals will obviously require data from larger sample sizes, given that we found a single *R. lacustris* (male) X *A. brama* (female) hybrid, while others were all *R. frisii* (female) X *R. lacustris* (male). The hybrids that we observed were probably natural, since in the Caspian Sea all the noted species are sympatric and, in some cases, syntopic (Abdoli, 2000).

Guidelines for conservation

Our study showed that the southern Caspian Sea populations of *R. lacustris*, while being connected by gene flow, may be considered as isolated populations (BFD results) with a genetic differentiation (F_{ST}) of 0.03. This genetic divergence, along with

a ~ 400 km geographic distance between them, if also existing in populations falling between the studied localities, can be interpreted as partial isolation of these populations and no panmixia. However, these two populations share mtDNA haplotypes and cannot be considered as different intra-specific lineages. The proportion of migrant individuals from southeast Caspian Sea in southwest Caspian Sea population is higher than the10% threshold proposed for demographic independence (see Palsbøll et al., 2007). Although, this migration rate is higher than the threshold proposed for demographic independence of populations, the migration is unidirectional which may be a result of small sample sizes used here or unreported fish transfer from southeast to southwest Caspian Sea. The genetic distance between southeast and southwest Caspian Sea populations is in line with the genetic differentiation-based criteria, but migration rate is contradictory to demographic independence criteria proposed for determination of management units (Moritz, 1994; Paetkau, 1999; Palsbøll et al., 2007). We propose the southeastern and southwestern Caspian Sea *R. lacustris* populations to be considered as two potentially separate management units. Anyway, as we did not include any specimen from sites located between the ones analysed here, and as the observed genetic differentiation is not as high as 0.06–0.72 reported for other migratory fish species (*S. caspius*) in the southern Caspian Sea basin (Vera et al., 2011; Hashemzadeh Segherloo et al., 2012; Tabatabaei et al., 2020b), this conclusion on management units of *R. lacustris* in the southern Caspian Sea should be further investigated using specimens from intermediate localities, since the identified pattern of genetic differentiation may simply be the ends of a population genetic continuum.

The *R. lacustris* population in the Aras River of the Kura River drainage was isolated from the Caspian Sea *R. lacustris* in nDNA-based analyses and its mtDNA haplotypes were different (2–3 bp). In addition to its realtively high level of differentiation (F_{ST}= 0.12), its separate population status was highly supported in BFD analysis. Furthermore, mtDNA haplotypes of the Aras River *R. lacustris* were not observed in the Caspian Sea samples in our study or in those of other studies comparing this population to southern Caspian Sea *R. lacustris* (Sharifi et al., 2016). Regarding low haplotype and nucleotide diversity in the Aras River population, this population may be a limnetic population that colonized the Aras Reservoir from upstream tributaries hosting similar populations, since the Aras River dam has blocked upstream migrations from downstream riches. Such colonization by *R. lacustris* of a recently constructed reservoir (2008 to 2010)—downstream of the Aras River dam (1971) has been reported (Levin & Roubenyan, 2012). We suggest that this population can be a limnetic form of *R. lacustris*, probably close to the *R. schelkovnikovi* inhabiting a tributary of the Aras River (Levin et al., 2017). These attributes of Aras River *R. lacustris* could reflect some level of adaptation to special habitats (river vs. sea habitat in this case), considerable reproductive isolation, significant divergence in allele frequencies, restricted gene flow, and local adaptation (Moritz, 1994; Fraser & Bernatchez, 2001; Robertson et al., 2014; Escobar et al., 2015). As such, these characteristics would justify considering the *R. lacustris* population and other populations/species inhabiting the upstream riches in the Aras River drainage as distinct conservation units. Hence, we propose the Aras River drainage population/s of the *R. lacustris* be considered as a separate unit, and that in future conservation or management plans, transfer of *R. lacustris* to the Aras River drainage from other localities including the southern Caspian Sea should be avoided.

Acknowledgements We thank G. Cote of the Bernatchez laboratory (Laval University) for his kind assistance during laboratory work, Gaétant Légaré, Sonia Pomerleau and Jérôme St-Cyr from the IBIS genomic analyses platform (Laval University, Québec City, Canada, http://www.ibis.ulaval.ca) for their assistance in library preparation and genotyping. The genomic data processing was performed using the KATAK server at IBIS and part of the data analysis and model checking was performed at the Shahr-e-Kord University Fast processing center (www.sku.ac.ir). We are also grateful to Associate editor C. Sturmbauer and two anonymous reviewers for their constructive comments on an earlier version of the manuscript. This work was supported by a NSERC (Canada) Discovery grant (http://www.nserc-crsng.gc.ca) to Louis Bernatchez; grant number 688MIGRD94 to Iraj Hashemzadeh Segherloo by Shahr-e-Kord University (www.sku.ac.ir); and a short term scholarship (V3 program) from the Fonds de Recherche Québécois sur la Nature et les Technologies (FRQNT: www.frqnt.gouv.qc.ca) to Iraj Hashemzadeh Segherloo; and a grant to F. Ghojoghi by Islamic Azad University, Azadshahr branch.

Data availability Demultiplexed DNA sequences are available at the SRA database (SRA Bioproject: PRJNA646736). For more data availability details, please see the text.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdoli, A., 2000. The Inland Water Fishes of Iran. Iranian Museum of Nature and Wildlife, Tehran. 378 p. In Persian.

Abed, A., G. Légaré, S. Pomerleau, J. St-Cyr, B. Boyle & F. J. Belzile, 2019. Genotyping-by-sequencing on the ion torrent platform in barley. Barley. Humana Press, New York, NY: 233–252.

Alda, F., V. A. Tagliacollo, M. J. Bernt, B. T. Waltz, W. B. Ludt, B. C. Faircloth, M. E. Alfaro, J. S. Albert & P. Chakrabarty, 2018. Resolving deep nodes in an ancient radiation of Neotropical fishes in the presence of conflicting signals from incomplete lineage sorting. Systematic Biology 68: 573–593.

Alexander, D. H., J. Novembre & K. Lange, 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19: 1655–1664.

Aljanabi, S. M. & I. Martinez, 1997. Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25: 4692–4693.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics, Babraham Institute, Cambridge.

Benestan, L., T. Gosselin, C. Perrier, B. Sainte-Marie, R. Rochette & L. Bernatchez, 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). Molecular Ecology 24: 3299–3315.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut & A. J. Drummond, 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537.

Breiman, L., 2006. randomForest: Breiman and Cutler's random forests for classification and regression. http://stat-www.berkeley.edu/users/breiman/RandomForests, R package version.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores & W. A. Cresko, 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124–3140.

Clement, M., D. Posada & K. A. Crandall, 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9: 1657–1659.

Eagderi, S., F. Ghojoghi & M. Nasri, 2017. Phylogenetic relationship of members of the Genus *Rutilus* Rafinesque, 1820 in the southern Caspian Sea (Bandar Turkman, and Shalman and Aras rivers) based on osteological characters. Journal of Fisheries (Iranian Journal of Natural Resources) 70: 179–188. **(in Persian with English abstract)**.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797.

Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw & C. M. Holzapfel, 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences USA 107: 16196–16200.

Escobar, L. M. D., J. Andrade-López, I. P. Farias & T. Hrbek, 2015. Delimiting evolutionarily significant units of the fish, *Piaractus brachypomus* (Characiformes: Serrasalmidae), from the Orinoco and Amazon river basins with insight on routes of historical connectivity. Journal of Heredity 106: 428–438.

Farris, J. S., M. Källersjö, A. G. Kluge & C. Bult, 1994. Testing significance of incongruence. Cladistics 10: 315–319.

Feng, C., Y. Tang, S. Liu, F. Tian, C. Zhang & K. Zhao, 2019. Multiple convergent events created a nominal widespread species: *Triplophysa stoliczkae* (Steindachner, 1866) (Cobitoidea: Nemacheilidae). BMC Evolutionary Biology 19(1): 177.

Fraser, D. J. & L. Bernatchez, 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Molecular Ecology 10: 2741–2752.

Geiger, M. F., C. Schreiner, G. B. Delmastro & F. Herder, 2016. Combining geometric morphometrics with molecular genetics to investigate a putative hybrid complex: a case study with barbels *Barbus* spp. (Teleostei: Cyprinidae). Journal of Fish Biology 88: 1038–1055.

Hall, T., 2001. BioEdit version 5.0. 6. Department of Microbiology, North Carolina State University, Raleigh, North Carolina. https://bioedit.software.informer.com/5.0/.

Hashemzadeh Segherloo, I., H. Farahmand, A. Abdoli, L. Bernatchez, C. R. Primmer, A. Swatdipong, M. Karami & B. Khalili, 2012. Phylogenetic status of brown trout *Salmo trutta* populations in five rivers from the southern Caspian Sea and two inland lake basins, Iran: a morphogenetic approach. Journal of Fish Fiology 81: 1479–1500.

Hashemzadeh Segherloo, I., E. Normandeau, L. Benestan, C. Rougeux, G. Coté, J.-S. Moore, N. Ghaedrahmati, A. Abdoli & L. Bernatchez, 2018. Genetic and morphological support for possible sympatric origin of fish from subterranean habitats. Scientific Reports 8(1): 2909.

Hayden, B., D. Pulcini, M. Kelly-Quinn, M. O'Grady, J. Caffrey, A. McGrath & S. Mariani, 2010. Hybridisation between two cyprinid fishes in a novel habitat: genetics, morphology and life-history traits. BMC Evolutionary Biology 10(1): 169.

Ivanova, N. V., T. S. Zemlak, R. H. Hanner & P. D. Hebert, 2007. Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes 7: 544–548.

Jombart, T. & C. Collins, 2015. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. MRC Centre for Outbreak Analysis and Modelling, Imperial College, London. http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf.

Kass, R. E. & A. E. Raftery, 1995. Bayes factors. Journal of the American Statistical Association 90: 773–795.

Keyvanshokooh, S. & M. R. Kalbassi, 2006. Genetic variation of *Rutilus rutilus caspicus* (Jakowlew 1870) populations in Iran based on random amplified polymorphic DNA markers: a preliminary study. Aquaculture Research 37: 1437–1440.

Keyvanshokooh, S., A. Ghasemi, M. Shahriari-Moghadam, R. M. Nazari & M. Rahimpour, 2007. Genetic analysis of *Rutilus rutilus caspicus* (Jakowlew 1870) populations in Iran by microsatellite markers. Aquaculture Research 38: 953–956.

Kiabi, B. H., A. Abdoli & M. Naderi, 1999. Status of the fish fauna in the South Caspian Basin of Iran. Zoology in the Middle East 18: 57–65.

Konopiński, M. K. & A. Amirowicz, 2018. Genetic composition of a population of natural common bream *Abramis brama* × roach *Rutilus rutilus* hybrids and their morphological characteristics in comparison with parent species. Journal of Fish Biology 92: 365–385.

Kotlik, P., S. Markov, A. L. Choleva, N. G. Bogutskaya, F. G. Ekmek & P. P. Ivanova, 2008. Divergence with gene flow between Ponto-Caspian refugia in an anadromous cyprinid Rutilus frisii revealed by multiple gene phylogeography. Molecular Ecology 17: 1076–1088.

Kottelat, M. & J. R. Freyhof, 2007. Handbook of European Freshwater Fishes. Publications Kottelat, Cornol.

Krück, N. C., I. R. Tibbetts, R. D. Ward, J. W. Johnson, W. K. Loh & J. R. Ovenden, 2013. Multi-gene barcoding to discriminate sibling species within a morphologically difficult fish genus (*Sillago*). Fisheries Research 143: 39–46.

Kumar, S., G. Stecher & K. Tamura, 2016. MEGA7: Molecular Evolutionary Genetics Analysis, version 7.0 for bigger datasets. Molecular Biology and Evolution 30: 2725–2729.

Kuparinen, A., M. Vinni, A. G. F. Teacher, K. Kähkönen & J. Merilä, 2014. Mechanism of hybridization between bream *Abramis brama* and roach *Rutilus rutilus* in their native range. Journal of Fish Biology 84: 237–242.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld & B. Calcott, 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34: 772–773.

Larmuseau, M. H. D., J. Freyhof, F. A. M. Volckaert & J. K. J. Van Houdt, 2009. Matrilinear phylogeography and demographical patterns of *Rutilus rutilus*: implications for taxonomy and conservation. Journal of Fish Biology 75: 332–353.

Leaché, A. D. & R. R. Bouckaert, 2018. Species trees and species delimitation with SNAPP: a tutorial and worked example. In: Workshop on Population and Speciation Genomics, Český Krumlov, Czech Republic.

Leaché, A. D., M. K. Fujita, V. N. Minin & R. R. Bouckaert, 2014. Species delimitation using genome-wide SNP data. Systematic Biology 63: 534–542.

Levin, B. A. & H. R. Roubenyan, 2012. New ichthyological findings in Armenia. Russian Journal of Biological Invasions 3: 105–109.

Levin, B., E. Simonov, O. Ermakov, M. A. Levina, E. A. Interesova, O. M. Kovalchuk, Y. A. Malinina, N. S. Mamilov, N. J. Mustafayev, D. V. Pilin, I. V. Pozdeev, N. I. Prostakov, H. R. Roubenyan, S. V. Titov & D. A. Vekhov, 2017. Phylogeny and phylogeography of the roaches, genus *Rutilus* (Cyprinidae), at the eastern part of its range as inferred from mtDNA analysis. Hydrobiologia 788: 33–46.

MacGuigan, D. J. & T. J. Near, 2018. Phylogenomic signatures of ancient introgression in a rogue lineage of darters (Teleostei: Percidae). Systematic Biology 68: 329–346.

Malinsky, M., 2019. D suite-fast D-statistics and related admixture evidence from VCF files. BioRxiv:634477.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17: 10–12.

Mascher, M., S. Wu, P. S. Amand, N. Stein & J. Poland, 2013. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS ONE 8(10): e76925.

Moritz, C., 1994. Defining 'evolutionarily significant units' for conservation. Trends in Ecology & Evolution 9: 373–375.

Mussmann, S. M., M. R. Douglas, T. K. Chafin & M. E. Douglas, 2019. BA3-SNPs: contemporary migration reconfigured in BayesAss for next-generation sequence data. Methods in Ecology and Evolution 10: 1808–1813.

Nzau Matondo, B., M. Ovidio, P. Poncin & J. C. Philippart, 2010. Eco-ethological characteristics of two natural hybrids of *Abramis brama* (L.) from the River Meuse basin. Environmental Biotechnology 6: 42–52.

Paetkau, D., 1999. Using genetics to identify intraspecific conservation units: a critique of current methods. Conservation Biology 13: 1507–1509.

Palsbøll, P. J., M. Berube & F. W. Allendorf, 2007. Identification of management units using population genetic data. Trends in Ecology & Evolution 22: 11–16.

Pembleton, L. W., N. O. Cogan & J. W. Forster, 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources 13: 946–952.

Reyhani, S., S. R. Gilkolaee, P. G. Mostafavi, H. Fazli, M. Porkazemi & N. Golestani, 2010. Microsatellite markers indicate a different structure among three p Identification of management units using population genetic data opulations of the Caspian Roach, *Rutilus rutilus caspicus* (Jakowlew, 1870), in the Caspian Sea: (Osteichthyes: Cyprinideae). Zoology in the Middle East 50: 67–73.

Rezvani, S., A. Eimanifar, R. Aghili & F. Laloei, 2006. PCR–RFLP analysis of mitochondrial DNA for identification of *Rutilus rutilus caspicus* populations on the southern coast of the Caspian Sea, Iran. Journal of the Marine Biological Association of the United Kingdom 86: 1463–1467.

Robertson, J. M., K. M. Langin, T. S. Sillett, S. A. Morrison, C. K. Ghalambor & W. C. Funk, 2014. Identifying evolutionarily significant units and prioritizing populations for management on islands. Monographs of the Western North American Naturalist 7: 397–411.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard & J. P. Huelsenbeck, 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.

Rozas, J., A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E. Ramos-Onsins & A. Sánchez-Gracia, 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34: 3299–3302.

Sharifi, M., P. Negatkhah Manavi, F. Chakmehdouz Ghasemi & S. Behmanesh, 2016. Population genetic structure of Caspian roach (*Rutilus rutilus caspicus*) using DNA sequencing in South Caspian Sea and Aras Lake. Journal of Aquaculture Development 10: 63–73. **(In Persian with English abstract)**.

Silvestro, D. & I. Michalak, 2012. raxmlGUI: a graphical front-end for RAxML. Organisms Diversity and Evolution 12: 335–337.

Sousa-Santos, C., A. M. Pereira, P. Branco, G. J. Costa, J. M. Santos, M. T. Ferreira, C. S. Lima, I. Doadrio & J. I. Robalo, 2018. Mito-nuclear sequencing is paramount to correctly identify sympatric hybridizing fishes. Acta Ichthyologica et Piscatoria 48: 123–141.

Swofford, D. L., 2002. PAUP: Phylogenetic Analysis using Parsimony (and other methods) 4.0b10. Sinauer, Sunderland, MA.

Tabatabaei, S. N., A. Abdoli, I. Hashemzadeh Segherloo, E. Normandeau, F. Ahmadzadeh, F. Nejat & L. Bernatchez, 2020a. Fine-scale population genetic structure of dndangered Caspian Sea trout, *Salmo caspius*: implications for conservation. Hydrobiologia. https://doi.org/10.1007/s10750-020-04334-7.

Tabatabaei, S. N., A. Abdoli, F. Ahmadzadeh, C. R. Primmer, A. Swatdipong & I. Hashemzadeh Segherloo, 2020b. Mixed stock assessment of lake-run Caspian Sea trout *Salmo caspius* in the Lar National Park, Iran. Fisheries Research 230: 105644.

Templeton, A. R., 2004. Using haplotype trees for phylogeographic and species inference in fish populations. Environmental Biology of Fishes 69: 7–20.

Toscano, B. J., D. Pulcini, B. Hayden, T. Russo, M. Kelly-Quinn & S. Mariani, 2010. An ecomorphological framework for the coexistence of two cyprinid fish and their hybrids in a novel environment. Biological Journal of the Linnaean Society 99: 768–783.

Vera, M., I. Sourinejad, C. Bouza, R. Vilas, A. Pino-Querido, M. R. Kalbassi & P. Martínez, 2011. Phylogeography, genetic structure, and conservation of the endangered Caspian brown trout, *Salmo trutta caspius* (Kessler, 1877), from Iran. Hydrobiologia 664: 51–67.

Wilson, G. A. & B. Rannala, 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177–1191.

Wyatt, P., C. Pitts & R. Butlin, 2006. A molecular approach to detect hybridization between bream *Abramis brama*, roach *Rutlius rutilus* and rudd *Scardinius erythrophthalmus*. Journal of Fish Biology 69: 52–71.