Received: 18 December 2019 | Revised: 25 September 2020 | Accepted: 14 October 2020

DOI: 10.1111/jzs.12438

ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

Hidden diversity—Delimitation of cryptic species and phylogeography of the cyprinid *Garra* species complex in Northern Oman

Sandra Kirchner[1,2] | Helmut Sattmann[3] | Elisabeth Haring[1,2] | Reginald Victor[4] | Luise Kruckenhauser[1,2]

[1]Department of Evolutionary Biology, University of Vienna, Vienna, Austria

[2]Central Research Laboratories, Natural History Museum Vienna, Vienna, Austria

[3]Third Zoological Department, Natural History Museum Vienna, Vienna, Austria

[4]Department of Biology, Sultan Qaboos University, Muscat, Oman

Correspondence
Sandra Kirchner, Evolutionary Biology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria.
Email: sandra.kirchner@nhm-wien.ac.at

Abstract

Organisms inhabiting extreme habitats are exposed to many threats. Consequently, they are especially prone to extinction in changing environments. In this study, the genetic variation and phylogeographic pattern of the cyprinid freshwater fish *Garra barreimiae* Fowler & Steinitz (1956) from Northern Oman was examined and compared with other species of the genus. Based on a comprehensive data set, we applied an integrative approach by combining information on mitochondrial and nuclear microsatellite data as well as morphology, abundance, and distribution. The analyses of mitochondrial (and control region [*CR*], *cytochrome c* oxidase subunit 1 [*COI*], *cytochrome b* [*Cytb*]) DNA sequences and nuclear data resulted in the same genetic groups corresponding to different areas alongside the Hajar mountains with distances comparable to other species of the genus. Although their distribution margins are often in close vicinity, our data suggest that gene flow between groups is rare. Interestingly, morphological investigations could not find clear and well-defined morphological traits for discrimination. Hence, these four newly detected genetic groups are addressed as cryptic species of the *Garra* species complex in Northern Oman. Threats by fragmented distribution are often underestimated, especially for cryptic taxa. Thus, conservational actions in order to preserve the so far undetected new species are urgently needed.

KEYWORDS

Garra barreimiae, Hajar mountains, microsatellites, population genetics

1 | INTRODUCTION

Phylogeographic studies are an important tool to help understanding a taxons' historic background and create knowledge about its genetic variation, distribution, and its limits which is crucial to evaluate biodiversity (Moyle & Leidy, 1992; Zachos & Habel, 2011).

The generally known major causes for biodiversity loss include habitat destruction, pollution, and climate change. Besides, also the

Contributing authors: Helmut Sattmann (helmut.sattmann@nhm-wien.ac.at), Elisabeth Haring (elisabeth.haring@NHM-WIEN.AC.AT), Reginald Victor (regvictor@gmail.com), Luise Kruckenhauser (luise.kruckenhauser@nhm-wien.ac.at)

lack of information on hidden species diversity as well as the non-description of newly found cryptic species has severe consequences as these unrecognized or uncatalogued taxa are not protected, leading further to a major loss of biodiversity (Zachos & Habel, 2011).

With the advance of DNA-based techniques for detection of genetic diversity, it became evident that biodiversity has been grossly underestimated in the past (Adams et al., 2014). Cryptic species have been overlooked by taxonomists for a long time, because conventional assessment was based mainly on morphological traits. Adams et al. (2014) demonstrated that even a well-studied organism like the Australian freshwater fish *Galaxias olidus* Günther, 1866 exhibited increased species diversity much higher than previously assumed after including molecular genetic techniques in the assessment. With this in mind, one can only speculate how much undetected diversity is hidden in taxa that are not that well studied.

In the present study, we exemplify this issue analyzing the genus *Garra* Hamilton, 1822 (Cyprinidae) from the southeastern Arabian Peninsula. According to Durand et al. (2002) the subfamily Cyprininae in Middle East is much more diversified than in the Euro-Mediterranean region. Por (1989) stated that this geographic region was an important center of radiation before the Pliocene orogenesis. Even though scientific positions are discordant whether the Middle East was also a center of speciation, there is broad agreement that it was a biogeographic interchange area of great importance. Hence, the Middle East has potentially a much higher, yet undiscovered species richness than so far assumed.

The landscape of Northern Oman is shaped by high mountain ridges up to 3,000 m asl and countless wadis (ephemeral riverbeds) that channel the water bodies resulting from the scarce precipitation events. The climate is hot and arid, and precipitation is irregular but can result in heavy rainfalls followed by massive floodings. The water masses pouring down during one of such torrential flash floods can have massive impacts and may exceed the average annual precipitation in this area (Al-Shaqsi, 2010; Gunawardhana & Al-Rawas, 2014). These conditions are challenging for freshwater dependant animals and only a few well-adapted organisms like the cyprinid barb *Garra barreimiae* Fowler & Steinitz, 1956 managed to inhabit this region permanently. This species is abundant throughout the Hajar mountains in Northern Oman and the United Arab Emirates (UAE) where it inhabits all kinds of water bodies, and even artificial water drainage systems (aflaj/falaj). Due to the highly structured landscapes and habitats, *G. barreimiae* populations are presumably quite isolated.

Kruckenhauser et al. (2011) were the first to analyze populations of *G. barreimiae* in Northern Oman genetically in order to assess the differentiation between the troglomorphic cave population inhabiting the Al Hoota cave and surface populations. In that study, an unexpected genetic diversity of mitochondrial (mt) lineages, separated by p-distances up to 11% were found. Moreover, due to the position of *Garra rufa* (Heckel, 1943) within *G. barreimiae*, the monophyly of *G. barreimiae* was questioned. Based on the data of Kruckenhauser et al. (2011), the subspecies *Garra barreimiae gallagheri* Krupp, 1988 was later raised to species status by Lyon et al. (2016). Another subspecies *Garra barreimiae shawkahensis* Banister & Clarke, 1977 from the United Arab Emirates had not been investigated genetically so far. Later, Pichler et al. (2018) investigated the morphological variation of the species throughout its distribution range in Oman and found only scarce morphological differentiation among populations.

The presented molecular genetic study followed up on the previous research on this group of taxa and phylogenetic lineages, which can be subsumed under the term *Garra* species complex of Northern Oman. With a comprehensive sample and an extended set of molecular genetic markers, we addressed the following questions: What is the extent of genetic variation within and between populations of the *Garra* species complex in Northern Oman? In particular, we asked whether the described subspecies *G. b. shawkahensis* is genetically differentiated from *Garra barreimiae barreimiae*? Can we detect gene flow between the lineages discovered by Kruckenhauser et al. (2011)? Can genetic groups be distinguished and if so, are such groups concordant with their geographic distribution? Is *G. barreimiae* non-monophyletic in the DNA-based phylogenetic tree as suggested earlier by Kruckenhauser et al. (2011)? Are the genetic results concordant with the previously published morphological data investigated by Pichler et al. (2018)?

In the course of finding answers to the questions posed, we analyzed three mt marker sequences as well as 19 polymorphic nuclear (nc) microsatellite markers in samples from numerous sites covering the whole distribution area of the *Garra* species complex in Northern Oman. Furthermore, genetic divergence to other species of the genus *Garra* was investigated.

The results presented here follow the taxonomy of the *Garra* species complex in Northern Oman as accepted in 2016, when the study had started. Our findings were, together with additional morphological and genetic analyses, the basis for a taxonomic revision of the *Garra* species complex in Northern Oman, which is presented elsewhere (Kirchner et al., 2020). A summary of the revised taxonomy is given at the end of this article.

2 | MATERIALS AND METHODS

2.1 | Study area and sampling

Fish were collected from 56 sampling sites covering the whole distribution area in Oman (Figure 1). Full names of sampling sites are provided in Table S1. Specimens were captured with fishing nets and anesthetized with clove oil and then fixed in 80% ethanol. Specimens and their DNA are stored in the Natural History Museum in Vienna (NHMW; Fish Collection and DNA and Tissue Collection). In addition to the samples already stored at the NHMW, 297 individuals of *G. barreimiae* were analyzed especially for this study.

Altogether, 632 specimens were included in the present study comprising *G. barreimiae* and *Garra gallagheri* as well as representatives of other taxa: *G. rufa* and *Garra sahilia* Krupp, 1983, which have been proven to be closely related to *G. barreimiae* (Hamidan et al., 2014; Lyon et al., 2016). As out-group we included

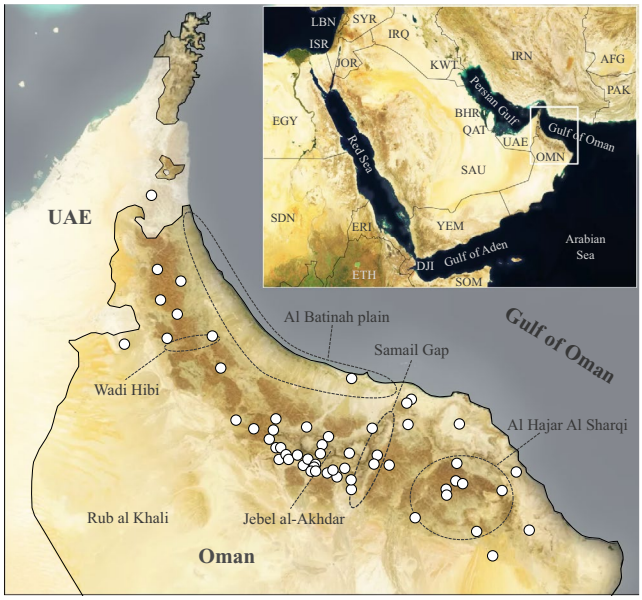

FIGURE 1 Study area and sampling sites. Overview of the study area on the Arabian Peninsula and detailed map of Northern Oman with sampling sites indicated as white dots scattered across the Hajar mountains and relevant geographic areas highlighted with dotted lines

Phreatichthys andruzzii Vinciguerra, 1924, which is close enough but clearly outside the *G. barreimiae* complex according to the results of Calderoni et al. (2016). Earlier studies (Hamidan et al., 2014; Lyon et al., 2016) provided the information that *G. gallagheri* is placed in the phylogenetic analysis within *G. barreimiae*; thus, it was included in our analysis of the intraspecific genetic variation. While 302 individuals were analyzed for the first time in the present study, the remaining individuals were already used in earlier studies. Table S1 provides details on specimens and the markers analyzed for each specimen. For the microsatellite analyses, altogether 524 individuals were analyzed, including also 213 specimens that had been already genotyped in a previous study (Kirchner et al., 2017). For the analyses of the *CR* sequences, 500 individuals were included, among them 117 sequences published by Kruckenhauser et al. (2011) (accession numbers: JF709013–JF709129). From 47 individuals, the *Cytb* and the *COI* genes were analyzed in addition to the *CR*.

2.2 | DNA extraction, PCR amplification, genotyping, and sequencing

Genomic DNA was extracted from tissue of fin clips preserved in 80% ethanol following the QIAGEN DNeasy Blood & Tissue extraction kit protocol. Partial sequences of three mt markers were analyzed: *COI*, *Cytb*, and *CR*. For the *CR*, a ~940-bp long amplicon (*CR*-long) was generated. The sequences were subsequently trimmed in order to include also shorter sequences (*CR*-short) generated in a previous study with the primers Gar-Thr1^{+}/TDK_DGbar1^{-} (Kruckenhauser et al., 2011). Primers and polymerase chain reaction (PCR) conditions used for PCR amplification of the various mt sequences are given in Table S2. DNA amplifications were conducted in 25 µl reaction volume containing 1 µl DNA solution, 5 µl 5× Q-Solution, 2.5 µl 10× PCR Buffer, 0.5 mM of each dNTP, 1.5 mM Mg^{2+}, 0.5 µM of each primer, 0.5 units TopTaq DNA polymerase, 13.9 µl AD for both the *COI* and *Cytb* section, and 1 µl DNA, 2.5 µl 10× PCR buffer, 0.5 mM of each dNTP, 1.5 mM Mg^{2+}, 0.5 µM of each primer, 0.5 units Platinum Taq DNA Polymerase, 18.9 µl AD for *CR*. The cycling protocol for the three different fragments (*COI*/*Cytb*/*CR*) was as follows: initial denaturation at 94°C for 3/2/3 min, 35 cycles denaturation at 94°C for 20/20/15 s, annealing at 58°C for 30 s/52°C for 10 s/65°C for 30 s, extension at 72°C for 60 s/72°C for 90 s/68°C for 60 s, and final extension at 72°C for 7 min/10 min.

The resulting PCR products were purified and sequenced in both directions at Microsynth AG (Balgach, Switzerland) using the PCR primers.

For the microsatellite analyses, samples were profiled with 19 microsatellite loci by using primers that were specifically designed for *G. barreimiae* (Kirchner et al., 2014). The same multiplex protocols have been applied as described in detail in Kirchner et al. (2017). In this manuscript, the term "genotype" is used to describe the microsatellite allele composition characteristic for a population.

2.3 | Alignment, sequence diversity statistics, and phylogenetic reconstructions

The obtained sequences were edited manually using BioEdit v. 7 (Hall, 1999) and aligned with the program MEGA v. 6 (Tamura et al., 2013). The software TrimAl v. 1.3 (Capella-Gutiérrez et al., 2009) implemented in the Phylemon 2.0 web tools was employed, using the "no gap" option for automated removal of gap regions for the final *CR*-long/*CR*-short alignment.

Pairwise p-distances as well as mean p-distances within and between clades were calculated with MEGA. The program DnaSP v. 5 (Librado & Rozas, 2009) was used to estimate nucleotide and haplotype diversity (Hd) as well as to calculate the mismatch distribution (MMD), Tajima' *D* (Tajima, 1989), Fu's *F* (Fu, 1997) as well as Ramos-Onsins' and Rozas' (R2) (Ramos-Onsins & Rozas, 2002) for each clade based on the *CR*-long alignment. The substitution saturation in the *CR*-long fragment have been examined with DAMBE v. 6 (Xia, 2017) (pairwise deletion) and employed the software PartitionFinder v. 2 (Lanfear et al., 2017) to determine the substitution model for the data set of each fragment separately (AICc model selection, search greedy, linked branch lengths).

The nucleotide substitution pattern showed that the sequences have not reached substitution saturation and are therefore well applicable for phylogenetic analyses. The best-fit substitution model for each amplicon and codon position was TVM + G for *CR*, GTR + I for *COI* (codon position 1/2/3 = TRN + G/TRN + I/F81 + I) and TRN + G for *Cytb* (codon position 1/2/3 = GTR+I + G/K80 + I + G/HKY + I). However, since both models TVM and TRN can not be applied in MrBayes, the GTR substitution model was applied instead.

Based on the *CR*-short alignment, haplotype (HT) networks were constructed using the software PopART v. 1.7 (Leigh & Bryant, 2015) and the median-joining (MJ) algorithm proposed by Bandelt et al. (1999) was applied. The neighbor joining (NJ) and maximum likelihood (ML) phylograms were calculated with MEGA and the BI phylograms were calculated with MrBayes v. 3 (Huelsenbeck & Ronquist, 2001). The results were visualized with FigTree v. 1.4.2 (Rambaut, 2007).

To evaluate the mt clades resulting from the phylogenetic tree reconstructions, species delimitation analyses were performed on the *COI* data set using two different approaches: the distance-based method "Automatic Barcode Gap Discovery" (ABGD) and the tree-based method "Bayesian Poisson Tree Process" (bPTP) model. bPTP modeling was used as implemented on the PTP species delimitation web server (https://species.h-its.org/ptp/) (Zhang et al., 2013) based on the unrooted ML tree with 100,000 MCMC generations. For the ABGD method, we applied the Jukes–Cantor model implemented on the ABGD web server (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) (Puillandre et al., 2012).

All HT sequences have been deposited in GenBank (accession numbers MT977717–MT978030 for *CR*-long, MT989476–MT989555 for *CR*-short, MT977670–MT977716 for *CytB*, MT900722–MT900750 for *COI*).

2.4 | Genetic diversity of nc microsatellite data

The nc microsatellite data were explored mostly by employing the statistical computing software environment R version 3.3.1 (R Development Core Team, 2015). After checking the data set for completeness (0.78% missing values in complete data set), basic information was calculated like number of alleles with *R base*, allelic richness with *R diveRsity* (Keenan et al., 2013), private alleles with *R poppr* (Kamvar et al., 2014), expected (H_e) and observed heterozygosity (H_o) per locus as well as the mean H_e and H_o over all loci per population with *R diveRsity*. In addition, the software Genepop v. 4 (Rousset, 2008) was used to test for genotypic linkage disequilibrium and the results showed that some locus pairs exhibit evidence of linkage disequilibrium in single populations, but never in all populations. Deviations from Hardy–Weinberg equilibrium (HWE) were estimated using both the global test and exact test in Genepop executing 10,000 Markov chain Monte Carlo (MCMC) runs for 100 batches, each with 5,000 iterations.

The software Micro-Checker v. 2.2 (Van Oosterhout et al., 2004) was used to check for the existence of microsatellite null alleles. In order to statistically describe the level of inbreeding or differentiation among and within populations, the fixation indices as implemented in the R-package *pegas* version 0.8.233 (Paradis, 2010) was calculated. This function computes F_{IT}, F_{ST}, and inbreeding coefficient (F_{IS}) for each locus in the data set according to the formulae of Weir and Cockerham (1984). Subsequently, the pairwise F_{ST} values between all populations have been calculated with the Microsoft Excel Add-In GenAlEx version 6.5 (Peakall & Smouse, 2012). The F_{IS} after Weir and Cockerham (1984) over all loci per population was calculated with the program FSTAT v. 2.9 (Goudet, 1995).

Allelic richness per locus and population was computed to take differing sampling sizes of populations into account. This was done by using 1,000 re-samples of the smallest sample in the data set (*gallagheri* with $N = 29$) via the *R diveRsity*. Private alleles were estimated with *R poppr* and the percentage of private alleles per population was calculated based on the actual number of alleles per population.

2.5 | Population structure analysis and differentiation

The software STRUCTURE v 2.3 (Pritchard et al., 2000) was used to perform population structure analyses on both the complete microsatellite data set as well as for separate subsets corresponding to the mt clades. The admixture model was employed and tested for clusters from $K = 1$ to $K = 7$ for the complete data set, and 20 iterations per run were performed for each assumed K to find the most reasonable number of clusters (K). Length of burnin was set to 100,000 and the number of MCMC repeats after burnin was 500,000. This analysis was repeated for each cluster (subsets) that resulted from the analysis of the complete data set and tested for as many K as sampling sites were included (Central: $K = 14$, $N = 255$; *gallagheri*: $K = 2$, $N = 29$; East: $K = 5$, $N = 67$; North: $K = 9$, $N = 62$; West: $K = 4$, $N = 35$).

All results were then evaluated with the online program STRUCTURE HARVESTER web version 0.6 (Earl & VonHoldt, 2012) to explore which K has the highest probability and ΔK. Concurrently, the structure results were loaded onto the CLUMPAK online pipeline (Kopelman et al., 2015) which automatically summarizes the results and offers a graphical representation.

In addition, a discriminant analysis of principal components (DAPC) implemented in *R adegenet* (Jombart, 2008) was performed based on the complete data set.

A spatial analysis of molecular variance (SAMOVA) based on the *CR*-long alignment was performed in order to define the number of clusters (K) that are maximally differentiated from each other without constraint for the geographic composition of the groups. We applied the software SAMOVA v. 2 (Dupanloup et al., 2002) and tested for $K = 3$ to $K = 9$ with 10 simulated annealing runs with samples grouped according to the site of their origin (44 populations).

3 | RESULTS

3.1 | Variation in mitochondrial DNA sequences

After trimming the primer regions, the length of the final alignments for each of the three sequences was 859 bp for the mt *COI*, 1,109 bp

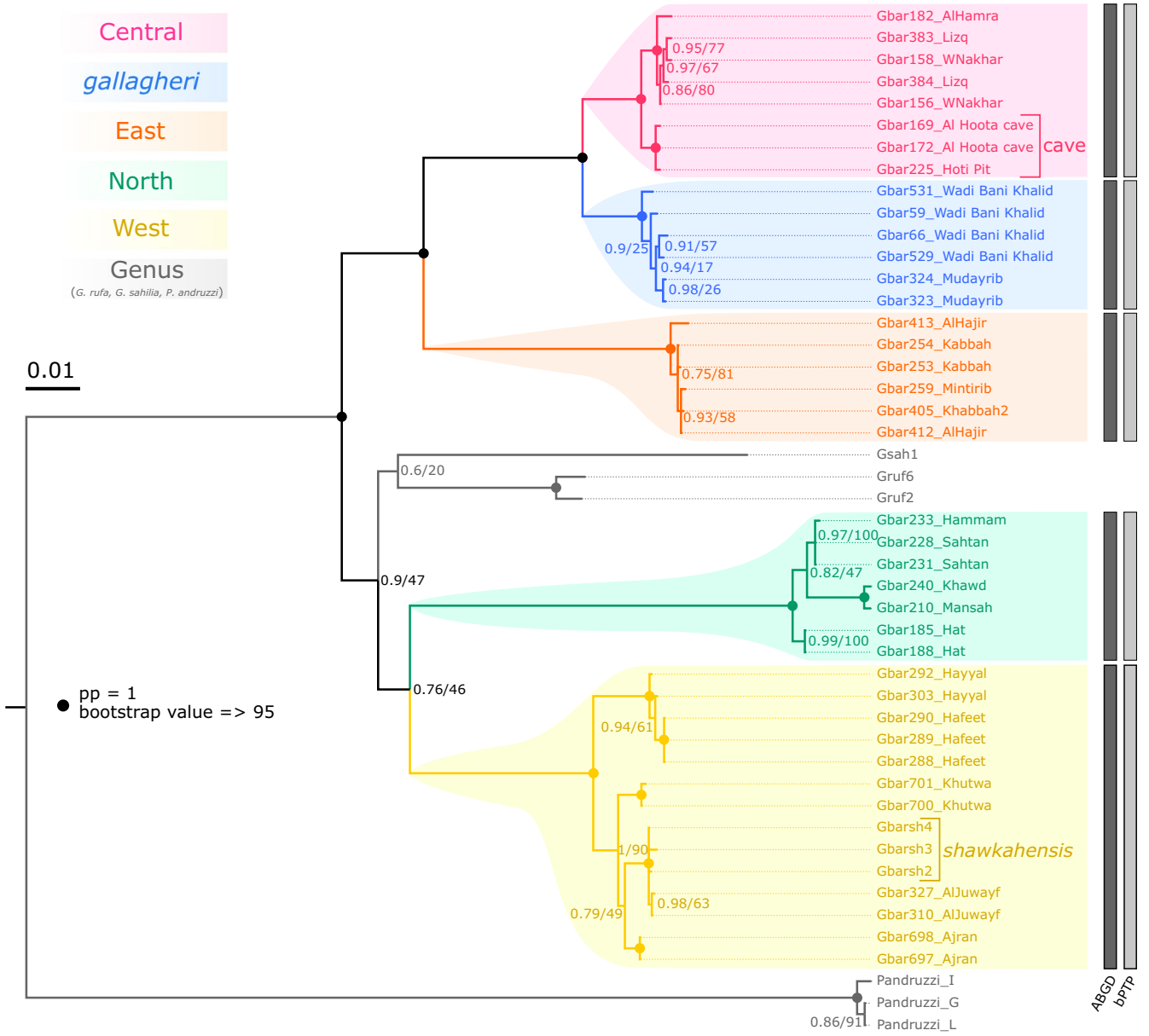

FIGURE 2 Phylogenetic BI tree based on the concatenated alignment (*COI* + *Cytb* + *CR*-long) of 44 individuals of the genus *Garra*and three individuals of *P. andruzzi* (out-group). Posterior probabilities (pp) and bootstrap values from the ML analysis are indicated at the nodes. Nodes which are highly supported by both analyses (pp = 1, bootstrap values ≥95) are marked with a filled dot. Individuals with "cave genotype" as well as sequences originated from the subspecies *Garra barreimiae shawkahensis*are indicated on the right side. Gray bars on the right represent results of two different species delimitation methods (ABGD = Atuomated Barcode Gap Detection, bPTP = Bayesian Poisson Tree Process)

for the mt *Cytb* and 818 bp for the mt *CR*. The alignment of the *CR*-short sequences was 401 bp. After trimming, the alignments *CR*-long and *CR*-short were 765 and 382 positions, respectively, in length.

All three mt markers were analyzed in 41 specimens of *G. barreimiae* from 23 sampling sites representing all clades. Combined with sequences from other taxa the concatenated alignment comprised 47 sequences and had a length of 2,739 bp (*COI*: position 1–859, *Cytb*: position 860–1968, *CR*-long: position 1,969–2,739).

The phylogenetic Bayesian inference (BI)/ML tree reconstruction based on the concatenated mt alignment resulted in trees presenting the same topology with five distinct and geographically separated clades corresponding to geographic regions (Figure 2). Accordingly, we will refer to them as Central, East, North and West clade as well as a *gallagheri* clade as this nomenclature was already established in the study of Pichler et al. (2018). The clade *gallagheri* comprised samples of *G. gallagheri* from Wadi Bani Khalid (WBK). The topology between the clades Central, East and *gallagheri* was highly supported whereas the relationship to both West and North clade remained unclear.

In the *CR*-long alignment comprising sequences of 314 individuals, 143 haplotypes were found with a max. genetic distance of ~13%. The resulting NJ tree (Figure S1) showed the same clades

TABLE 1 General information on analyzed sequences, mean and maximum p-distances, nucleotide and haplotype diversity of *Garra barreimiae*

Group	All	Central	*gallagheri*	East	North	West
N sequence	309	177	12	35	21	54
N haplotype	132	55	7	14	15	34
Hd	0.968	0.916	0.879	0.734	0.962	0.973
π	0.070	0.012	0.008	0.003	0.012	0.023
mean distance	7.0	1.2	0.8	0.3	1.2	2.3
max. distance	12.7	3.3	1.6	1.4	2.4	4.6

Note: Nucleotide (π) and haplotype diversity (Hd) as well as mean and max. p-distances within mt clades of *Garra barreimiae* (*CR*-long).

as the BI/ML tree. The split between Central and *gallagheri* clade is highly supported, the topology of the remaining clades received only poor support. Concerning the sequences of *G. rufa* and *G. sahilia*, their placement within the phylogenetic tree remains unclear.

Both species delimitation methods (ABGD, bPTP) yielded five clusters corresponding to the genetic clades obtained in the mt phylogenetic tree reconstructions (see Figure 2).

The genetic variation was quite high in all clades, and within group distances were highest in the West clade (mean p-distances: *CR*: 2.3%, *Cytb*: 0.8%) and North clade (*COI*: 0.7%) whereas the lowest were detected in the East and *gallagheri* clade (*CR*: 0.3%, *Cytb*: 0.2%, *COI*: 0.1%). Concerning Hd, the highest values were found within the West clade (*CR*: 0.97, *COI*: 0.87) and the *gallagheri* clade (*Cytb*: 1.0), nucleotide diversity (π) was highest in the West clade (*CR*: 0.023, *Cytb*: 0.008) and the North clade (*COI*: 0.007). Again, lowest values of Hd were detected in the East clade (*CR* 0.734, *Cytb*: 0.6) and the *gallagheri* clade (*COI*: 0.6), *gallagheri* was also the group that exhibits lowest π values for all three fragments (*CR*: 0.008, *COI*: 0.001, *Cytb*: 0.002; Table 1, Table S3).

When comparing the genetic groups to each other, highest between group differences were detected between North/Central (*COI*: min. 8.4%) and North/*gallagheri* (CR: min. 10.8%; *Cytb*: min. 8.2%), whereas lowest between group differences were found between Central/*gallagheri* in all three fragments (*CR*: min. 3.3%; *COI*: min. 1.9%; *Cytb*: min. 1.6%) (Table 2). For mean pairwise p-distances between the genetic groups, see Table S4.

In order to test for neutral evolution and to infer on past demographic changes within each clade, we calculated Tajima's *D*, which was negative in the clades Central and East but positive in the remaining clades *gallagheri*, North and West. However, the value of the East clade was the only one that significantly differed from zero (−1.859), which indicates an excess of rare alleles and hence a recent population expansion. In addition, we also calculated Fu's *F* and Ramos-Onsins' and Rozas' R2 statistics, which are considered to be more powerful to detect population fluctuations. The R2 statistics resulted in small positive values (0.06–0.21) in all clades, but again, only the East clade showed significant values. Fu's *F* statistics, on the other hand, resulted in negative values in all clades but clade *gallagheri* (detailed results are shown in Table S5). Again, only the values of clades East and Central were significant. In summary, these results rejected the hypothesis of constant population size in the clades East and Central. In contrast, the mismatch distribution exhibits a multimodal distribution in all clades indicating that they are in demographic equilibrium (Figure S2). However, these results should be treated with caution as we found rather high degrees of substructure in at least some of the clades (North, West). This can bias the outcome of such analyses significantly as these methods assume panmictic populations.

All MJ networks of the mt clades are depicted in Figures 3 and 4. The network of the Central clade reflects no clear pattern between the geographic origin of the samples. There is a "cave haplogroup" containing all samples from Al Hoota cave, all but two (Gbar504 and Gbar505) from the Hoti Pit cave, as well as most specimens with intermediate phenotype from WF and from GT (some are in HT 1 and 9).

TABLE 2 Maximum intra- and minimum interspecific p-distances (%) between clades based on three mt fragments (CR-long, COI and Cytb): C Central, G *gallagheri*, N North, E East, W West

	CR (*N* = 309)					*COI* (*N* = 42)					*Cytb* (*N* = 42)				
	C	G	N	E	W	C	G	N	E	W	C	G	N	E	W
Central	3.3					0.3					0.6				
gallagheri	3.3	1.6				1.9	0.2				1.6	0.4			
North	10.7	10.8	2.4			8.4	7.8	1.4			8.1	8.2	1.4		
East	9.5	8.9	10.2	1.4		5.5	5.7	7.8	0.3		5.3	5.3	8.0	0.5	
West	9.4	9.2	9.4	9.8	4.6	6.9	6.5	8.1	6.0	1.2	7.3	7.1	7.4	7.3	1.6

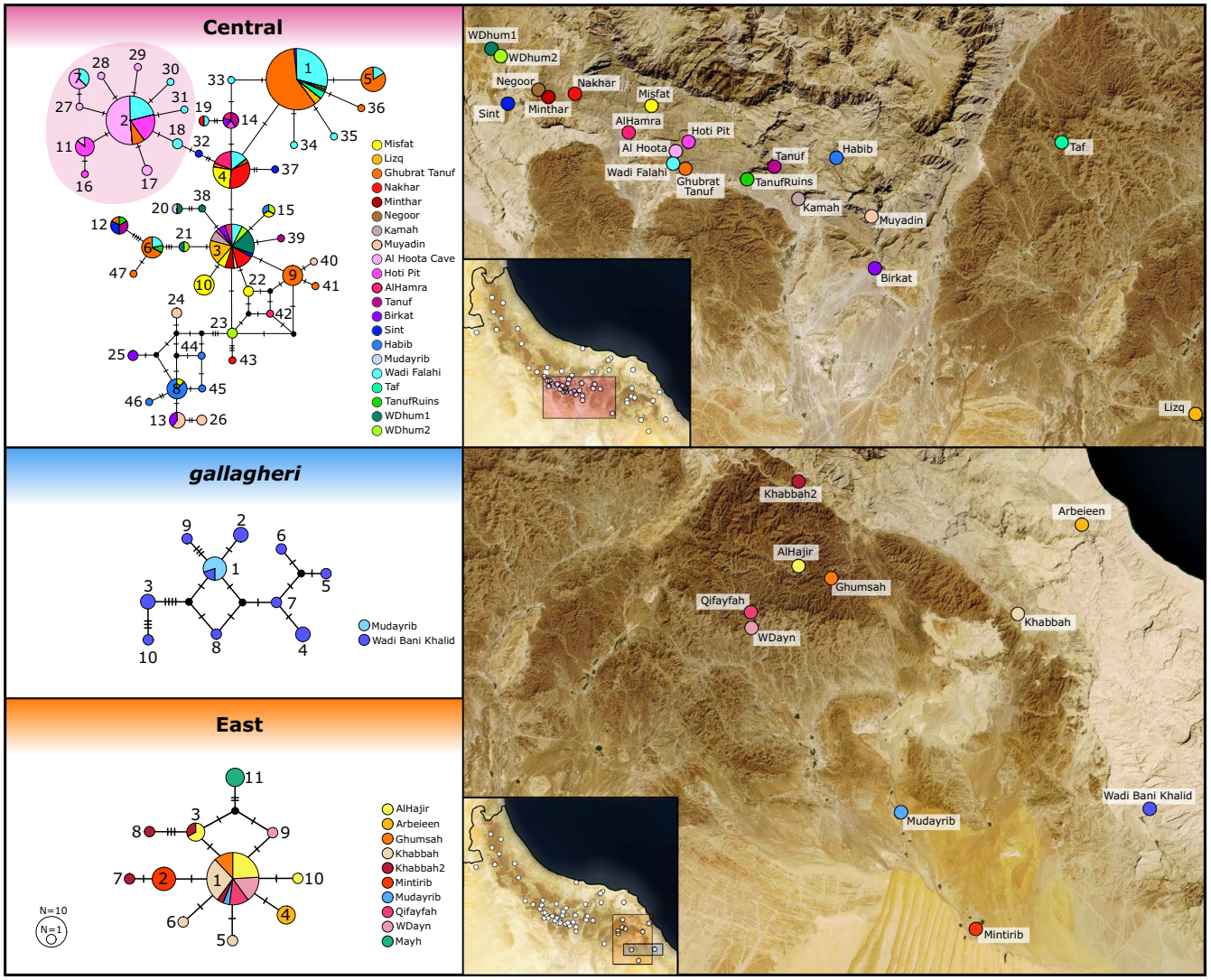

FIGURE 3 Median-joining networks based on *CR*haplotypes of mt clades (Central = pink, *gallagheri* = blue, East = orange). Each circle represents a haplotype, size of circles corresponds to the number of individuals sharing the haplotype. Colors indicate sampling sites. Every crossbeam on the connecting lines between haplotypes represents a substitution. Small black dots between haplotypes symbolize hypothetic haplotypes not present in the data set. Distribution area of each mt clade (small map) as well as detailed geographic origin of the samples are depicted in the maps to the right

The *gallagheri* network reveals no geographic structured pattern at all. The samples that originated from the second site Mudayrib all share the same HT. This HT is also found in samples retrieved from Wadi Bani Khalid (HT 1).

Depicted in the East clade network is one central HT (HT 1) comprising 60% of the samples and containing samples from all sampling sites except Arbeieen, Mintirib, and Mayh. The remaining 40% are divided into nine haplotypes which all but one (HT 3) comprise samples from only one sampling site.

The network of the North clade is more structured and comprises 19 haplotypes. It can roughly be divided into two geographically separated parts: The first comprising a central (HT 2) and nine associated haplotypes arranged in a star-like pattern (HT 7, 8, and 10–16) containing exclusively all samples from the northeastern located sites Khawd, Mahal, Manal, Samail, and Mansah. The site Mayh is even further east but is associated with the second part of the network, that comprises a major (HT 1) and several minor haplotypes containing samples from more southwestern located sites.

From the network of the West clade, it can be seen that this clade is significantly more differentiated compared to the other clades. It comprises 25 haplotypes and shows also some geographical structuring. All samples from the most northwards located sites (Wadi Shawkah, AlJuwayf, and Zihaymi) are clustered exclusively in five haplotypes (HT 2, 11, 23–25). At least seven mutational steps are between this "northern cluster" and the next closest haplotypes (HT 7 or 19). Another 14 mutational steps separate a geographically rather broad haplogroup comprising all and exclusively samples from more southwards lying sites (HT 1, 3–6, 8, 12–17).

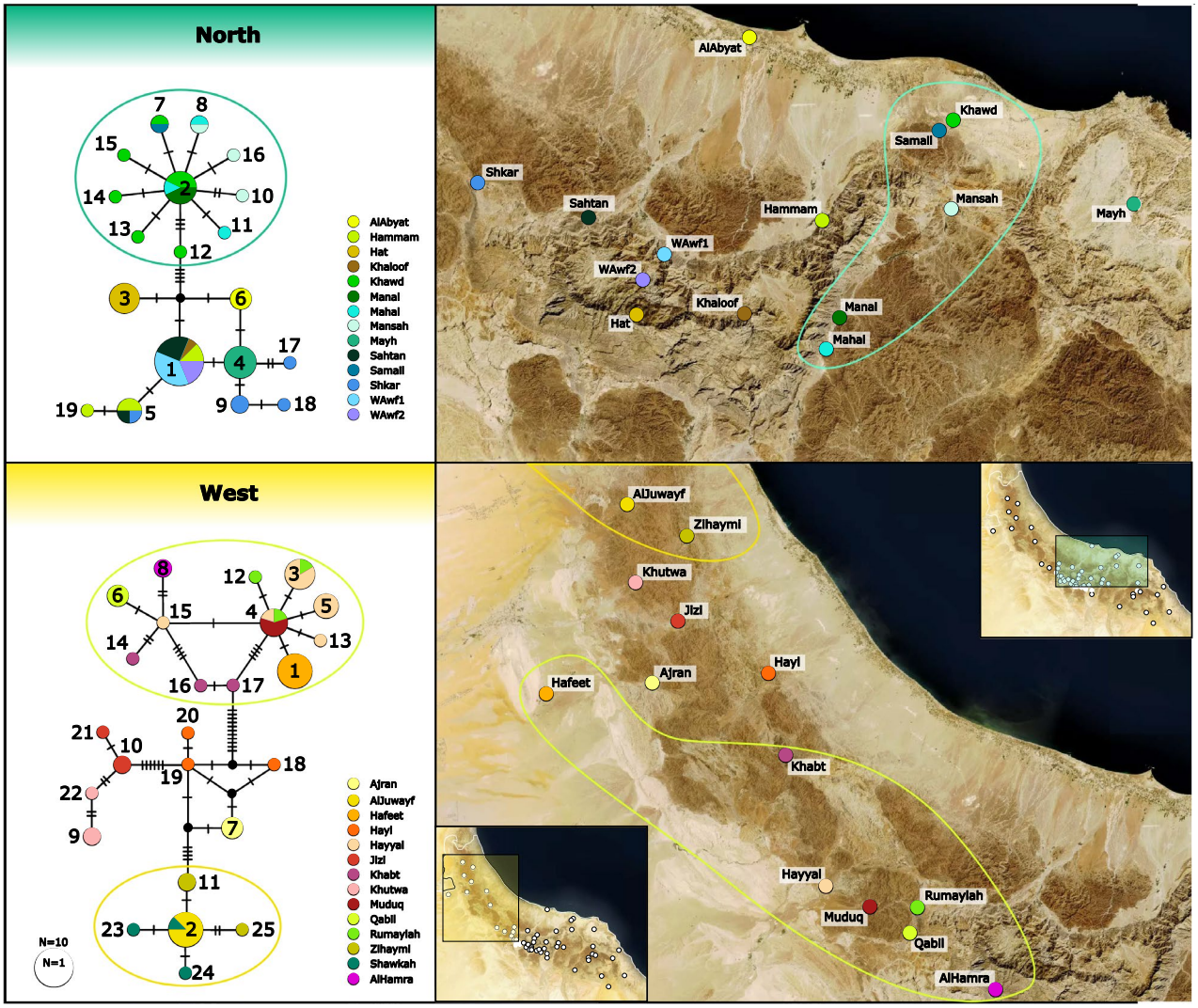

FIGURE 4 Median-joining networks based on CR haplotypes of mt clades (North = green, West = yellow). Each circle represents a haplotype, size of circles corresponds to the number of individuals sharing the haplotype. Colors indicate sampling sites. Every crossbeam on the connecting lines between haplotypes represents a substitution. Small black dots between haplotypes symbolize hypothetic haplotypes not present in the data set. Distribution area of each mt clade (small map) as well as detailed geographic origin of the samples are depicted in the maps to the right

3.2 | Microsatellite variation (population structure analysis)

3.2.1 | Genetic differentiation between groups and potential gene flow

We employed nuclear markers to test whether the groups detected via mt sequence data would be supported by microsatellite analyses. Given that the clusters of the initial STRUCTURE analysis resulted in similar groupings as the mt data analysis, the subsets of the microsatellite data were prepared accordingly. Since not all of the 19 microsatellite loci (Kirchner et al., 2014) worked equally good in the different clades, subsets were used for the analysis of the genetic structure within the clades. Hence, the non-working loci were excluded leaving: 17 in subset Central, 18 in subset *gallagheri*, 18 in subset East, 13 in subset North and 12 in subset West. Only nine loci worked well across all subsets and were used for STRUCTURE analysis of the complete data set. The reviewed genotypes of all individuals analyzed are listed in Table S1.

After inspecting the results of the initial STRUCTURE analysis, it became evident that it was necessary to reduce the number of samples belonging to the Central subset ($N = 331$) since the number of samples comprised in the remaining subsets is far lower and unequal sampling can affect the outcome of the analysis (Sebastien, 2016). Hence, for analysis we chose sampling sites covering the whole distribution area where at least 10 specimens were available. Furthermore, also samples with full (Al Hoota Cave: 37, Hoti Pit: 17) or partial cave genotypes (Ghubrat Tanuf: 11, Wadi Fallahi: 14) from the Central subset were excluded, because the differentiation between cave and surface populations produces a stronger signal than

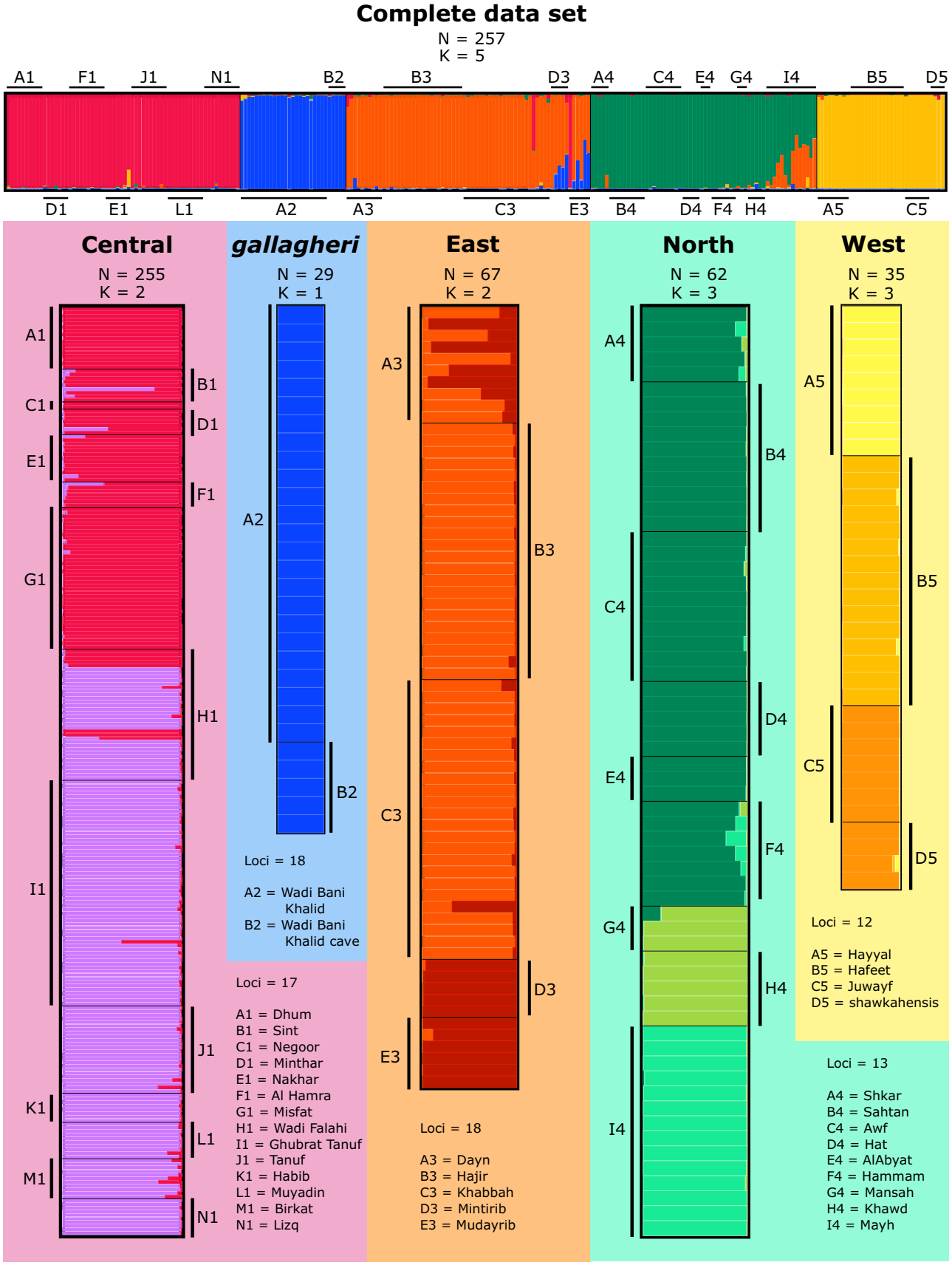

FIGURE 5 Population structure analyses based on the complete data set (top) and the subsets (bottom). For the complete data set, the most probable number of clusters was five. Concerning the subsets, STRUCTURE analyses resulted in $K = 2$ in Central (magenta), $K = 1$ in *gallagheri*(blue), $K = 3$ in North (green) and $K = 3$ in West clade (yellow)

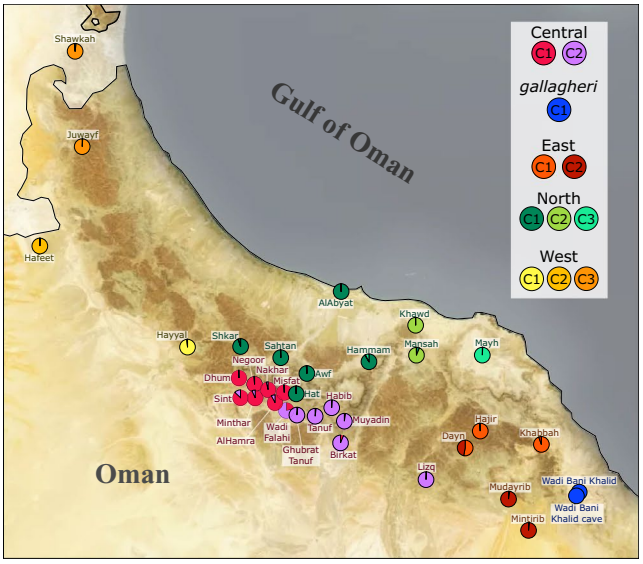

FIGURE 6 Geographic map of study area in Northern Oman. Sampling sites used for microsatellite analyses are marked via pie chart. The different colors of the pie chart fractions resemble the portion of cluster affiliation from this sampling site in each clade

any other population structure within the subset Central. Besides, these samples have already been examined in detail by Kirchner et al. (2017). This left us with 64 samples from the subset Central that were included in the STRUCTURE analysis of the complete data set and with 255 samples for the analysis of the Central subset only. The other subsets comprised 29 (*gallagheri*), 67 (East), 62 (North), and 35 (West) samples.

The reduced data set yielded the highest probability for $K = 5$. These five clusters correspond to the mt clades. The analyses were repeated with these five subsets corresponding to these clusters, which resulted in $K = 2$ for subset Central ($N = 255$, without cave/intermediate individuals), $K = 1$ for *gallagheri* ($N = 29$), $K = 2$ for East ($N = 67$), $K = 3$ for North ($N = 62$), and $K = 3$ for West ($N = 35$) as depicted in Figure 5. A map illustrating the distribution of the clusters from each clade is shown in Figure 6. Only few sites showed ambiguity of belonging (<90% affiliation to one cluster): two sites from Central clade (Sint, WF) and one from East clade (Dayn).

There are a few sites where STRUCTURE results indicated some degree of gene flow between the clusters. The samples from Mintirib and Mudayrib, which belong to the mt East clade show mixed genotypes between the clusters East and *gallagheri*. In addition, one individual from Mudayrib belongs 100% to the Central cluster according to nc and mt data. Individuals which originated from Mayh of the mt North clade, show high degree of mixed signals from both North as well as East group (Figure 5, $K = 5$).

Five clusters were identified with the "find.clusters" function of the DAPC based on the complete data set. Only a few individuals were grouped ambiguously when comparing the DAPC clusters with the mt clades (Figure S3).

3.2.2 | Genetic variability within groups/clusters

In order to assess genetic variation among and between populations, we calculated the descriptive statistics (Table 3) for the nine polymorphic unlinked microsatellite loci that worked well across the complete data set. In general, variation was high with 32–142 alleles per cluster across all loci. These results were corrected by taking the sample size into account and calculating the allelic richness according to the group with the fewest individuals. The highest allelic richness was found within the cluster Central_C2 (8.35), whereas the lowest was detected in West_C2 (2.25). Furthermore, the calculated mean H_e over all loci was highest in North_C3 (0.69) and lowest in West_C3 (0.24). Overall, the H_e was higher than the H_o in all clusters but West_C1 and West_C2. Most private alleles were detected in West_C3 (16.67%) and fewest in West_C1 (4.76%).

Null alleles were detected at least at one locus in most clusters. However, since none of these loci showed a constant significant deviation in all clusters, we did not exclude these loci from our analyses. Deviations from HWE (5% level) were detected in at least one locus in almost all clusters but West_C1 and West_C2, whereas Central_C1 had most loci with significant deviations. However, the reason for deviations from HWE in the majority of these loci is most likely due to the presence of null alleles. The results from the exact test indicate that all clusters are in HW equilibrium.

The F_{IS} is positive in all populations except the West_C1 and West_C2, indicating at least some degree of heterozygote deficiency especially in the cluster East_C2 (0.27). The pairwise F_{ST} values between pairs of clusters of each subset (Table 4) revealed highest values between West_C2 and West_C3 (0.36), whereas lowest F_{ST} values were found between the two clusters of the Central clade (0.053).

3.3 | Geographic distribution of the clades

With more than 50 sampling sites spread throughout Northern Oman, one can delimit distribution areas of each clade of *G. barreimiae* quite precisely (Figure S4). The gathered data gives detailed information on the distribution margins and thereby enables to identify potential barriers or paths to gene flow.

The Central clade with its distribution area amidst the Hajar mountains on the Saiq Plateau was the most extensively sampled group and shows moderate degree of variation. The microsatellite data reveals a clear differentiation between a western (Central_C1) and eastern (Central_C2) cluster, a pattern that could be explained by the change of direction of water drainage, one arm of the wadi close to the city Al Hamra leading southwards, another extending eastwards comprising all sites belonging to the eastern cluster (Central_C2). The distribution range of the Central clade is clearly restricted by geographic boundaries. All (but two) sampling sites lie on the southern slope of the Al Jebel al-Akhdar mountain range, which reaches altitudes up to 3,000 m asl (highest peak in Arabia:

TABLE 3 Population genetic parameters of the five subsets and their corresponding clusters

Pop	*N*	Na	A	Ar	pA	pA%	H_o	H_e	F_{IS}	p	Ld	Lp%	HWE
Central - 17 Loci													
All	255	173	10.18	—	24	13.87	0.51	0.63	0.18	0.08	JMLR	100.00	0.00
C1	101	130	7.65	7.07	9	6.92	0.49	0.59	0.17	0.09	PH8A, DL6X, JMLR	100.00	0.00
C2	154	142	8.76	8.35	15	10.56	0.53	0.60	0.11	0.19	—	100.00	0.00
gallagheri - 18 Loci													
C1	29	130	7.22	—	15	11.54	0.52	0.60	0.13	0.25	JQSO, 3Z7I, 3N43, I6G2	100.00	0.00
East - 18 Loci													
All	67	171	9.50	—	11	6.43	0.51	0.61	0.16	0.02	UHPE, HOLN, 9XNC, 3N43, I6G2, 3ROZ	100.00	0.00
C1	51	138	7.67	5.18	11	7.97	0.52	0.55	0.07	0.34	JMLR, 9XNC, I6G2	94.44	0.00
C2	16	116	6.44	5.47	0	—	0.50	0.65	0.27	0.13	3Z7I, 3ROZ	100.00	0.00
North - 13 Loci													
All	62	147	11.31	—	28	19.05	0.49	0.68	0.28	0.01	UHPE, 1EHE, CJHG, JMLR, 3N43, 2PUM, CB75	100.00	0.00
C1	40	94	7.23	4.07	13	13.83	0.45	0.54	0.18	0.20	UHPE, CB75	92.31	0.00
C2	8	60	4.62	3.78	9	15.00	0.48	0.53	0.15	0.44	—	84.62	0.00
C3	14	72	5.54	4.41	6	8.33	0.62	0.69	0.14	0.35	JQSO, JMLR	100.00	0.00
West - 12 Loci													
All	35	69	5.75	—	13	18.84	0.32	0.49	0.35	0.19	JQSO	91.67	0.00
C1	9	42	3.50	3.07	2	4.76	0.48	0.43	−0.05	0.69	—	75.00	0.62
C2	15	32	2.67	2.25	5	15.63	0.32	0.31	−0.02	0.64	—	66.67	0.61
C3	11	36	3.00	2.38	6	16.67	0.20	0.24	0.22	0.60	HOLN	75.00	0.00

Note: Pop Population, *N* Number of analyzed individuals, Na number of alleles per population, A mean number of alleles per locus, Ar allelic richness, pA number of private alleles, pa% percentage of private alleles per population based on the actual amount of alleles per population, H_o mean observed heterozygosity over all loci, H_e mean expected heterozygosity over all loc, F_{IS} inbreeding coefficient, p (mean) p-value for Hardy–Weinberg equilibrium (exact test, mean over all loci), Ld Loci, that deviated from HW-equilibrium and do not exhibit potential microsatellite null alleles, Lp% percentage of polymorphic loci, HWE p-value for Hardy–Weinberg equilibria (global test).

Jebel Shams) (Kusky et al., 2005). This mountain ridge limits the distribution of the Central clade to the North. To the east lies the Samail Gap, a valley stretching in north-south direction from the city Bidbid to Izki, that separates the eastern part of the Hajar mountains (Al Hajar Al Sharqi). South to the Central clade distribution the sand desert Rub al Khali ("empty quarter") expands, which acts as a natural boundary for freshwater fish migration. The western foothill of Jebel Shams represents the westernmost occurrence of the Central clade.

The sister group of the Central clade is the *gallagheri* clade which consists of all samples from the easternmost site Wadi Bani Khalid and most samples from nearby located Mudayrib. Since the geographic distribution of this clade is the most restricted, the genetic variation within this group is also rather low.

The East clade is genetically the most homogenous clade with lowest mt distances and very little substructure. This can be explained by the fact that the distribution covers only the eastern part of the Hajar mountains Al Hajar Al Sharqi and appendant geological structures which is not disrupted by natural barriers like deep valleys or high mountain ridges. Besides the *gallagheri* clade, the distribution of the East clade is geographically the most isolated one seemingly not connected to any other clade. It is likely that aquatic passageways between sampling sites of the East clade are more permanent and thereby, creating contact zones between populations and hence, preventing diversification within the clade. Concerning nc data, we found two clusters that roughly follow a north (East_C1) to south (East_C2) pattern but the presence of mixed genotypes at the site Dayn in the middle of the distribution area as well as very low F_{ST} values (Table 4) between these clusters indicate little genetic differentiation (Wright, 1978).

The distribution of the North clade stretches across the northern part of the mountains and reaches the coastal area at the Gulf of Oman. This clade exhibits high genetic variation and can be divided

TABLE 4 Pairwise F_{ST} values between the clusters of *Garra barreimiae*

Cluster 1	F_{ST}	Cluster 2
Central_C2	0.053	Central_C1
East_C2	0.069	East_C1
North_C2	0.131	North_C1
North_C3	0.176	North_C1
North_C3	0.202	North_C2
West_C2	0.167	West_C1
West_C3	0.274	West_C1
West_C3	0.366	West_C2

Note: Pairwise F_{ST} values between the clusters of *Garra barreimiae* based on 17 (Central), 18 (East), 13 (North), 12 (West) microsatellite loci. Calculated with GenAlEx, all values are highly significant.

in at least two subgroups. Again, the Samail Gap seems to form a barrier dividing a western cluster (North_C1) from an eastern cluster (North_C2). In addition, the easternmost site Mayh represents a third cluster based on nc data (North_C3) but shows no such differentiation based on mt data. Since the distribution of the North clade extends from high mountain peaks to coastal lowland areas of the Al Batinah plain, it comprises high habitat diversity including many extensive wadis that channel water directly toward the Gulf of Oman. These geological structures could be accountable for the high genetic structure, preventing contact between populations of different wadis even during high water level periods.

The West clade is distributed along the western ridge of the Hajar mountains and reaches northwards at least until Wadi Shawkah in the UAE. This clade shows highest within group differences, which are well reflected in the phylogenetic tree and network reconstructions (Figures 2 and 3, Figure S1). Two main subclades were found, one northern exhibiting high degree of substructure and including samples from the subspecies *G. b. shawkahensis* (also represented by nc cluster West_C3), and a second comprising all southern samples (reflected by two nc cluster West_C1 and West_C2). In this case, the distribution limits of the subclades run along the Wadi Hibi, which leads from the mountains in northeastern direction toward the coast. The highest F_{ST} values were detected within the West clade indicating more genetic differentiation than in any other clade (Table S6). The Al Batinah plain also overlaps with the northeastern distribution of the West clade. Therefore, we assume that these geological barriers caused the high degree of genetic differentiation within the West clade.

SAMOVA results revealed highest Fct values (0.93434) at $K = 4$, hence identified four clusters. These clusters are mostly consistent with the phylogenetic reconstructions of mt sequences, with exception that clade *gallagheri* (comprising two sampling sites) is part of the Central cluster in the SAMOVA. However, pooling the clades *gallagheri* and Central into the same cluster is not surprising, since these are also sister groups in the mt phylogenetic reconstructions exhibiting the lowest between group p-distances (0.04), albeit the geographic distances are rather high. All the other clades are well supported by SAMOVA clusters.

4 | DISCUSSION

4.1 | Phylogeography

The most striking result of this study was the high genetic variation within *G. barreimiae* found throughout the rather small geographic range around the Hajar mountains. Using a comprehensive data set of 632 individuals and employing both mt sequences as well as microsatellite data, five genetically distinct groups were detected, which were named due to their distinct geographic distribution North, East, West, Central and, according to the taxon name, *gallagheri* clade. We found strong correlations between genetic and geomorphological structures (see chapter "Geographic Distribution of the Clades" in Results).

Within the Central clade, the mt tree presents a "cave subclade" comprising only troglomorphic fish originated from Al Hoota cave and Hoti Pit as well as hybrid individuals between cave and nearby surface sites (Kirchner et al., 2017). This supports the findings of the previous study by Kruckenhauser et al. (2011). Originally, the species *G. gallagheri* (former subspecies of *G. barreimiae*) was described only from the site Wadi Bani Khalid and thereby represents the geographically most restricted clade. However, members of this taxon have been discovered for the first time also at another site (Mudayrib), together with individuals from the East clade. The East clade is genetically the most homogenous clade. In contrast, both the North clade and the West clade exhibit a much higher genetic variation in respect of mt as well as nc data. Pichler et al. (2018) investigated the morphological variation among the clades of *G. barreimiae* and found that the West clade is also the most diverging clade. It would be highly interesting to study whether the genetic substructure of this clade is reflected also in morphological differences (this clade contains individuals sampled closest to the *locus typicus* of the nominate form as well as from the *locus typicus* of *G. b. shawkahensis*). For this purpose, it would be necessary to measure also individuals assigned to *G. b. shawkahensis* from the Wadi Shawkah in the UAE. Unfortunately, only tissue samples from these specimens were available for this study.

Although most sampling sites of each cluster are not connected by any perennial water body, the microsatellite analysis shows no genetic differentiation between single sampling sites. Therefore, we assume that there must be some recent gene flow between populations belonging to the same cluster, either due to underground water connections in the karst area, as it has been reported for other cyprinids (Palandačić et al., 2012) or because of sporadic heavy rainfalls, which quickly fill wadis with water and thereby generate temporal surface water connections.

4.2 | Hybridization

Although the nuclear and mt data give concordant results, following the geographic distribution, there are some exceptions: a few individuals possess mixed genotypes from two clades, as in microsatellite

analyses hybridization with backcrossing cannot be detected after few generations, this indicates recent hybridization. In the present study, indication for hybridization between the clades was only detected at three out of 37 sites that were analyzed with nc microsatellite data. This affects the closely related clades East and *gallagheri* but also the more distant clades East and North. Most individuals from the site Mudayrib cluster with the *gallagheri* clade in the mt tree but exhibit a mixed nc genotype between the clades East and *gallagheri*. A similar pattern is found in the close-by site Mintirib where all samples belong to the East clade (mt) but show mixed genotype between the clades East and *gallagheri* (nc). Both sites are located on the southeastern margin of the distribution area, southbound their distribution is limited in general by the sand desert Ramlat al-Wahiba. In this case, when the water level is rising due to rainfall, it seems plausible that also distribution expands along temporary water routes and thereby enables sporadic contact zones between geographically separated taxa. The third site where hybrids were detected is Mayh. Based on mt data, half of the individuals originating from Mayh belong to the North clade whereas the other half belongs to the East clade. Most individuals of this site exhibit mixed nc genotypes irrespective of their mt clade affiliation.

Interspecific hybridization is common in cyprinids. Genetic evidence is not that frequent, but there are some striking examples, like hybrids between taxa of the European *Phoxinus* species complex (Palandačić et al., 2017), sometimes also over multiple generations like between bighead and silver carp (Lamer et al., 2015) or even between different genera like roach and bream in Ireland (Hayden et al., 2010) or goldfish and introduced common carp in Australia (Haynes et al., 2012).

Overall, it can be stated that hybridization between the clades of *G. barreimiae* and between *G. barreimiae* and *G. gallagheri* is not a common event and does not lead to blurring of the genetic patterns. It seems plausible that the sites in which hybridization occurs are zones of secondary contact between the clades. This might be due to anthropogenic influence. At least the samples of Mudayrib and Mintirib were collected from aflaj, which are manmade irrigation canal systems transporting water over kilometers from mountain springs to the cultivated gardens (personal communication with a local from Mintirib). Introductions either deliberately or as a consequence of channeling water via falaj might play a role.

Another indication that hybridization occurs only very rarely is that even though three out of the five clades occur in close vicinity around the mountain peak of Jebel Shams (West, North, Central), not a single individual with mixed genotype is detected via microsatellite analyses from this area. There are two individuals from the site Al Hamra that belong to the West clade based on mt data but are grouped perfectly with cluster C1 of the Central clade, to which all other individuals from Al Hamra belong.

4.3 | Relationship between clades and possible origin and age of clades

Kruckenhauser et al. (2011) questioned the monophyly of *G. barreimiae*. Despite the fact that in the tree presented in the current study, *G. rufa* as well as *G. sahilia* appear rather close to the *Garra* species complex in Northern Oman, with similar genetic distances found between the clades, the presented data did not provide sufficient information regarding the monophyly.

Collectively, analyses of mt and nc data clearly identified five well differentiated groups. The SAMOVA resulted in four significantly divergent groups almost identical to the groups detected by the phylogenetic tree and STRUCTURE analyses with the only difference that the clades Central and *gallagheri* formed one cluster in the SAMOVA. This result can be reasonably explained by the comparatively low genetic but high geographic distances between the Central and *gallagheri* clades.

Although we found clearly distinct clades within our data set, the phylogenetic relationships between these clades as well as to other closely related species of the genus are still not completely resolved and therefore, we cannot evaluate whether the five investigated clades are monophyletic with respect to *G. rufa*.

Because of the unresolved relationships between the clades, it cannot be concluded if the Arabian Peninsula was colonized in one or several independent migration events. It is difficult to deduce when and where the first migration of *Garra* into the Arabian Peninsula occurred, but it is quite likely that migration happened along the Persian Gulf during a glacial maximum when the sea level was lower than today. The flooding of Persian Gulf Basin at the end of the last glacial maximum (LGM; 20 kya) marks the latest possible time of migration along this route. With a maximum depth of 110 m the Persian Gulf was exposed for over 70 kya until sea level rose after the end of the last Glaciation (Clark et al., 2009). During that time the climate was subtropical, and the Persian Gulf formed the fertile Ur-Schatt River Valley that was recharged by the rivers Euphrat, Tigris, Karun, and Wadi Batin meaning the majority of freshwater of Southwest Asia drained into the Ur-Schatt River Valley (Rose, 2010). Considering the distances found between clades it is unlikely that this amount of differentiation arose after a migration onto the Arabian Peninsula during the LGM. The question whether several invasions occurred over a longer period of time remains open. To further address this and other questions, comparative genome-wide analyses of a broader data set including more closely related species from this region, specifically from Iran as well as southern Oman would be necessary.

In order to get insight into the phylogenetic history of *Garra* in Northern Oman it would be of high interest to get an estimate when the splits happened. Yet, a molecular clock needs to be calibrated with an independent source of information such as fossils. Unfortunately, no fossil records were found so far from the wide distribution range of extant *Garra* spp. in Africa and Asia. The only known fossil evidence of *Garra* was retrieved from Armenia and dated back to the Pliocene (Vasilyan & Carnevale, 2013) and therefore was not applicable as a molecular clock calibration for the present tree. However, as a first approximation the substitution rates from the literature for the *Cytb* gene were used to estimate the approximate age of the clades. Two different substitution rates have been proposed concerning cyprinids: 0.53% per million years (Boore, 1999)

and 0.76% per million years (Dowton et al., 2009). We roughly estimated the age by applying the two substitution rates to the mean p-distances of *Cytb* sequences of each clade. Three time intervals of diversification were identified: the first interval reflects the split between the Central clade and the *gallagheri* clade, which corresponds to an estimate of to 1.3 (respectively 1.8) Myr and represents the youngest split. The second interval was dated to 3.6 (5.2) Myr and comprises the split between East clade and Central/*gallagheri* clades. The splits between both the West and North clade and all other clades represent the oldest ones, and estimates range from 4.9 up to 5.6 (7.2–8.0) Myr. Accordingly, the age of the cave population in comparison to the genetically closest surface population (Nakhar) most probably has occurred not earlier than the Middle Pleistocene (0.8–0.1 Myr). In those days, the climate in Oman was characterized by gradually increasing aridity interrupted by several pluvial periods caused by spikes in Indian Ocean monsoon intensity (Blechschmidt et al., 2009; Burns et al., 1998). It is plausible that these climate changes triggered the colonialization of the Al Hoota cave, either to retreat to the few remaining water reservoirs when temperature was rising, and surface water level dropped, or to occupy new freshwater habitats established in the course of heavy rain periods.

4.4 | Species delimitation

In the past few years, several new species of the genus *Garra* from the Middle East have been described: *Garra jordanica* (Hamidan et al., 2014) from the Dead Sea Basin, *Garra mondica* (Sayyadzadeh et al., 2015) and *Garra amirhosseini* (Esmaeili et al., 2016) from Iran, as well as *Garra sindhi* (Lyon et al., 2016) and *Garra smarti* (Krupp & Budd, 2009) from southern Oman. Moreover, new cave species have been described from Iran, for example, *Garra lorestanensis* (Mousavi-Sabet & Eagderi, 2016) and *Garra tashanensis* (Mousavi-Sabet et al., 2016). Obviously, the species diversity within this genus had been underestimated in the past and needs thorough revision. Since this genus comprises over 130 described species (Froese & Pauly, 2018) and has a vast Afro-Asian distribution, a proper revision based on an integrative approach is unfeasible to conduct at once. Thus, this study is one of many steps necessary to elucidate the inestimable diversity this genus holds. Generally, it must be emphasized that the rising number on studies revealing cryptic diversity highlight also the problem of underestimating the actual (so far overseen) biodiversity, which has a major effect on conservation efforts.

A study by Pichler et al. (2018) stated, that the clades of *G. barreimiae* are morphologically quite variable and not to distinguish on the basis of single characters. The West clade could be discriminated by a few characters like number of gill rakers, precaudal and intermediate vertebrae, and branched pelvic fin rays. But in general, the variation of characters of all clades overlapped considerably. Hence, the genetically well-separated clades of *G. barreimiae* could not be distinguished morphologically based solely on single traits that are commonly accepted and prevalently used to classify Cypriniformes. Our results illustrate that the level of genetic variation is not necessarily coherent with morphological differentiation, emphasizing once more the importance of thorough and careful taxonomic decisions based on data from multiple sources in the sense of integrative taxonomy (Schlick-Steiner et al., 2010).

Eventually, a decision whether the here described clades should be classified as separate species depends on the species concept applied. De Queiroz (2005, 2007) presented a unified species concept that proposes that the existence of separately evolving metapopulation lineages is the only necessary property of species and the common element reconciling alternative concepts of species. Yet, this concept still embraces the various properties that have been considered defining alternate species concepts (e.g., phylogenetic species concept) and sees them as evidence for separate evolving lineages. In the case of *G. barreimiae*, taxa are morphologically similar and delimiting species boundaries based solely on few morphological characters is futile. Molecular genetic methods, on the other hand, revealed well differentiated and geographically separated clades which were also well supported by two different approaches of species delimitation methods. The finding of no or very few incidences of gene flow provides relevant evidence for the assumption of distinct (cryptic) species.

With enhanced usage of molecular genetic methods in biodiversity research, especially since DNA barcoding intensified, the term "cryptic species" or "pseudocrpytic species" has been utilized increasingly in the past decade. Yet, there are no commonly agreed criteria what actually defines a cryptic species. Hence, the term has been used inconsistently depending on which organism group was investigated and on previous knowledge about the taxon (Fišer et al., 2018). Fišer et al. (2018) identified three mechanisms causing cryptic diversity: recent divergence, phylogenetic niche conservatism (morphological stasis) and morphological convergence. Considering the genetic distances and age estimations, it becomes evident that the clades appear old enough to expect that morphological changes could have taken place. Yet, the phylogenetic relationships among the clades are close enough to exclude convergence as an explanation. It seems more likely that long-lasting extreme environmental conditions caused stabilizing selection on morphology, as suggested by Bickford et al. (2007). They stated that evolving under severe environmental extremes could limit changes in morphology. The environmental fluctuations in Northern Oman affected all clades of *G. barreimiae* similarly, thus it might have prevented morphological diversification. This can also be confirmed for water quality parameters (temperature, conductivity, pH) as well as habitat structures, which are strongly varying between different sampling sites according to our field observations. Following the suggestions of Bickford et al. (2007) who considered species to be cryptic if they have been classified as a single nominal species and cannot be distinguished by morphology, we regard the term cryptic species as justified in this case at least for the North, Central and East clades.

5 | CONCLUSIONS

In the present study, the genetic variability and phylogeographic pattern within the *Garra* species complex of Northern Oman was

investigated. The high genetic distances between mt clades comparable to distances between other species of this genus together with the correspondence of clades with the genetic groups inferred from microsatellite data lets us assume that they represent separately evolving lineages in the sense of De Queiroz (2007) and thus should be considered as distinct species. The fact that morphological characters to discriminate clades unambiguously are lacking confirms the view that the *Garra* species complex in Northern Oman is composed of cryptic species in the sense of Bickford et al. (2007). The taxon *G. barreimiae* is currently listed as "Least Concern" according to the IUCN Red List. Yet, as desertification, habitat destruction as well as water shortage and potential chemical pollution pose serious threats to organisms like *G. barreimiae* occupying extreme habitats. Hence, the updated species status of the five independent lineages of what was so far considered as *G. barreimiae* may provide an adequate basis for a conservation reassessment.

6 | POST SCRIPTUM

A taxonomic revision of the *Garra* species complex from Northern Oman was performed in Kirchner et al. (2020). In this publication we revised, on the basis of the data presented in the present study and additional morphological and genetic analyses, the taxonomic status of the described clades as follows: *G. barreimiae* Fowler & Steinitz, 1956 (West clade), *G. gallagheri* Krupp, 1988 (*gallagheri* clade), *Garra longipinnis* Banister & Clarke, 1977 (Central clade), *Garra shamal* Kirchner et al., 2020 (North clade) and *Garra sharq* Kirchner et al., 2020 (East clade). Eventually, this update on the species delimitation and taxonomy may lead to a re-evaluation of the IUCN Red List conservation status of the *Garra* species from Northern Oman, according to which *G. barreimiae* is currently assigned to "Least Concern".

ACKNOWLEDGEMENTS

The study was funded by the University of Vienna (uni:docs fellowship), the Natural History Museum of Vienna (Central Research Laboratories), by the Ministry of Commerce and Industry of the Sultanate of Oman and by the Ministry of Tourism of the Sultanate of Oman. The Federal Ministry for Transport, Innovation and Technology Austria provided financial support in the course of a FEMtech practical. The authors want to thank Johannes Rath and Christian Baranyi from the University of Vienna for providing help with the Sequence Analyzer, Harald Krenn (University of Vienna) for his general help and support, Arthur Pichler for his valuable input and overall help, Barbara Däubl (NHM Vienna) and Julia Schindelar (NHM Vienna) for technical assistance in the laboratory, Robert Illek (NHM Vienna) for support in practical field work, Lukas Plan (NHM Vienna) and Katharina Bürger (NHM Vienna) for their speleological and geographical expertise and for helping with collecting samples from the cave sites, Frank Zachos (NHM Vienna) and Stephan Koblmüller (University of Graz) for help and support with population genetic analyses, and Cristiano Bertolucci (University of Ferrara) and Silvia Fuselli (University of Ferrara) as well as Jörg Freyhof (IGB Leibniz) and Matthias Geiger (ZFMK Bonn) for providing additional samples. We want to express our gratitude to the Ministry of Environment and Climate Affairs of Oman for their cooperation, support and the sampling permits and to Michael Robinson, a zoologist at the Sultan Qaboos University in Muscat, for his intellectual input.

ORCID

Sandra Kirchner https://orcid.org/0000-0001-8715-6064
Elisabeth Haring https://orcid.org/0000-0002-5411-1879

REFERENCES

Adams, M., Raadik, T. A., Burridge, C. P., & Georges, A. (2014). Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology*, *63*(4), 518–533. https://doi.org/10.1093/sysbio/syu017

Al-Shaqsi, S. (2010). Care or Cry: Three years from Cyclone Gonu. What have we learnt? *Oman Medical Journal*, *25*(3), 162–167. https://doi.org/10.5001/omj.2010.50

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, *22*(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Blechschmidt, I., Matter, A., Preusser, F., & Rieke-Zapp, D. (2009). Monsoon triggered formation of Quaternary alluvial megafans in the interior of Oman. *Geomorphology*, *110*(3–4), 128–139. https://doi.org/10.1016/j.geomorph.2009.04.002

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, *27*(8), 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Burns, S. J., Matter, A., Frank, N., & Mangini, A. (1998). Speleothem-based palaeoclimate record from northern Oman. *Geology*, *26*(6), 499–502. https://doi.org/10.1130/0091-7613(1998)026<0499

Calderoni, L., Rota-Stabelli, O., Frigato, E., Panziera, A., Kirchner, S., Foulkes, N. S., Kruckenhauser, L., Bertolucci, C., & Fuselli, S. (2016). Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization. *Heredity*, *117*(5), 383–392. https://doi.org/10.1038/hdy.2016.59

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl Tutorial. *Bioinformatics (Oxford, England)*, *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., McCabe, A. M., & Maximum, T. L. G. (2009). *Science (New York, N.Y.)*, *325*(5941), 710–714. https://doi.org/10.1126/science.1172873

de Queiroz, K. (2005). A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences*, *56*(18), 196–215. https://doi.org/10.1073/pnas.0502030102

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

Dowton, M., Cameron, S. L., Dowavic, J. I., Austin, A. D., & Whiting, M. F. (2009). Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, *26*(7), 1607–1617. https://doi.org/10.1093/molbev/msp072

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11*(12), 2571–2581. https://doi.org/10.1046/j.1365-294X.2002.01650.x

Durand, J. D., Tsigenopoulos, C. S., Ünlü, E., & Berrebi, P. (2002). Phylogeny and biogeography of the family cyprinidae in the Middle East inferred from cytochrome b DNA - Evolutionary significance of this region. *Molecular Phylogenetics and Evolution*, *22*(1), 91–100. https://doi.org/10.1006/mpev.2001.1040

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Esmaeili, H. R., Sayyadzadeh, G., Coad, B. W., & Eagderi, S. (2016). Review of the genus *Garra* Hamilton, 1822 in Iran with description of a new species : A morpho- molecular approach (Teleostei : Cyprinidae). *Iranian Journal of Ichthyology*, *3*(2), 82–121.

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, *27*(3), 613–635. https://doi.org/10.1111/mec.14486

Froese, R., & Pauly, D. (2018). *FishBase*. www.fishbase.org.version (06/2018). World Wide Web Electronic Publication.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2), 915–925.

Goudet, J. (1995). Fstat Version 1.2.: A computer program to calculate F-statistics. *Journal of Heredity*, *86*, 485–486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Gunawardhana, L. N., & Al-Rawas, G. A. (2014). Trends in extreme temperature and precipitation in Muscat, Oman. *Proceedings of the International Association of Hydrological Sciences*, *364*(June), 57–63. https://doi.org/10.5194/piahs-364-57-2014

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hamidan, N. A., Geiger, M. F., & Freyhof, J. (2014). *Garra jordanica*, a new species from the Dead Sea basin with remarks on the relationship of *G. ghorensis*, *G. tibanica* and *G. rufa* (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, *25*(3), 223–236.

Hayden, B., Pulcini, D., Kelly-Quinn, M., O'Grady, M., Caffrey, J., McGrath, A., & Mariani, S. (2010). Hybridisation between two cyprinid fishes in a novel habitat: Genetics, morphology and life-history traits. *BMC Evolutionary Biology*, *10*(1), 169. https://doi.org/10.1186/1471-2148-10-169

Haynes, G. D., Gongora, J., Gilligan, D. M., Grewe, P., Moran, C., & Nicholas, F. W. (2012). Cryptic hybridization and introgression between invasive Cyprinid species *Cyprinus carpio* and *Carassius auratus* in Australia: Implications for invasive species management. *Animal Conservation*, *15*(1), 83–94. https://doi.org/10.1111/j.1469-1795.2011.00490.x

Huelsenbeck, J. P., & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Keenan, K., Mcginnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kirchner, S., Kruckenhauser, L., Pichler, A., Borkenhagen, K., & Freyhof, J. (2020). Revision of the *Garra* species of the Hajar Mountains in Oman and the United Arab Emirates with the description of two new species (Teleostei: Cyprinidae). *Zootaxa*, *4751*(3), 521–545. https://doi.org/10.11646/zootaxa.4751.3.6

Kirchner, S., Sattmann, H., Haring, E., Plan, L., Victor, R., & Kruckenhauser, L. (2017). Investigating gene flow between the blind cavefish *Garra barreimiae* and its conspecific surface populations. *Scientific Reports*, *7*(1), 5130. https://doi.org/10.1038/s41598-017-05194-3

Kirchner, S., Weinmaier, T., Rattei, T., Sattmann, H., & Kruckenhauser, L. (2014). Characterization of 19 new microsatellite loci for the Omani barb *Garra barreimiae* from 454 sequences. *BMC Research Notes*, *7*, 522. https://doi.org/10.1186/1756-0500-7-522

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kruckenhauser, L., Haring, E., Seemann, R., & Sattmann, H. (2011). Genetic differentiation between cave and surface-dwelling populations of *Garra barreimiae* (Cyprinidae) in Oman. *BMC Evolutionary Biology*, *11*(1), 172. https://doi.org/10.1186/1471-2148-11-172

Krupp, F., & Budd, K. (2009). A new species of the genus *Garra* (Teleostei: Cyprinidae) from Oman. *International Journal of Ichtyology*, *15*(2), 117–120.

Kusky, T., Robinson, C., & El-Baz, F. (2005). Tertiary-Quaternary faulting and uplift in the northern Oman Hajar Mountains. *Journal of the Geological Society*, *162*(5), 871–888. https://doi.org/10.1144/0016-764904-122

Lamer, J. T., Ruebush, B. C., Arbieva, Z. H., McClelland, M. A., Epifanio, J. M., & Sass, G. G. (2015). Diagnostic SNPs reveal widespread introgressive hybridization between introduced bighead and silver carp in the Mississippi River Basin. *Molecular Ecology*, *24*(15), 3931–3943. https://doi.org/10.1111/mec.13285

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. https://doi.org/10.1093/molbev/msw260

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lyon, R. G., Geiger, M. F., & Freyhof, J. (2016). *Garra sindhi*, a new species from the Jebel Samhan Nature Reserve in Oman (Teleostei: Cyprinidae). *Zootaxa*, *4154*(1), 79–88. https://doi.org/10.11646/zootaxa.4154.1.5

Mousavi-Sabet, H., & Eagderi, S. (2016). *Garra lorestanensis*, a new cave fish from the Tigris River drainage with remarks on the subterranean fishes in Iran (Teleostei: Cyprinidae). *FishTaxa*, *1*(1), 45–54.

Mousavi-Sabet, H., Vatandoust, S., Fatemi, Y., & Eagderi, S. (2016). Tashan Cave a new cave fish locality for Iran; and *Garra tashanensis*, a new blind species from the Tigris River drainage (Teleostei: Cyprinidae). *FishTaxa*, *1*(3), 133–148.

Moyle, P. B., & Leidy, R. A. (1992). Loss of biodiversity in aquatic ecosystems: Evidence from fish faunas. In: P. L. Fiedler, & S. K. Jain (Eds.), *Conservation biology* (pp. 127–169). Springer. https://doi.org/10.1007/978-1-4684-6426-9_6

Palandačić, A., Matschiner, M., Zupančič, P., & Snoj, A. (2012). Fish migrate underground: The example of *Delminichthys adspersus* (Cyprinidae). *Molecular Ecology*, *21*(7), 1658–1671. https://doi.org/10.1111/j.1365-294X.2012.05507.x

Palandačić, A., Naseka, A., Ramler, D., & Ahnelt, H. (2017). Contrasting morphology with molecular data: An approach to revision of species complexes based on the example of European *Phoxinus* (Cyprinidae). *BMC Evolutionary Biology*, *17*(184), 1–17. https://doi.org/10.1186/s12862-017-1032-x

Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pichler, A., Ahnelt, H., Kirchner, S., Sattmann, H., Haring, E., Handschuh, S., Freyhof, J., Victor, R., & Kruckenhauser, L. (2018). The morphological diversity of *Garra barreimiae* [Teleostei : Cyprinidae]. *Environmental Biology of Fishes*, *101*, 1053–1065. https://doi.org/10.1007/s10641-018-0758-7

Por, F. D. (1989). *The legacy of tethys. An aquatic biogeographic of the Levant*. Kluwer Academic.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics Society of America*, *155*(2), 945–959.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, *21*(8), 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

R Development Core Team (2015). *R: A language and environment for statistical computing*. http://www.r-project.org

Rambaut, A. (2007). *FigTree, a graphical viewer of phylogenetic trees*. http://tree.bio.ed.ac.uk/software/figtree/

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Rose, J. I. (2010). New light on human prehistory in the Arabo-Persian Gulf Oasis. *Current Anthropology*, *51*(6), 849–883. https://doi.org/10.1086/657397

Rousset, F. (2008). Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Sayyadzadeh, G., Esmaeili, H. R., & Freyhof, J. (2015). *Garra mondica*, a new species from the Mond River drainage with remarks on the genus *Garra* from the Persian Gulf basin in Iran (Teleostei: Cyprinidae). *Zootaxa*, *4048*(1), 75–89.

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology*, *55*(1), 421–438. https://doi.org/10.1146/annurev-ento-112408-085432

Sebastien, P. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*, 608–627. https://doi.org/10.1111/1755-0998.12512

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. F. (2004). Micro-Checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vasilyan, D., & Carnevale, G. (2013). The Afro-Asian labeonine genus *Garra* Hamilton, 1822 (Teleostei, Cyprinidae) in the Pliocene of Central Armenia: Palaeoecological and palaeobiogeographical implications. *Journal of Asian Earth Sciences*, *62*, 788–796. https://doi.org/10.1016/j.jseaes.2012.11.033

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Wright, S. (1978). *Evolution and the genetics of populations, volume 4. Variability within and among natural populations*. University of Chicago Press.

Xia, X. (2017). DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. *Journal of Heredity*, *108*(4), 431–437. https://doi.org/10.1093/jhered/esx033

Zachos, F. E., & Habel, J. C. (2011). *Biodiversity hotspots. Distribution and protection of conservation priority areas*. Springer. https://doi.org/10.1007/978-3-642-20992-5

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Figure S1. NJ tree based on *CR*-long sequences ($N = 314$).

Figure S2. Mismatch distribution (MMD).

Figure S3. Discriminant Analysis of Principal Components (DAPC).

Figure S4. Sampling sites and clade affiliations.

Table S1. Excel File including three tables.

Table S2. Primer sets used in the present study.

Table S3. General information on analyzed sequences, mean and maximum p-distances, nucleotide and haplotype diversity of *Garra barreimiae*.

Table S4. Mean p-distances between clades based on three mt fragments (*CR*-long, *COI* and *Cytb*).

Table S5. Results of neutrality tests.

Table S6. Pairwise F_{ST} values between the five nc clusters of *Garra barreimiae*.

Alignment S1. *CR*-long alignment (816 bp) comprising sequences of 314 individuals.

Alignment S2. *CR*-short alignment (401 bp) comprising sequences of 481 individuals.

Alignment S3. *COI* alignment (859 bp) comprising sequences of 47 individuals.

Alignment S4. *Cytb* alignment (1,109 bp) comprising sequences of 47 individuals.

How to cite this article: Kirchner S, Sattmann H, Haring E, Victor R, Kruckenhauser L. Hidden diversity—Delimitation of cryptic species and phylogeography of the cyprinid *Garra* species complex in Northern Oman. *J Zool Syst Evol Res*. 2021;59:411–427. https://doi.org/10.1111/jzs.12438