**SAPIENZA UNIVERSITY OF ROME**

**FACULTY OF MATHEMATICAL, NATURAL AND PHYSICAL SCIENCES**

**Department of Environmental Biology**

**Department of Biology and Biotechnologies 'Charles Darwin'**

**PhD in Environmental and Evolutionary Biology**

*Animal Biology Curriculum*

# Genetic and Morphological diversity in *Sarmarutilus rubilio* (Bonaparte, 1837), an Italian endemic freshwater fish

Candidate:
**Gerardo Petrosino**

Tutor:
**Prof. Anna Rita Rossi, Department of Biology and Biotechnologies 'Charles Darwin'**

Co-tutor:
**Dr. Paolo Colangelo, National Research Council, Research Institute on Terrestrial Ecosystems**

XXXV cycle (2019 – 2022)

Title: Genetic and Morphological diversity in *Sarmarutilus rubilio* (Bonaparte, 1837), an Italian endemic freshwater fish

INDEX

**EXTENDED ABSTRACT** ..... **1**

**ANSWERS TO THE REVIEWERS** ..... **6**

**INTRODUCTION** ..... **8**

Intraspecific diversity in freshwater fishes ..... 8

Study area: the Italian Peninsula ..... 9

The target species: *Sarmarutilus rubilio* (Bonaparte, 1837) ..... 10

Scientific gaps in *Sarmarutilus rubilio* and main objectives of the thesis ..... 11

Author's contribution, thesis structure, specific aims and results ..... 12

References ..... 15

**CHAPTER 1 - Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny** ..... **21**

Introduction ..... 22

Materials and Methods ..... 25

Results ..... 27

Discussion ..... 34

Conclusions ..... 39

References ..... 40

Supplementary Materials ..... 45

**CHAPTER 2 - Phenotypic plasticity over genetic diversity: ecomorphological patterns revealed in the eurytopic and threatened Italian endemic freshwater fish *Sarmarutilus rubilio* (Bonaparte, 1837)** ..... **59**

Introduction ..... 60

Materials and Methods ..... 63

Results ..... 68

Discussion ..... 74

Conclusions ..... 77

References ..... 78

Supplementary Materials ..... 85

**CHAPTER 3 - Microsatellite polymorphism in *Sarmarutilus rubilio*: insights into the complex phylogeographic history of an Italian endemic freshwater fish** ..... **93**

Introduction ..... 94

Materials and Methods ..... 97

Results ..... 102

Discussion ..... 106

Conclusions and Management Implications ..... 110

References ..... 111

Supplementary Materials ..... 119

**GENERAL CONCLUSIONS** ..... **131**

Phylogenetic and phylogeographic analysis ..... 131

Genetic and environmental contributions to morphological differences ..... 132

Population structure and gene flow ..... 132

Implications for conservation and management ..... 133

## EXTENDED ABSTRACT

Intraspecific diversity, i.e. the genomic and phenotypic diversity present within and among populations, is the basis for species adaptation and diversification, thus understanding its magnitude and the mechanisms that originate and maintain it, is fundamental for the conservation of biodiversity.

Such diversity is particularly high in freshwater fishes, as severe limitations to populations' gene flow, between different hydrographic networks or even within the same basin, are responsible for high intraspecific genetic differentiation. In addition in these taxa, several mechanisms of morphological adaptations are known to generate phenotypic variations across populations inhabiting different environments within species' ranges.

Considering that the Mediterranean area is a hotspot of freshwater fish diversity with many threatened species, I decided to investigate the genetic and morphological diversity within the South European roach *Sarmarutilus rubilio*, an endemic Italian species that is currently classified as Vulnerable by the International Union for the Conservation of Nature. Due to its wide ecological niche, this species is present in several basins, which are characterized by different environmental features and past hydrogeological history; so *S. rubilio* was considered a good model for testing the existence of still undetected intraspecific diversity, the knowledge of which will help future conservation interventions for this species.

## Chapter 1 - Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny

Italy hosts 27 endemic and sub-endemic freshwater fish species and their origin and distribution is strongly related to its complex paleogeographic and paleoclimatic history: tectonic movement and sea-level variations during Miocene and Pleistocene promoted genetic diversification and allopatric speciation, and mountainous ridges isolated native species from the postglacial expansions of Eastern-European ones. Geological events determined different fish species assemblages, geographically localized within the Italian peninsula, so that three distinct ichthyogeographic districts were recognized, named Padany-Venetian (PV),

Tuscany-Latium (TL), and Apulia-Campania (AC), but their borders are still not clearly defined. The South European roach *Sarmarutilus rubilio* inhabits various freshwater environments across the three Italian districts, and sightings of this species were reported also in Albanian lakes. Thus its distribution could likely disguise a strong genetic structure, which was also suggested by preliminary data that evidenced high intraspecific clusters divergence.

Starting with these assumptions, I investigated both the complete mitochondrial control region (CR) in almost two hundred specimens from the whole species range and the barcoding portion of the cytochrome oxydase I (COI) in a subset of Italian and Albanian *Sarmarutilus*-like individuals, to test the presence of different South European roach genetic clusters in Italy and assess the relationship between Italian and Albanian fishes. For the latter purpose comparisons with COI sequences from almost all other roach species (which inhabit the basins from Central Europe to Ponto-Caspian region) were also performed to better assess their taxonomical status. In addition, I analyzed two nuclear markers (Cyprinid formerly unknown nuclear Polymorphism – Cyfun P, and 1° intron of the ribosomal protein S7 – RpS7) in a few Italian specimens showing a highly divergent COI profile to obtain further insights about *S. rubilio* phylogeography.

The results highlighted a marked genetic divergence between *S. rubilio* and all other roach species, including Albanian individuals, and among Italian samples, they revealed the existence of three deeply divergent geographic haplogroups (lineages), named A, B and C, whose distribution only partially matched the extension of Italian ichthyogeographic districts. Haplogroup C likely corresponds to a new putative cryptic species (its status is supported by divergence in both mitochondrial and nuclear markers despite the lack of morphological differences) and is located exclusively at the northern border of the South European roach range (Magra-Vara basin); haplogroup B is restricted to South Italy and fixed in the Fondi plain; haplogroup A is widespread across the entire range, both in Tyrrhenian and Adriatic slope of Italy, and in some sites, it is in co-occurrence with C or B. These lineages likely originated in allopatry, as a consequence of the tectonic uplifting of the Apuan Alps in the north and of the Colli Albani Volcano in the south during the Pleistocene, which promoted isolation and vicariance. Moreover, subsequent secondary contacts through sea level variations and even river captures along the Apennine Mountains were responsible for the observed lineage coexistence in some of the investigated sites. Finally, considering the overall complex

genetic structure of *S. rubilio* and the genetic uniqueness of populations in Magra Vara basin (where lineage C was found) and Fondi plain in Southern Lazio (where lineage B was fixed), I suggested focusing conservation interventions on these two geographic areas and to avoid translocation of *Sarmarutilus* between different basins for sustaining the declining ones.

## Chapter 2 - Phenotypic plasticity over genetic diversity: ecomorphological patterns revealed in the eurytopic and threatened Italian endemic freshwater fish *Sarmarutilus rubilio* (Bonaparte, 1837)

Correct assessment of diversity among freshwater fishes, which is crucial for their conservation, could be a challenging task. Indeed numerous cases of cryptic species are reported, i.e. species that can be recognized and discriminated only by molecular approaches, and this can lead to underestimation of the actual diversity; on the other hand, several taxa show intraspecific morphological diversity across their geographic range, as a response faced by different populations to changes in environmental features (i.e. phenotypic plasticity), and misinterpreting this phenomenon can cause species inflation. Moreover genetic and morphological diversity may be correlated, especially when differences in phenotypic traits are genetically fixed between diverging populations. Within the South European roach *Sarmarutilus rubilio*, a freshwater fish endemic of the Italian peninsula, three different genetic lineages (named A, B, C) with distinct geographic distributions were found; lineage C divergence from the others was high suggesting it could belong to a new species, but no morphological differences were observed. In addition, the South European roach is an omnivorous species with a broad ecological niche: it can be found in different freshwater ecosystems, from streams to lakes and reservoirs, from sea level to 1250 m above the sea; therefore any attempt to investigate morphological differences among lineages should take into account the putative effect of phenotypic plasticity.

In this study, I analysed 202 *S. rubilio* specimens from 13 sampling sites (12 lotic and 1 lentic freshwater environments), to assess the degree of body shape variation in this species and test the existence of morphological differences associated with different genetic backgrounds, i.e. between lineages, and/or

environmental parameters. In detail geometric morphometrics analysis was performed to evaluate morphological differences in fishes' body shape, using seventeen landmarks identified on specimens' left side. Genetic diversity was assessed using the mitochondrial Control Region (CR), while differences among sampling sites were defined using measures of seventeen environmental parameters, such as site elevation, stream depth, percentage of pools and vegetation in the investigated watercourse.

Low statistical support was found for the relationship between morphological and genetic differences, anyway lineage B specimens showed deepest body shapes than those belonging to A and C, and no morphological differences were observed between the latter two, despite their high genetic divergence. Conversely, the correlation between morphological and overall environmental parameters was more relevant and revealed some interesting outcomes. Streamlined body shapes were observed in sites scarcely altered by human intervention and with fast water flow, and on the opposite deeper body shapes were observed in canals and one reservoir with slow/still water flow. This is consistent with the "shape-water velocity" patterns observed in other fish species: indeed streamlined body shapes are suited for swimming in high water flow, as they minimize drag; conversely, deep body shapes are fitted for complicated locomotor patterns where water flow is slower. In conclusion, the results suggested morphological diversity in the South European roach is mainly the result of phenotypic plasticity in response to hydrodynamic patterns rather than of different genetic backgrounds, and in particular that water velocity may be the major driver in determining morphological responses. Knowledge about *S. rubilio* intraspecific morphological variation will be useful to predict the effects of habitat alterations (e.g. water abstraction and climatic changes) on lineages' and populations' survival and thus for their sustainable management and conservation.

## Chapter 3 - Microsatellite polymorphism in *Sarmarutilus rubilio*: insights into the complex phylogeographic history of an Italian endemic freshwater fish

In freshwater fish species microsatellite loci are very useful in disentangling genetic structure. Due to their high mutation rate, they can provide insights about recent isolation or contemporary gene flow between populations and complement mitochondrial investigations, which generally better depict historical

processes. Phylogeographic investigations in *Sarmarutilus rubilio*, based on mitochondrial markers analysis, revealed that three different lineages originated in allopatry during the Pleistocene and that they currently coexist in some sites, both in northernmost (Magra-Vara basin) and southernmost (South Italy) areas of this species' native range. Moreover, open questions remained, i.e. whether the observed coexistence was the result of a past temporary connection between basins, or the consequence of human-mediated translocation of specimens in the last centuries, and whether further genetic subdivision exists within lineages for populations belonging to different basins. Therefore I further investigated the genetic structure of *S. rubilio* using 9 microsatellite loci in the same populations previously characterized for the mitochondrial markers. The results revealed strong correlation with previously observed mitochondrial genetic structure, and also highlighted differentiation between populations where two lineages coexist and the others, suggesting their independent evolution after the occurred temporary connection between basins. The strongest differentiation was observed for populations from a restricted area between Central and South Italy (i.e. Fondi plain) where lineage B was fixed, likely due to their long-term isolation. Alleles strictly associated with lineage B were found, but none was conversely observed for lineage C. Thus its putative cryptic species status, highlighted by previous phylogeographic investigation, was not supported. As expected further genetic substructures were revealed also at the interbasin level within lineage A; noteworthy the lack of differentiation between the Tiber River (in the Tyrrhenian slope) and Tronto River (in the Adriatic one) isolated by the Apennine chain could be likely due to the translocations of specimens.

Finally, microsatellite investigations highlighted the genetic uniqueness of populations in the Magra-Vara basin and Fondi plain, which have limited geographic extension, and therefore they may be more susceptible to habitat alterations due to climate change or human activities, or to other threats such as the introduction of allochthonous species. It is highly recommended to consider these populations as distinct Management Units and prioritize them in conservation actions required by the European Habitats Directive for *S. rubilio.*

## ANSWERS TO THE REVIEWERS

In this section I reported the main changes that I made in the first version of the thesis according to the reviewers' suggestions, for which I would like to really thank them.

In the Introduction section, I provided further information about factors affecting neutral genetic diversity and fixed misleading sentences about freshwater fish diversity in the Mediterranean basin, also providing new references. In addition, I clarified my contribution to the collection of data and the presentation of general information.

Since the first chapter was already published and found to be well done by the reviewers, there was no need for changes. However I appreciated hints by Dr. Fruciano about testing the hypothesis of genetic structure, and that will be very useful for the preparation of the paper on microsatellites and future studies.

In the second chapter, I checked out all the suggestions, which were here provided only by Dr. Fruciano: I recognized that the correlation between genetic and morphological distances had to be calculated using sampling site means rather than the pairwise comparisons between specimens, and that disparity in variation between different genetic lineages had to be corrected removing the effect of the sampling site. While the results didn't changed for genetic-morphological distances (i.e. again no correlation between them was found), the new disparity test rejected the significant differences found in the first version of the thesis, thus being also more congruent with the other outcomes of this chapter. I also double-checked other minor issues raised by the reviewer with further recalculations, that confirmed the observed results and thus demonstrated the robustness of the methods used.

In the third chapter minor issues were indicated, regarding the redundant hypothesis of genetic diversity (AMOVA) and missing values (NMDS): these were easily and rapidly addressed. On the other hand, also considering that additional microsatellite loci analysis is undergoing (to reach a higher number of lociand submit reliable and robust results for publication), the reviewer suggested to re-use quantitatively the data from the first two chapters, to improve the quality of the paper. I began this analysis and I tested the correlation between microsatellites and mitochondrial genetic structures (I added this result). Coupling

microsatellite and morphological data turned out to be more complicated. I explored the correlation between the two datasets using PLS analysis, but the results were likely biased by the strong microsatellite genetic structure. Other analysis considered not strictly necessary for the purpose of completing/defending the PhD by the reviewers (like testing the eventual correlation between loci putatively under selection and environmental factors) will be performed using linear models, when all the microsatellite genotypes will be available, and will be included in the publication.

The final correction was made in the conclusions section according to the suggestion of Dr. Sabatini, i.e. making the indications of conservation and management more consistent with the context of the thesis.

# INTRODUCTION

## Intraspecific diversity in freshwater fishes

Intraspecific diversity, i.e. the genomic and phenotypic diversity found within and among populations, is the basis for species adaptation and diversification. Mechanisms that originate and maintain it across and within taxa are therefore fundamental for the full understanding of evolutionary models (Fruciano *et al.*, 2016; Salisbury *et al.*, 2020) and the conservation of biodiversity (Caballero & García-Dorado, 2013; Forsman & Wennersten, 2016; Mimura *et al.*, 2017).

In freshwater fishes, the amount and distribution of intraspecific neutral genetic diversity (i.e. that which has no direct effect on individual fitness) differ across geographic regions, as they are mainly linked to past hydrogeological processes, i.e. geological evolution of geographic regions that affected the connections between hydrogeographic basins, determining populations' isolation and genetic divergence, migrations, secondary contact events as well as present gene flow (Won, Jeon, & Suk, 2020; Perea *et al.*, 2021; MacGuigan, Orr, & Near, 2023). Furthermore, distinct patterns of genetic structure can be identified between species, as the result of the interactions between (hydro)geological events and species-specific characteristics, like life-history traits (Martinez, Willoughby & Christie, 2018), dispersal capabilities and habitat preferences (Pilger *et al.*, 2017; Hu et al., 2022), as well as stochastic processes, like genetic drift in small populations (Pavlova et al., 2017).

In addition to intraspecific genetic diversity, phenotypic variation within freshwater fish species is often observed across their geographic range. This could be the consequence of diverging characteristics fixed through genetic drift (Adkison, 1995; Gallant *et al.*, 2011) or natural selection, the latter providing local adaptations of populations that evolved in different environments (Hendry, Taylor, & McPhail, 2002; Jeukens & Bernatchez, 2012). On the other hand observed variations may be the response faced by different populations to changes in environmental features, i.e. phenotypic plasticity (Langerhans & Makowicz, 2009; Kelley *et al.*, 2017; Kristjánsson *et al.*, 2018; Lema *et al.*, 2019).

## Study area: the Italian Peninsula

Mediterranean basin hosts 460 native freshwater fish species 292 of which are endemic, as recorded by Tierno de Figueroa *et al.* in 2013, and since then several new species were recognized (Denys *et al.*, 2014; Mangit & Yerli, 2018; Lorenzoni *et al.*, 2021). Such diversity is the result of millions of years of complex paleogeographic and paleoclimatic events: tectonic movement in Miocene, sea-level variations occurred during Messinian Salinity Crisis and Pleistocene glacial/interglacial periods, promoted genetic diversification and allopatric speciation across Mediterranean peninsulas (Perea *et al.*, 2010; Seifertová *et al.*, 2012); afterwards, mountainous ridges (e.g. Alps, Pyrenees) isolated native freshwater fishes from the post-glacial expansion of East-European species (Lévêque *et al.*, 2008).

Italy hosts 61 native species, 27 of which are endemic or sub-endemic (AIIAD, 2021). Their distribution is not homogeneous across the Italian peninsula and most of them inhabit only a few adjacent basins. Based on their distinct species assemblage three different ichthyogeographic districts were recognized: Padany-Venetian (PV, i.e. Po valley and northern Adriatic basins), Tuscany-Latium (TL, i.e. Thyrrenian basins from Eastern Liguria to Tiber River), and Apulia-Campania (AC, i.e. approximately Tyrrhenian and Adriatic basins of South Italy) (Bianco, 1995, 2014).

Such diversity is threatened by the introduction of invasive species, whose number equals that of the native ones (AIIAD, 2021), and by other anthropogenic impacts such as habitat degradation, water pollution and abstraction (Rondinini, Battistoni, & Teofili, 2022); moreover, impending climate changes are expected to increase the natural high seasonal fluctuations of water flow of Mediterranean rivers, posing another threat to the survival of fish populations (Vardakas *et al.*, 2017).

In this context data about intraspecific genetic and morphological diversity are fundamental for identifying populations with unique characteristics across species ranges and making predictions about their response to habitat alterations, which is the background for planning effective management and conservation interventions (Berbel-Filho *et al.*, 2016; Kaus *et al.*, 2019; Potts, Mandrak, & Chapman, 2021; Finger *et al.*, 2022).

## The target species: *Sarmarutilus rubilio* (Bonaparte, 1837)

The target species of this study is the South European roach *Sarmarutilus rubilio* (Bonaparte, 1837), the unique species of the genus, classified as *Rutilus rubilio* until the identification of significant morphological and genetic differences with the other species of the *Rutilus* genus (Bianco & Ketmaier, 2014). It's a small-medium sized freshwater fish species, belonging to the Leuciscidae family - previously included in the Cyprinidae, see (Schönhuth *et al.*, 2018) - endemic to the Italian peninsula: it is distributed from North-Central to Southern Italy, with most of the population localized in basins of the Tyrrhenian slope (from River Magra North to River Sele South, i.e. TL and AC district) and some in the Adriatic basin (from River Tronto in the North to River Trigno in the South, i.e. PV and AC district) (Bianco & Ketmaier, 2014). Therefore, although it is mainly present in the TL ichthyogeographic district, its range includes also the other two districts. Moreover, *S. rubilio* was translocated in many basins of Sicily and the southern part of the Italian Peninsula (e.g. Calabria region) (Bianco *et al.*, 2013) so the border between the native and introduced ranges remains unclear. The South European roach was translocated even in Tunisia (Djait *et al.*, 2019) and Turkey (Keskin, Unal, & Atar, 2016), and in the past decades there were also claims about this species (or subspecies) being present in Albanian lakes (Marić, 2010), although the taxonomical status of these fishes is still debated (Milošević *et al.*, 2011; Bianco & Ketmaier, 2014; Geiger *et al.*, 2014).

The South European roach is an omnivorous species with a broad ecological niche (Balestrieri *et al.*, 2006) and it could be found in still waters (both natural lakes and artificial dam reservoirs), small rivers, streams and canals, on substrates with rocks, pebbles, sand and gravel, but also muddy and rich in submerged vegetation, mainly in lowland and hilly areas, though some populations were also found at 1250 m above sea level (Bianco *et al.*, 2013; Di Tizio & Di Felice, 2016).

Despite being a eurytopic species, the number and size of its populations are declining and its range reduced by over 30% in 10 years (Bianco *et al.*, 2013), mainly due to the introduction of allochthonous species, such as the morphologically and ecologically similar roaches *Leucos aula* and *Rutilus rutilus* (Di Tizio & Di Felice, 2016). Indeed when I started working on this species, in 2019, *S. rubilio* was still classified in the Near-Threatened IUCN risk assessment category (Bianco *et al.*, 2013), while recently its conservation status

has been updated to Vulnerable (Rondinini *et al.*, 2022). Finally, the species is listed in Annex II of the European Union Habitats Directive, which includes species of community interest whose conservation requires the designation of special areas of conservation (EEC, 1992).

## Scientific gaps in *Sarmarutilus rubilio* and main objectives of the thesis

Considering that the South European roach inhabits basins across three different ichthyogeographic districts, which evolved mostly independently of each other, and that the genetic structure of freshwater fishes is strongly dependent on past geological, climatic and hydro-morphological events, the distribution of this species could likely disguise a strong genetic differentiations across its native range.

Indeed a previous barcoding study on freshwater fish taxa across the Mediterranean area, listed *S. rubilio* among those species characterized by high intraspecific diversity (Geiger *et al.*, 2014), but since then no further investigations explored this outcome. Moreover, despite morphological variations between *S. rubilio* populations reported by local fishermen (Tancioni 2019, personal communication), no data in the literature was found about this topic.

Bridging the knowledge gap on intraspecific diversity in the South European roach, a threatened species will thus provide insights into the role of geological events and environmental features in shaping divergence across its range. These data are fundamental for the identification of populations with unique characteristics and thus for defining management units for conservation (Allendorf, 2017).

In detail, the aims of this research are:

- To identify the genetic structure of the species at different geographical scales, i.e. test the existence of distinct and geographically localized evolutionary lineages and assess the amount of genetic differentiation between populations, thus the effects of historical and recent events in shaping it **(Chapters 1 and 3)**
- To test the existence of morphological differentiation between populations and assess whether it has a genetic base and/or is mainly due to the interaction between sampling sites' environmental parameters and specimens' phenotypic plasticity **(Chapter 2)**

- On the bases of previous points, to provide tips for species conservation **(Chapters 1, 2 and 3, see also General conclusions)**

## Author's contribution, thesis structure, specific aims and results

During the three years of my doctorate course, I collected on the field tissues and photos for about three hundred specimens of *S. rubilio*, thanks to the collaboration with Prof. Lorenzo Tancioni from the Department of Biology, University of Rome "Tor Vergata", who provided the equipment and the know-how for sampling activities. Subsequently, I performed all the analyses reported in the thesis, with the fundamental help of my tutors, Prof. Anna Rita Rossi and Dr. Paolo Colangelo, which gave me suggestions about the most appropriate approaches for data investigations and supervised the writing of the thesis/scientific papers. Finally, master's degree students Annarita Ricci and Camilla Polinori helped me with laboratory activities. The thesis is structured into three chapters, as most of my investigations were organized as scientific papers: the first was already published by the journal *Genes*, the second has been submitted to the *Biological Journal of the Linnean Society*, and the third needs additional data to be submitted ( i.e. results from four further microsatellite loci and additional analyses, which are currently undergoing to provide more reliable results).

**Chapter 1** – using mitochondrial (COI and CR) and nuclear (CyfunP and S7) markers to test the existence of different genetic groups within *S. rubilio* and make inferences about the species' phylogeography. Three intraspecific divergent lineages were found, one being a putative new cryptic species; they likely originated in allopatry, subsequently, secondary contact events between them occurred in some sites.

**Chapter 2** – geometric morphometrics analysis was performed to assess the amount of body shape variation between specimens and populations. To investigate the contribution of genetic and environmental features, the possible correlation of data with CR lineages and seventeen different environmental parameters was also assessed. The data suggested that morphological diversity in the South European roach is mainly the result of phenotypic plasticity in response to different environmental drivers (i.e. lotic vs lentic hydrodynamic patterns) rather than to different genetic backgrounds.

**Chapter 3** – specimens were genotyped at nine microsatellite loci to assess the amount of genetic differentiation between populations, and make inferences about recent isolation and divergence events and/or ongoing gene flow among them. Results mostly confirmed the outcomes of mitochondrial analysis (Chapter 1), however, the existence of a cryptic species was not confirmed by microsatellites. On the contrary, typical alleles were associated with one of the other lineages. Data revealed that secondary contact between lineages didn't occur in recent times and highlighted further genetic substructure between geographically close basins. Data on four further microsatellite loci will be added and a new analysis combining microsatellites diversity with morphological data from the previous chapter will be performed, to refine and deepen the outcomes.

Considering that the number of sites and specimens analyzed was mostly the same but slightly differed among the three chapters to best achieve their specific aims and adapt to the ongoing analysis and outcomes, in the following table there is a summary of the markers/approaches and number of specimens for each sampling site

**Summary Table of sites and *S. rubilio* specimens analysed with the different main approaches.**

In the mtDNA column: in black data reported in Chapter 1, in red new data obtained and reported in Chapter 2 for comparison with geometric morphometric analysis.

| Drainage basin | Sampling site | mtDNA-CR | Geometric Morphometrics | Microsatellites |
|---|---|---|---|---|
| Magra-Vara | VAR1 | 13 | -- | -- |
| Magra-Vara | VAR2 | 20 + 1 | 21 | 20 |
| Tronto | TRO (TRO2 in chapter 2) | 18 | 12 | 15 |
| Tronto | TRO1 | 6 | 6 | -- |
| Foro | FOR (FOR2 in chapter2) | 16 | 9 | 17 |
| Foro | FOR1 | 7 | 7 | -- |
| Arrone | ARR | 15 | 15 | 20 |
| Tiber | TIB1 | 33 | 31 | -- |
| Tiber | TIB2 | 4 + 6 | 6 | 18 |
| Tiber | TIB3 | 27 | 27 | 25 |
| Fondi | SET | 20 | 20 | 20 |
| Fondi | FON | 16 | 16 | 15 |
| San Puoto | SPU | 7 | -- | -- |
| Santa Croce | SCR | 32 | 8 | 22 |
| Liri-Garigliano | GAR | 24 | 24 | 18 |
| Noce | NOC | 6 | -- | 6 |

## References


**Adkison MD**. **1995**. Population differentiation in Pacific salmons: local adaptation genetic drift, or the environment? *Canadian Journal of Fisheries and Aquatic Sciences* **52**: 2762–2777.

**AIIAD**. **2021**. Principi guida riguardanti le immissioni di fauna ittica nelle acque interne italiane. : 1–31.

**Allendorf FW**. **2017**. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* **26**: 420–430.

**Balestrieri A, Prigioni C, Remonti L, Sgrosso S & Priore G**. **2006**. Feeding ecology of *Leuciscus cephalus* and *Rutilus rubilio* in southern Italy. *Italian Journal of Zoology* **73**: 129–135.

**Berbel-Filho WM, Martinez PA, Ramos TPA, Torres RA & Lima SMQ**. **2016**. Inter- and intra-basin phenotypic variation in two riverine cichlids from northeastern Brazil: potential eco-evolutionary damages of São Francisco interbasin water transfer. *Hydrobiologia* **766**: 43–56.

**Bianco PG**. **1995**. Factors Affecting the Distribution of Freshwater Fishes Especially in Italy. *Cybium*.

**Bianco PG, Caputo V, Ferrito V, Lorenzoni M, Nonnis Marzano F, Stefani F, Sabatini A & Tancioni L. 2013**. *Rutilus rubilio*. In: Rondinini C, Battistoni A, Peronace V, Teofili C, eds. *Lista Rossa IUCN dei Vertebrati Italiani.* . Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare. http://www.iucn.it/scheda.php?id=-251974278.

**Bianco PG**. **2014**. An update on the status of native and exotic freshwater fishes of Italy. *Journal of Applied Ichthyology* **30**: 62–77.

**Bianco PG & Ketmaier V**. **2014**. A revision of the *Rutilus* complex from Mediterranean Europe with description of a new genus, *Sarmarutilus*, and a new species, *Rutilus stoumboudae* (Teleostei: Cyprinidae). *Zootaxa* **3841**: 379–402.

**Caballero A & García-Dorado A**. **2013**. Allelic Diversity and Its Implications for the Rate of Adaptation. *Genetics* **195**: 1373–1384.

**Denys GPJ, Dettai A, Persat H, Hautecœur M & Keith P. 2014.** Morphological and molecular evidence of three species of pikes *Esox spp.* (Actinopterygii, Esocidae) in France, including the description of a new species. *Comptes Rendus Biologies* **337(9)**: 521-534,

**Djait H, Bahri-Sfar L, Laouar H, Missaoui N & Ben Hassine OK**. **2019**. Dietary comparison of pike-perch, *Sander lucioperca* (Linnaeus, 1758) and catfish, *Silurus glanis* Linnaeus, 1758 in Sidi Salem dam reservoir (Tunisia). *Cybium* **43**: 61–69.

**EEC**. **1992**. Council Directive 92/43/ECC. *Official Journal of the European Union* **94**: 40–52.

**Finger AJ, Benjamin A, Crookshanks C, Campbell MA & Sağlam İK**. **2022**. Broad- and fine-scale structure across the distribution of the relict dace (*Relictus solitarius*) in the Great Basin desert, USA. *Conservation Science and Practice* **4**.

**Forsman A & Wennersten L**. **2016**. Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography* **39**: 630–648.

**Fruciano C, Franchini P, Kovacova V, Elmer KR, Henning F & Meyer A**. **2016**. Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nature Communications* **7**: 12736.

**Gallant JR, Arnegard ME, Sullivan JP, Carlson BA & Hopkins CD**. **2011**. Signal variation and its morphological correlates in *Paramormyrops kingsleyae* provide insight into the evolution of electrogenic signal diversity in mormyrid electric fish. *Journal of Comparative Physiology A* **197**: 799–817.

**Geiger MF, Herder F, Monaghan MT, Almada V, Barbieri R, Bariche M, Berrebi P, Bohlen J, Casal-Lopez M, Delmastro GB, Denys GPJ, Dettai A, Doadrio I, Kalogianni E, Kärst H, Kottelat M, Kovačić M, Laporte M, Lorenzoni M, Marčić Z, Özuluğ M, Perdices A, Perea S, Persat H, Porcelotti S, Puzzi C, Robalo J, Šanda R, Schneider M, Šlechtová V, Stoumboudi M, Walter S & Freyhof J**. **2014**. Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes. *Molecular Ecology Resources* **14**: 1210–1221.

**Hendry AP, Taylor EB & McPhail JD**. **2002**. Adaptive divergence and the balance between selection and

gene flow: lake and stream stickleback in the Misty system. *Evolution* **56**: 1199–1216.

**Hu J, Zhao J, Sui X, Zhu R & He D. 2023.** Across the highest mountain on earth: discordant phylogeographic patterns and recent dispersal of Tibetan stone loaches (*Triplophysa*) in the Himalayas. *Journal of Fish Biology*, 1–13.

**Jeukens J & Bernatchez L**. **2012**. Regulatory versus coding signatures of natural selection in a candidate gene involved in the adaptive divergence of whitefish species pairs (*Coregonus spp.*). *Ecology and Evolution* **2**: 258–271.

**Kaus A, Michalski S, Hänfling B, Karthe D, Borchardt D & Durka W**. **2019**. Fish conservation in the land of steppe and sky: Evolutionarily significant units of threatened salmonid species in Mongolia mirror major river basins. *Ecology and Evolution* **9**: 3416–3433.

**Kelley JL, Davies PM, Collin SP & Grierson PF**. **2017**. Morphological plasticity in a native freshwater fish from semiarid Australia in response to variable water flows. *Ecology and Evolution* **7**: 6595–6605.

**Keskin E, Unal EM & Atar HH**. **2016**. Detection of rare and invasive freshwater fish species using eDNA pyrosequencing: Lake Iznik ichthyofauna revised. *Biochemical Systematics and Ecology* **67**: 29–36.

**Kristjánsson BK, Leblanc CAL, Skúlason S, Snorrason SS & Noakes DLG**. **2018**. Phenotypic plasticity in the morphology of small benthic Icelandic Arctic charr (*Salvelinus alpinus*). *Ecology of Freshwater Fish* **27**: 636–645.

**Langerhans RB & Makowicz AM**. **2009**. Shared and unique features of morphological differentiation between predator regimes in *Gambusia caymanensis*. *Journal of Evolutionary Biology* **22**: 2231–2242.

**Lema SC, Bock SL, Malley MM & Elkins EA**. **2019**. Warming waters beget smaller fish: evidence for reduced size and altered morphology in a desert fish following anthropogenic temperature change. *Biology Letters* **15**: 20190518.

**Lévêque C, Oberdorff T, Paugy D, Stiassny MLJ & Tedesco PA**. **2008**. Global diversity of fish (Pisces) in

freshwater. *Hydrobiologia* **595**: 545–567.

**Lorenzoni M, Carosi A, Quadroni S, De Santis V, Vanetti I, Delmastro GB & Zaccara S. 2021.** Cryptic diversity within endemic Italian barbels: revalidation and description of new *Barbus* species (Teleostei: Cyprinidae). Journal of Fish Biology **98**: 1433–1449.

**MacGuigan DJ, Orr OD & Near TJ**. **2023**. Phylogeography, hybridization, and species discovery in the *Etheostoma nigrum* complex (Percidae: Etheostoma: Boleosoma). *Molecular Phylogenetics and Evolution* **178**: 107645.

**Mangit F & Yerli SV. 2018.** Systematic evaluation of the genus *Alburnus* (Cyprinidae) with description of a new species. *Hydrobiologia* **80:**, 297–312.

**Marić D**. **2010**. *Rutilus albus* sp. n. (Teleostei: Cyprinidae) from Lake Skadar. *Periodicum Biologorum* **112**: 153–158.

**Martinez AS, Willoughby JR & Christie MR. 2018.** Genetic diversity in fishes is influenced by habitat type and life‐history variation. *Ecology and Evolution* **8**: 12022‐12031.

**Milošević D, Winkler KA, Marić D & Weiss S**. **2011**. Genotypic and phenotypic evaluation of *Rutilus spp.* from Skadar, Ohrid and Prespa Lakes supports revision of endemic as well as taxonomic status of several taxa. *Journal of Fish Biology* **79**: 1094–1110.

**Mimura M, Yahara T, Faith DP, Vázquez-Domínguez E, Colautti RI, Araki H, Javadi F, Núñez-Farfán J, Mori AS, Zhou S, Hollingsworth PM, Neaves LE, Fukano Y, Smith GF, Sato YI, Tachida H & Hendry AP**. **2017**. Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications* **10**: 121–139.

**Myers N, Mittermeler RA, Mittermeler CG, Da Fonseca GAB & Kent J**. **2000**. Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858.

**Pavlova A, Beheregaray LB, Coleman R, Gilligan D, Harrisson KA, Ingram BA, Kearns J, Lamb AM,**

**Lintermans M, Lyon J, Nguyen TTT, Sasaki M, Tonkin Z, Yen JDL & Sunnucks P. 2017.** Severe consequences of habitat fragmentation on genetic diversity of an endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications* **10**: 531– 550.

**Perea S, Böhme M, Zupančič P, Freyhof J, Šanda R, Özulu M, Abdoli A & Doadrio I**. **2010**. Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evolutionary Biology* **10**: 1–27.

**Perea S, Sousa-Santos C, Robalo J & Doadrio I**. **2021**. Historical biogeography of the Iberian Peninsula: multilocus phylogeny and ancestral area reconstruction for the freshwater fish genus *Squalius* (Actinopterygii, Leuciscidae). *Journal of Zoological Systematics and Evolutionary Research* **59**: 858–886.

**Pilger TJ, Gido KB, Propst DL, Whitney JE & Turner TF. 2017.** River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. *Molecular Ecology* **26**: 2687-2697.

**Potts LB, Mandrak NE & Chapman LJ**. **2021**. Coping with climate change: Phenotypic plasticity in an imperilled freshwater fish in response to elevated water temperature. *Aquatic Conservation: Marine and Freshwater Ecosystems* **31**: 2726–2736.

**Rondinini C, Battistoni A & Teofili C (Eds.)**. **2022**. *Lista rossa IUCN dei vertebrati italiani 2022*. Roma: Comitato Italiano IUCN e Ministero dell'Ambiente e della Sicurezza Energetica.

**Salisbury SJ, McCracken GR, Perry R, Keefe D, Layton KKS, Kess T, Nugent CM, Leong JS, Bradbury IR, Koop BF, Ferguson MM & Ruzzante DE**. **2020**. Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Molecular Ecology* **29**: 4280–4294.

**Schönhuth S, Vukić J, Šanda R, Yang L & Mayden RL**. **2018**. Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). *Molecular Phylogenetics and Evolution* **127**: 781–799.

**Seifertová M, Bryja J, Vyskočilová M, Martínková N & Šimková A**. **2012**. Multiple Pleistocene refugia and post-glacial colonization in the European chub (*Squalius cephalus*) revealed by combined use of nuclear and mitochondrial markers. *Journal of Biogeography* **39**: 1024–1040.

**Tierno de Figueroa JM, López-Rodríguez MJ, Fenoglio S, Sánchez-Castillo P & Fochetti R**. **2013**. Freshwater biodiversity in the rivers of the Mediterranean Basin. *Hydrobiologia* **719**: 137–186.

**Di Tizio L & Di Felice PL**. **2016**. *Rutilus rubilio* (Bonaparte, 1837) (Rovella). In: Stoch F, Genovesi P, eds. *Manuali per il monitoraggio di specie e habitat di interesse comunitario (Direttiva 92/43/CEE) in Italia: specie animali*. ISPRA, 164–165.

**Vardakas L, Kalogianni E, Papadaki C, Vavalidis T, Mentzafou A, Koutsoubas D & Skoulikidis Th. N**. **2017**. Defining critical habitat conditions for the conservation of three endemic and endangered cyprinids in a Mediterranean intermittent river before the onset of drought. *Aquatic Conservation: Marine and Freshwater Ecosystems* **27**: 1270–1280.

**Won H, Jeon HB & Suk HY**. **2020**. Evidence of an ancient connectivity and biogeodispersal of a bitterling species, *Rhodeus notatus*, across the Korean Peninsula. *Scientific Reports* **10**: 1011.

# CHAPTER 1

# Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny

genes

MDPI

*Article*

# Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny

**Gerardo Petrosino [1,*], Lorenzo Tancioni [2], Martina Turani [1], Arnold Rakaj [2], Luca Ciuffardi [3] and Anna Rita Rossi [1]**

[1] Department of Biology and Biotechnology "C. Darwin", Sapienza University of Rome, 00161 Rome, Italy; turanimartina@gmail.com (M.T.); annarita.rossi@uniroma1.it (A.R.R.)
[2] Department of Biology, University of Rome "Tor Vergata", 00133 Rome, Italy; tancioni@uniroma2.it (L.T.); arnoldrakaj@gmail.com (A.R.)
[3] Center for BioNatural Studies srl, 16132 Genova, Italy; info@lucaciuffardi.it
* Correspondence: gerardo.petrosino@uniroma1.it

**Abstract:** Italy hosts a large number of endemic freshwater fish species due to complex geological events which promoted genetic differentiation and allopatric speciation. Among them, the South European roach *Sarmarutilus rubilio* inhabits various freshwater environments in three different ichthyogeographic districts. We investigated the genetic diversity of *S. rubilio* using two different mitochondrial markers (COI and CR), aiming to define its relationship with other similar taxa from the Balkan area and, from a phylogeographic perspective, test the effects of past hydrogeological dynamics of Italian river basins on its genetic structure and demographic history. Our analysis highlighted a marked genetic divergence between *S. rubilio* and all other roach species and, among Italian samples, revealed the existence of three deeply divergent geographic haplogroups, named A, B and C. Haplogroup C likely corresponds to a new putative cryptic species and is located at the northern border of the South European roach range; haplogroup B is restricted to Southern Italy; and haplogroup A is widespread across the entire range and in some sites it is in co-occurrence with C or B. Their origin is probably related to the tectonic uplifting of the Apuan Alps in the north and of the Colli Albani Volcano in the south during the Pleistocene, which promoted isolation and vicariance followed by secondary contacts.

**Keywords:** Italian ichthyogeographic districts; Albanian lakes; *Rutilus*; *Leucos*; freshwater fish; Cyfun P; cytochrome oxidase I; control region; vicariance events; secondary contacts; conservation genetics



**Citation:** Petrosino, G.; Tancioni, L.; Turani, M.; Rakaj, A.; Ciuffardi, L.; Rossi, A.R. Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny. *Genes* **2022**, *13*, 1071. https://doi.org/10.3390/genes13061071

Academic Editor: Ze Zhang

Received: 4 May 2022
Accepted: 13 June 2022
Published: 15 June 2022



## 1. Introduction

In recent decades, the integration of molecular approaches when investigating the diversity and distribution of animal species [1] has provided new insights into the taxonomy and evolution of freshwater fish, facilitating phylogenetic reconstruction [2,3], solving taxonomical controversies [4–6], identifying cryptic species [7–9], detecting the introduction of allochthonous species [10,11] and clarifying the conservation status of threatened species [12]. In addition, a strong relationship emerged between the genetic diversity of freshwater fishes, geography and past hydrogeological events [13–15].

The Mediterranean area is considered a biodiversity hotspot [16] and hosts 25% of the strictly freshwater fish species registered in the Palearctic region [17] due to its complex paleogeographic and paleoclimatic history. Indeed, tectonic movement in the Miocene and sea-level variations which occurred during the Messinian Salinity Crisis and Pleistocene glacial/interglacial periods, promoted genetic diversification and allopatric speciation across Mediterranean peninsulas, as these territories were alternatively linked and isolated from each other, and acted as glacial refugia during the Pleistocene [2,18]. Afterwards,

mountainous ridges (e.g., Alps, Pyrenees) isolated native freshwater fishes from the post-glacial expansion of Eastern-European species [19], thus explaining the presence of the huge number of endemic species in this region (526 nominal and 490 confirmed [20]).

In the Italian peninsula, three different ichthyogeographic districts are currently recognized, named Padany-Venetian (PV), Tuscany-Latium (TL), and Apulia-Campania (AC) (see Figure 1), whose borders are still not clearly defined [21]. These districts were shaped by river confluence and isolation mechanisms during the Pleistocene and host different fish species assemblages [22], corresponding to 61 native species, 27 of which are endemic or sub-endemic [23]. Most of them are threatened by invasive species, which are equal to the number of native ones [23], and by other anthropogenic impacts such as habitat degradation, water pollution and abstraction [24]. Among the existing Italian native species, the Leuciscidae family (previously included in the Cyprinidae, see [25]), is the most represented group with 14 species (23% of native ones). Despite the use of molecular markers providing useful information about these species in the last decade [26–28], genetic data are still required for many of them for effective management, and conservation action plans are still lacking [29].

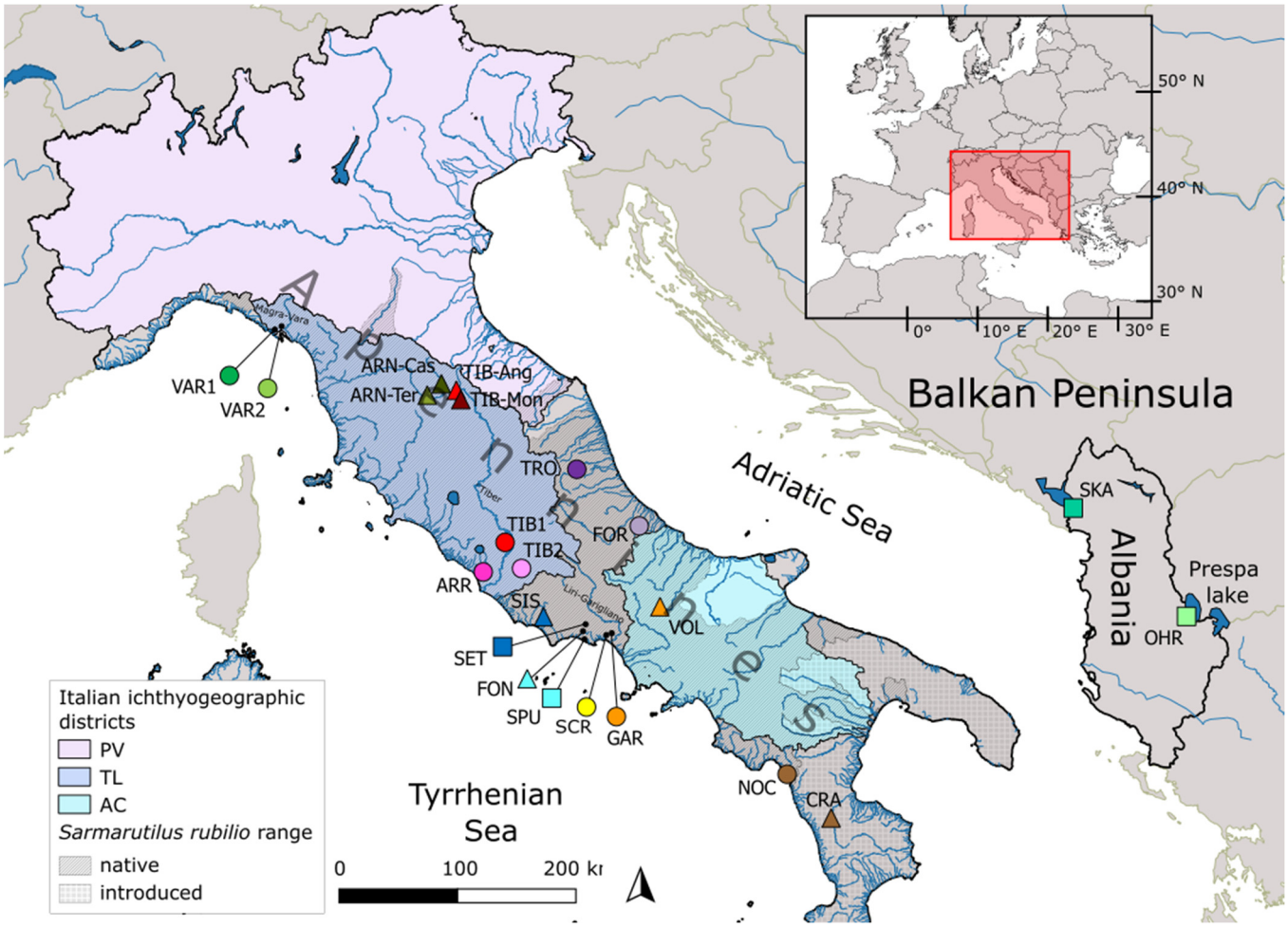



**Figure 1.** Geographic location of the study area (black and red rectangles, top right), along with *S. rubilio* distribution according to IUCN (https://www.iucnredlist.org/species/19786/9014268, accessed on 11 April 2022), and sampling sites: circle indicates rivers and square lakes. For abbreviations refer to Table 1. Triangles indicate sites corresponding to sequences retrieved from the Genbank database and included in the COI phylogenetic reconstruction. Names of major basins discussed in the paper are also indicated.

**Table 1.** Sampling details for twelve populations of South European roach from Italy and two *Rutilus* sp. from Albania and summary of genetic variation for the mitochondrial control region (CR) and cytochrome oxidase I (COI) subset.

| | | | | | | | CR | | | | COI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Drainage Basin** | **River/Lake** | **Pop ID** | **Lat (°N)** | **Lon (°E)** | **Protected Area** | **N** | **Hp (% Private)** | **Hp Rich** | **Hd (±s.e.)** | **π % (±s.e.)** | **N** | **Hp** |
| Italy | | | | | | | | | | | | |
| Magra-Vara | Riccò | VAR1 | 44.161 | 9.758 | | 13 | 3 (0) | 3.00 | 0.51 (±0.14) | 1.51 (±0.81) | 11 | 4 |
| | Graveglia | VAR2 | 44.191 | 9.790 | | 20 | 4 (0) | 3.54 | 0.63 (±0.08) | 1.65 (±0.86) | 8 | 4 |
| Tronto | Tronto | TRO | 42.802 | 13.465 | SAC | 18 | 2 (0) | 1.72 | 0.11 (±0.10) | 0.01 (±0.02) | 3 | 2 |
| Foro | Foro | FOR | 42.246 | 14.186 | | 16 | 2 (50) | 1.98 | 0.23 (±0.13) | 0.03 (±0.03) | 3 | 2 |
| Arrone | Arrone | ARR | 41.914 | 12.265 | | 15 | 4 (50) | 3.73 | 0.60 (±0.11) | 0.09 (±0.08) | 3 | 2 |
| Tiber | Rio Martino | TIB1 | 42.173 | 12.545 | | 33 | 4 (50) | 3.35 | 0.63 (±0.06) | 0.08 (±0.07) | 2 | 2 |
| | Fosso Passerano | TIB2 | 41.932 | 12.732 | | 4 | 3 (0) | NA | 0.83 (±0.22) | 0.11 (±0.11) | 2 | 2 |
| Fondi | Settecannelle | SET | 41.368 | 13.421 | Regional Park | 20 | 5 (60) | 3.84 | 0.44 (±0.13) | 0.05 (±0.05) | 3 | 2 |
| San Puoto | San Puoto | SPU | 41.285 | 13.408 | SPA | 7 | 2 (0) | NA | 0.57 (±0.12) | 0.06 (±0.06) | 7 | 1 |
| Santa Croce | Santa Croce | SCR | 41.287 | 13.716 | SAC | 32 | 5 (40) | 3.47 | 0.60 (±0.07) | 0.21 (±0.13) | 3 | 2 |
| Liri-Garigliano | Ausentello | GAR | 41.303 | 13.743 | | 24 | 3 (0) | 2.46 | 0.30 (±0.11) | 0.04 (±0.04) | 3 | 1 |
| Noce | Pamafi | NOC | 39.934 | 15.752 | | 6 | 2 (50) | NA | 0.33 (±0.22) | 0.58 (±0.38) | 6 | 2 |
| Total | | | | | | 208 | 21 | | | | 54 | 14 |
| Albania | | | | | | | | | | | | |
| Skadar | Skadar | SKA | 42.059 | 19.455 | | 8 | 8 (100) | NA | 1.00 (±0.06) | 0.39 (±0.25) | 8 | 4 |
| Ohrid | Ohrid | OHR | 40.963 | 20.640 | | 7 | 4 (100) | NA | 0.71 (±0.18) | 0.12 (±0.1) | 7 | 3 |
| Total | | | | | | 15 | 12 | | | | 15 | 7 |

Abbreviations: sample size (N), number of haplotypes and percentage of private haplotypes (Hp), haplotype richness (Hp rich), haplotype diversity (Hd), nucleotide diversity (π). Standard errors (s.e.) are given in parentheses for Hd and π. SAC = Special Area of Conservation, according to EC Habitat Directive; SPA = Special Protection Area, according to EC Birds Directive. Datum = WGS84 for geographic coordinates.

The South European roach *Sarmarutilus rubilio* (Bonaparte, 1837)—a unique species of the genus—is a leuscid previously known as *Rutilus rubilio*. Nomenclature change of this taxon was driven by the identification of diagnostic morphological differences scarcely appreciable with the naked eye and by separate clustering when compared to any other genetic lineage included in the former *Rutilus* genus [30]. It derived its name from the Sarmatic Sea, or Lago Mare, an ancient central European inner freshwater sea where it likely originated in the Middle Miocene, between 5 [30] and 9.7 [2] million years ago (Mya) when most of the Italian Peninsula was still under sea level. The South European roach has a broad ecological niche. It is eurytopic, moderately rheophilic and thermophilic, mainly living in hilly and lowland stretches of streams and rivers, and more rarely in lentic environments. It is distributed from North-Central to Southern Italy, with most of the population localized in the basins of the Tyrrhenian slope (from River Magra north to

River Sele south) and some in the Adriatic basin (from River Tronto in the north to River Trigno in the south) according to [30] (see Figure 1 below); thus, its range includes the three different Italian ichthyogeographic districts, although it is mainly found in the TL district. In addition, *S. rubilio* was also translocated in many basins of the southern part of the Italian Peninsula and Sicily—even in Tunisia [31] and Turkey [32]—and the border between the native and introduced ranges still remains unclear. Presently, the number and size of its populations is declining and its range decreased by over 30% in 10 years [33], mainly due to the introduction of allochthonous species, such as the morphologically and ecologically similar roaches *Leucos aula* and *Rutilus rutilus* [34]; thus, even though the species is still classified in the Near-Threatened IUCN risk assessment category, it is likely going to be qualified as Vulnerable in the near future [33] and is listed in Annex II of the European Union Habitats Directive [35]. The presence of the species (or subspecies) was also reported in Albanian lakes [36] as external meristic traits between Italian *S. rubilio* and Albanian roaches overlap; therefore, debates over the taxonomical status of Albanian roach are ongoing [20,30,37]. *Sarmarutilus rubilio* distribution could likely disguise a strong genetic structure, which is also suggested by the evidence that this species is listed among those Mediterranean freshwater fish taxa (about 98% barcoded) characterized by high intraspecific clusters divergence; therefore qualifying it as a potential candidate species, i.e., cryptic species [20] (Appendix S1). Starting with these assumptions, we investigated both the complete mitochondrial control region (CR) in almost two hundred specimens from the whole species range, and the barcoding portion of the cytochrome oxydase I (COI) in a subset of Italian and Albanian *Sarmarutilus*-like individuals. In addition, we analyzed two nuclear markers (Cyprinid formerly unknown nuclear Polymorphism, and first intron of the ribosomal protein S7) in a few Italian specimens showing a highly divergent COI profile. These analyses aimed to (1) assess the relationship between Italian *S. rubilio* specimens from Albania with a similar morphology and other roach species (all species previously included in the genus *Rutilus* and now split among *Leucos*, *Sarmarutilus* and *Rutilus*); (2) test the presence of geographically localized South European roach genetic clusters in Italy; (3) make inferences on how the hydrogeological evolution of Italian peninsula influenced *S. rubilio* lineage diversification. The results will be useful in constructing future conservation action plans for this species and will provide further data on the relationships between Italy's hydrogeological history and the genetic structure of freshwater fishes, thus contributing to the debate aimed at clarifying and eventually rearranging the Italian ichthyogeographic districts' borders, in relation to the divergent processes of other endemic species.

## 2. Materials and Methods

### *2.1. Sampling, Morphological Identification and DNA Extraction*

Overall, 223 specimens were collected, with 208 *S. rubilio* individuals from 12 sites in the Italian peninsula, covering most of the known supposed native range, from the Liguria region at the northern border of species distribution to the southern one in Basilicata. In addition, 15 specimens showing *S. rubilio*-like morphology were retrieved from two Albanian lakes, where this species was also reported in recent decades (Figure 1; Table 1).

Fishes were sampled using electrofishing and anaesthetized with a 0.035% MS 222 (Tricaine Methanesulfonate) solution, in accordance with the relevant legislation (CEN EN 131 14011/2003-Water quality-Sampling of fish with electricity) and standards for species listed in Annex II of Habitats Directive [38], as authorized by Regional Directions responsible for Hunting and Fishing activities and by Directions of Protected areas (Prot. n.: Lazio G10101, 25 July 2019; Abruzzo DPD023/171, 12 April 2021; Marche 213, 13 April 2021; Liguria 5166–2021, 30 August 2021; Aurunci protected area 0002963.U, 01 October 2021). Specimens were morphologically identified according to descriptions in the literature [39]. The tips of pelvic fins were taken from each specimen and fixed in 96% ethanol to further perform genetic analyses. After tissue collection and recovery from the anaesthetic, the specimens were released into the wild. Some entire specimens for each sampling site were preserved in 70% ethanol and deposited in the Museum of Comparative Anatomy "Battista

Grassi", Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome (AC1687-98). Total genomic DNA extraction was carried out using the salting-out method [40].

### 2.2. PCR Amplification

A fragment of approximately 620 base pairs (bp) of the mitochondrial cytochrome c oxidase subunit 1 (COI), widely used for animal species barcoding [41], was amplified in a subset of 2 to 11 specimens for each site to validate the morphological taxonomic identification. For other specimens (from VAR sites), molecular identification using COI was not resolutive, so we amplified and sequenced two different non-coding nuclear markers which were successfully used for species identification in Leuciscidae fishes [2,4], with 156–218 base pairs (bp) fragments of Cyprinid formerly unknown nuclear Polymorphism (Cyfun P, [42]), and 313–851 bp fragments of the first intron of the ribosomal protein S7 (S7).

To investigate genetic diversity across *S. rubilio* populations, the entire mitochondrial control region (CR, 930 bp), a non-coding fragment involved in the initiation of the replication and transcription of the entire mtDNA, was amplified in all Italian samples.

PCR reactions for each marker were carried out in a total volume of 20 µL in a TC9610 Multigene OptiMax Thermal Cycler. Reagents, primers description and thermal cycling condition parameters for each marker are described in Table S1. Amplicons were purified and sequenced using an external service (Microsynth AG, Balgach, Switzerland, www.microsynth.ch, accessed on 11 April 2022).

The obtained sequences were then compared to those available for almost all Leuciscidae species in BOLD [43] and Genbank using the Basic Local Alignment Search Tool (BLAST) [44].

### 2.3. Data Analysis

Sequences for all markers were aligned using Clustal X 2.0 [45] and the polymorphic sites were checked manually.

Mitochondrial haplotypes for both COI and CR were identified using DnaSP v6 [46] and haplotype networks were built using the TCS algorithm [47] in Popart [48].

To investigate the phylogenetic relationships with closely related taxa, we used COI sequences and applied the Maximum Likelihood (ML) and Bayesian Inference (BI) methods. Our COI dataset was enriched with all *S. rubilio* sequences stored in GenBank (19 in total) and with 4 sequences from each *Rutilus* and *Leucos* available species, which were taken from different basins when possible (Table S2); however, "unverified" sequences were not included. *Squalius cephalus* (accession number NC031540) was used as an outgroup. An ML analysis was performed using W-IQ-TREE [49], with the substitution model TIM + F + G4 defined by ModelFinder [50] according to the minimum BIC score, and by performing an ultrafast bootstrap analysis [51] for 1000 iterations for statistical support. BI analyses were performed using MrBayes 3.2.7 [52] with the GTR + I substitution model, selected by the AIC (Akaike Information Criterion) and implemented in JModelTest v.2.1.10 [53]. The analysis was carried out in three heated and one cold Metropolis-coupled Markov chain Monte Carlo (MCMC) in two independent runs of 1 million generations each, with sampling performed every 1000 generations and 10% of the trees discarded as burn-in. The effective sample size (ESS) of the two runs was evaluated using Tracer v1.7.1 [54] and each parameter exceeded 500. All trees obtained in the ML and BI analyses were visualized using FigTree v1.4.4 (Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, http://tree.bio.ed.ac.uk/software/figtree/, accessed on 11 April 2022).

After phylogenetic reconstruction, all the analyses of genetic diversity, population structure and demography were performed based on CR sequences of *S. rubilio* from Italy. Arlequin 3.1 [55] was used to calculate standard molecular indices (Hd and $\pi$, respectively [56]), both for populations and haplogroups identified by the network, and to calculate the genetic differentiation between populations as pairwise ΦST (Kimura 2-

parameters distance–K2P [57]). Significance was tested by performing 10,000 permutations and adjusting *p*-values for multiple testing with Bonferroni correction. Additionally, inter- and intra-haplogroup genetic distances (K2P) were calculated using MEGA 11 [58]. Haplotype richness was calculated using the PopGenReport package [59] in R software [60] by randomly sampling 13 individuals per population to account for different sample sizes.

Population differences were visualized by performing non-metric multidimensional scaling (NMDS) in PAST 3.26 [61] on the ΦST matrix. The NMDS analysis was also repeated considering the presence of more than one haplogroup for some populations (see results).

An AMOVA (Analysis of Molecular Variance) with Arlequin 3.1 was used to test the various hypothesis of population structure, which were as follows: (1) no groups, i.e., panmixia, or population subdivision according to (2) ichthyogeographic district (3 groups, PV, TL, AC attributing sites located in the undefined area according to recent literature [27]), (3) number of haplogroups (3 groups, different from the previous hypothesis), (4) NMDS output (7 groups, due to the presence of subgroups within haplogroups).

The divergence time between CR haplogroups was estimated using the BI method implemented in BEAST v.1.10.4 [62], and by setting the HKY+I + G substitution model selected using JModelTest and the Coalescent Constant Size as Tree Prior; and 3.84 and 8.48% per million years (My) divergence rates were used, i.e., the minimum and maximum rates estimated for CR in Leuciscidae species [63]. *Rutilus rutilus*, the only roach species for which the CR sequence was available, and *Squalius squalus* (GenBank Accession Numbers AP010775 and NC031540, respectively), were used as outgroups. MCMC was run for 10 million steps and sampled every 1000 steps, with 10% of trees discarded as burn-in. The EES and trees were evaluated as in the COI analysis.

The ancestral state of haplogroup distribution, dispersal and vicariance events were inferred using the statistical DIVA (S-DIVA) method [64] implemented in RASP v.3.1 software [65]. CR Bayesian tree topology was used to map ancestral distributions and the tips of the tree (CR haplotypes) were coded using the following five main Italian hydrogeographic areas, identified a priori according to geographic distances and barriers between them: (1) Magra-Vara basin at the species northern border (VAR sites); (2) central Adriatic slope (TRO and FOR); (3) central Tyrrhenian slope (TIB sites and ARR); (4) Fondi plain (SET, SPU); (5) south Tyrrhenian slope (SCR, GAR, NOC). Outgroup species distribution was labeled as "continental Europe".

To make inferences about past demographic events in *S. rubilio*, the following three steps were conducted for each previously identified CR haplogroup: (1) Arlequin 3.1 for the mismatch analysis and to test the demographic expansion hypothesis, calculating Harpending's raggedness index Hri [66] and the sum of squared deviations SSD [67], which were both assessed with a parametric bootstrap of 10,000 replicates; and the expansion parameter $\tau$ and effective population size $\theta_0$ and $\theta_1$ before and after the expansion were also calculated; (2) the same software was used to perform Tajima's D [68] and Fu's F [69] neutrality tests together with the R2 neutrality test [70] implemented in DnaSP v6. When the significative signature of demographic expansion was found, the time since expansion was calculated as $t = \tau/(2\mu * \text{generation time} * k)$, where $2\mu$ is the divergence rate and K is sequence length [71]; the generation time for *S. rubilio* is one year [34].

## 3. Results

### *3.1. Molecular Identification and Phylogeny*

The analysis of COI (624 bp) sequences from a subset of 69 specimens revealed 21 COI haplotypes (GenBank accessions: OM974277-97). A total of 7 of them were identified among the 15 sequences from Albanian specimens (differing from each other for one-three site-specific mutational steps) and 14 haplotypes from the sequences corresponding to 54 Italian samples (Table 1 and Figure S1). All variable sites between haplotypes were synonymous mutations. Comparisons with GenBank sequences for the Albanian samples showed > 99% similarity to sequences labelled as *Rutilus albus*, *R. prespensis*, *R. ohridanus* and only ~93% with *S. rubilio*. Considering Italian specimens, the morphological attribution

to *S. rubilio* was confirmed by COI sequences for 46 out of the 54 individuals (from 97.91 to 100% similarity). The remaining eight specimens, collected in VAR1 and VAR2 sites (HpC01 and HpC02), showed 95.01–95.97% similarity with *S. rubilio* and lower values than any other species present in GenBank (92.72–94.55% when compared with other *Rutilus*/*Leucos* species). The BOLD identification engine was also unable to match any records in the selected database for these VAR specimens.

A phylogenetic analysis including *Rutilus* and *Leucos* species showed the same tree topology for both ML and BI (Figure 2). All Albanian haplotypes from lake Skadar clustered with most *Leucos albus* specimens while those from lake Ohrid showed no clear assignment to a distinct group, but were still included with strong support in the *L. albus*/*R. ohridanus*/*R. prepspensis* cluster. Despite *R. ohridanus* (as well as *R. prespensis*) being currently considered a synonym of *Leucos basak* (see Discussion), our trees showed these putative taxa did not form a monophyletic group exclusive of *L. aula* or *L. albus*, and *L. basak* from the Neretva basin did not cluster with Albanian ones. Once it was ascertained that samples from Albania did not correspond to *S. rubilio*, they were excluded from further analysis.

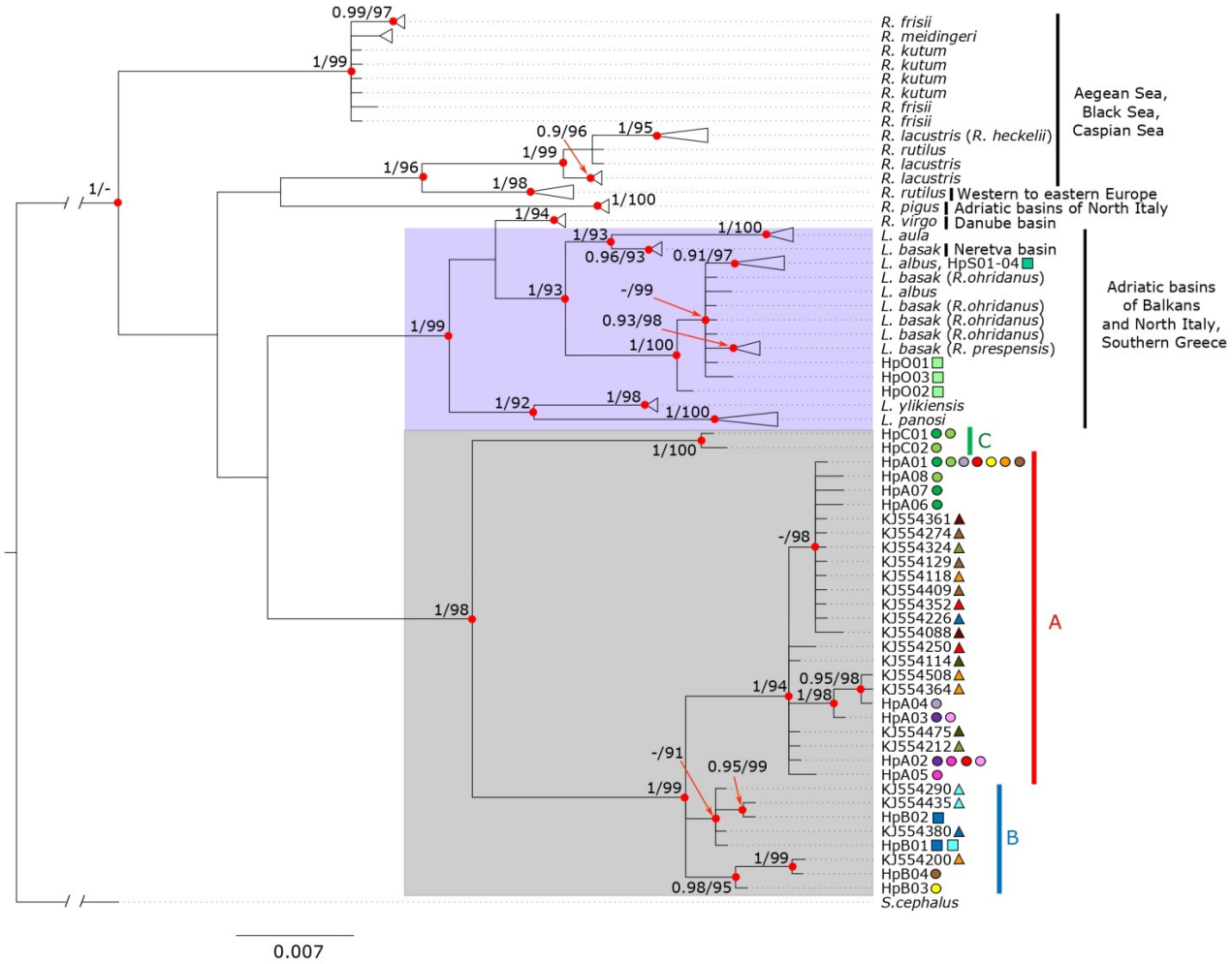



**Figure 2.** ML and BI phylogenetic tree based on mitochondrial sequences of COI including sequences of all the species originally included in the genus *Rutilus* (see Table S2 for Accession Number). Haplotypes obtained in this study are indicated and represented by circles (rivers) and squares (lakes); sequences retrieved from Genbank are indicated by triangles. Rectangles on the tree highlight currently recognized *Leucos* species (violet) and *Sarmarutilus* (grey); for the latter, three haplogroups are indicated. Posterior probabilities > 0.90 (BI) and bootstrap values > 90 (ML) at each node are shown.

All Italian sequences clustered in a well-supported monophyletic clade (hereafter *Sarmarutilus*); however, its relationships with *Leucos* and *Rutilus* clades were not resolved, as basal nodes in our tree were not supported. Within *Sarmarutilus*, different lineages were identified, with specific geographic distributions. The most differentiated one includes only two haplotypes from the sampling sites at the northern border of the species range (VAR, lineage C) and differs for 26–27 mutational steps from the most common haplotypes (HpA01 and HpB01) (Table S3a and Figure S1). The second (lineage A) includes the majority of the sequences and is widespread across the peninsula. The remaining sequences (group B) are exclusive of the southern Tyrrhenian slope and do not form a monophyletic cluster, although they are clearly grouped in the COI haplotype network (Figure S1), due to the low number of mutations compared to haplotypes of lineages A and C. From now on, we refer to haplogroups A, B and C. None of the previously deposited sequences clustered in haplogroup C.

Considering that COI haplogroup A (common *S. rubilio*) and C (unknown) coexist in the same VAR sites, we investigated whether (a) individuals from the two genetic lineages represent remnants of isolation and secondary contact events, (b) nuclear divergence among them still exist and/or (c) whether they hybridize. To test these hypotheses, nuclear markers were amplified in a subset of individuals from both mtDNA lineages (A and C), from VAR and sites in Central Italy. The nuclear S7 (GenBank accessions: OM966282-97) region did not show sufficient variability, and thus it did not provide an answer. On the contrary, the Cyfun P fragment (GenBank accessions: OM966266-81) was more variable and provided diagnostic mutations/indels that distinguished two groups of nuclear sequences among VAR specimens, one of which shared samples from Central Italy and the other was exclusive of VAR (Table S4). Sequences found in both VAR and Central Italy showed 97.02–100.00% similarity with the ones available in GenBank for *S. rubilio* (JQ286163-4, from Tiber river), while for those exclusive of VAR sites similarity values dropped to 90.30–92.20%. The genetic distance between our two groups of Cyfun P sequences (K2P, MEGA 11) was equal to approximately 4%, with VAR specific ones also distinguished by the deletion of nine base pairs. Of note, six out of seven individuals showing the typical Cyfun P VAR sequences belonged to the mitochondrial haplogroup C (Table S4). We re-examined the morphology of VAR specimens (pictures and vouchers) showing both mitochondrial and nuclear sequences belonging to lineage C, with no differences from *S. rubilio* observed (Figure S2) and meristic counts in the range of *S. rubilio* (number of branched rays of dorsal, anal and pelvic fins, scales on lateral line and number of pharyngeal teeth [30]).

### *3.2. Genetic Variability and Demographic History*

We identified 21 CR haplotypes from 208 Italian specimens (GenBank accessions: OM966233-53, Figure 3), whose spatial distribution mirrored that observed with COI. Haplogroup A (12 haplotypes) is the most widespread of these and is present in 10 out of 12 Italian sampling sites, from north to south, both in Tyrrhenian and Adriatic slopes; haplogroup B (7 haplotypes) is present in south Tyrrhenian slope, fixed in SET and SPU sites, and rare in SCR and NOC (one specimen each), where it coexists with haplogroup A; haplogroup C (2 haplotypes) is limited to VAR sites, always found together with haplogroup A but less abundant (representing 31% and 40% in VAR1 and VAR2, respectively) (Figure S3). As for COI, a high number of mutational steps was observed between CR sequences belonging to haplogroup C and the most frequent haplotypes of haplogroup A and B: 30 between HpC01 and HpA01, 27 between HpC01 and HpB01; additionally, sequences of haplogroup C were one base shorter (929 bp) than those of the other two haplogroups (930 bp) (Table S3b).

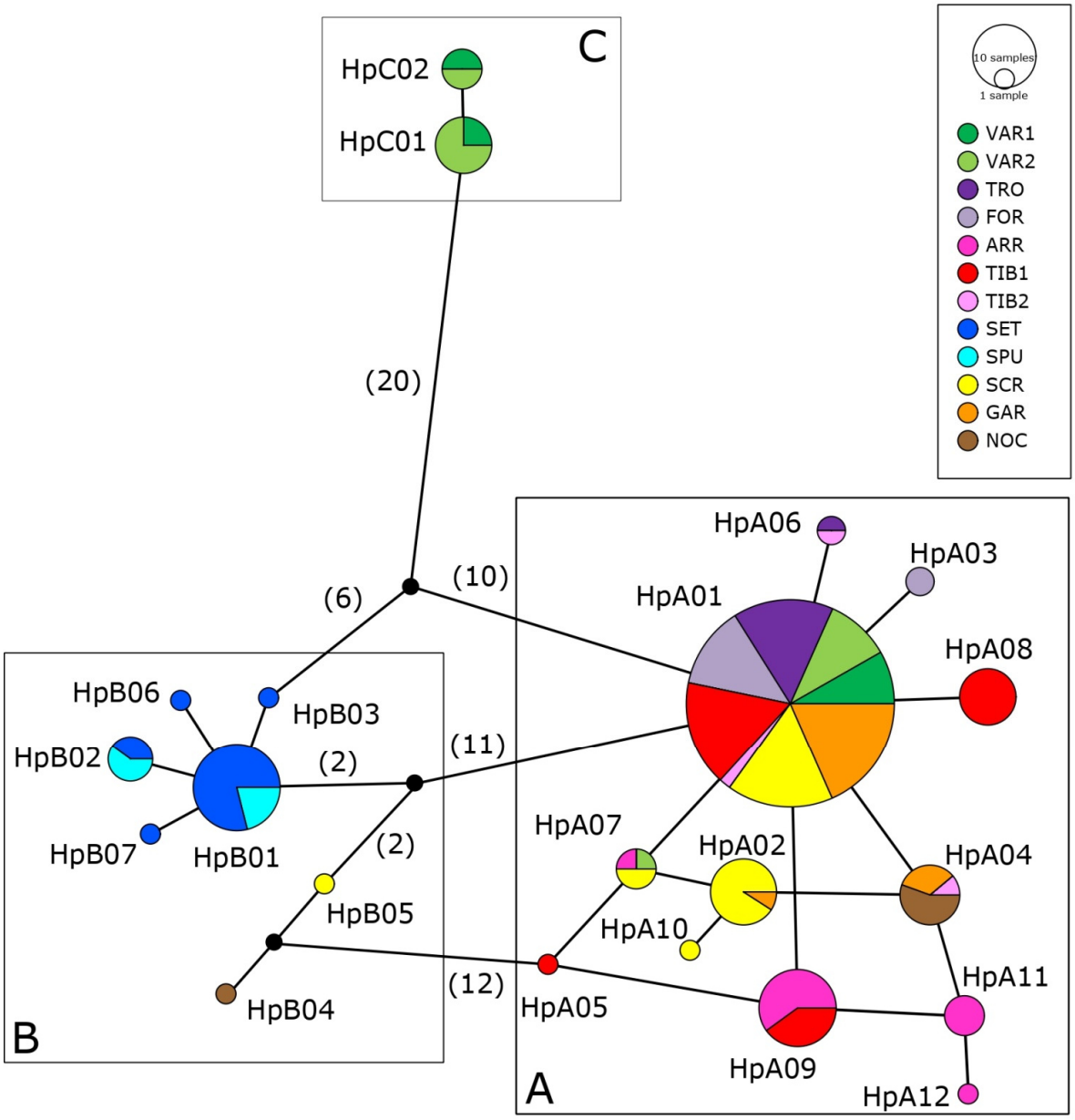



**Figure 3.** Haplotype network based on CR sequences. Each circle corresponds to one haplotype and its dimension is proportional to the haplotype frequency. The number of nucleotide substitutions between haplotypes is indicated in parenthesis. Boxes highlight the three haplogroups found in this study (A, B and C, see Results). Population abbreviations refer to Table 1.

The CR haplotype network (Figure 3) showed a star-like topology for both haplogroup A and B (the latter mainly due to SET specific haplotypes), with few mutations at the intra-haplogroup level (1–4 in A, 1–7 in B, 1 in C). HpA01 is the most common haplotype, being shared by individuals from different basins and slopes; seven haplotypes are shared by individuals from different basins (HpA02, HpA04, HpA06, HpA07, HpA09, HpB01 and HpB02), and others within a single basin (e.g., HpC01 and HpC02 in VAR sites). Haplotype and nucleotide diversity for each site (Table 1) showed the lowest values of Hd in sites from the Adriatic slope (TRO, FOR) and in two of the southernmost ones (GAR, NOC), which also showed the lowest values of $\pi$% and haplotype richness, except for NOC, where haplogroup A and B coexist. Higher values of $\pi$% were observed in VAR sites, where haplogroup A and C coexist. When calculations were repeated for each haplogroup, similar

parameters were obtained for haplogroups A and B, and lower values for haplogroup C, which included the smallest number of samples (Table 2).

**Table 2.** Summary of genetic variation, neutrality test and mismatch distribution parameters for the mitochondrial control region (CR) for each haplogroup and all samples.

| Haplogroup | N | Hp | Hd (±s.e.) | π % (±s.e.) | Tajima's D | Fu's F | R2 | SSD | Hri | τ | $\theta_0$ | $\theta_1$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 167 | 12 | 0.56 (±0.04) | 0.09 (±0.07) | −0.919 | −6.561 * | 0.052 | 0.001 | 0.056 | 0.926 | 0.100 | 3.196 |
| B | 29 | 7 | 0.55 (±0.10) | 0.13 (±0.09) | −1.722 * | −2.237 | 0.062 ** | 0.014 | 0.119 | 0.725 | 0.000 | 99,999.000 |
| C | 12 | 2 | 0.48 (±0.11) | 0.05 (±0.05) | 1.066 | 1.003 | 0.242 | 0.019 | 0.236 | 0.734 | 0.000 | 99,999.000 |
| Tot. | 208 | 21 | 0.71 (±0.03) | 0.75 (±0.39) | −0.107 | 1.541 | 0.081 | 0.035 | 0.040 | 0.072 | 1.816 | 99,999.000 |

Abbreviations: sample size (N), number of haplotypes (Hp), haplotype diversity (Hd), nucleotide diversity (π), standard errors (s.e.), sum of squared deviations (SSD), Harpending's raggedness index(Hri), time since expansion in mutation units (τ), population size estimators before and after the expansion ($\theta_0$ and $\theta_1$). Significance thresholds: * $p < 0.05$; ** $p < 0.01$.

Data from the mismatch distribution analysis for all samples did not agree with the expectation of demographic expansion of a homogeneous population; conversely, three different peaks were observed following the presence of the three identified haplogroups (Figure S4). The analysis was then repeated for each haplogroup and a unimodal distribution was obtained for haplogroup A and B. Tajima's D and Fu's F consistently has negative values (indicating population size expansion); however, when considering all demographic parameters together, only those referring to haplogroup B agreed with the expansion hypothesis, i.e., it had a significant Tajima'D and R2, $\theta_1$ equal to 99,999 (Table 2) and the times of demographic growth for B corresponded to 20-9 Kya. No clear pattern was obtained for the small (12 individuals) haplogroup C.

### *3.3. Genetic Structure and Phylogeography*

Global ΦST (0.577, $p < 0.001$) indicated high genetic differentiation among sites. Calculation of pairwise ΦST values (Table S6a) showed that the highest significant values were observed in comparisons involving sites with a homogenous composition, i.e., between B-only and A-only sites ($0.95 < \Phi ST < 0.98$); however, the results were partly biased at sites where individuals belonging to different haplogroups were present (e.g., VAR). It is noteworthy that significant values were also obtained among sites within haplogroup A (e.g., see ARR, TIB1). To deepen the analysis on genetic differentiation between the main haplogroups and sites, pairwise ΦST calculation was repeated by separating the different haplogroup components, when they occur in the same site simultaneously (in VAR, SCR, NOC) (Table S6b). Values obtained between different haplogroups were high ($0.93 < \Phi ST < 1.00$) but not always significant, likely due to a very small sample size. NMDS plots were obtained by considering all the sites (Figure 4a), and the main haplogroups within the sites (Figure 4b) allowed us to identify the main haplogroups pattern and further subdivisions within the most represented ones (haplogroups A and B). Indeed, within haplogroup A, all populations were placed together but NOC_A and ARR appeared to be separated from the others, similarly to haplogroup B, where SCR_B and NOC_B were separated from SET/SPU but in different directions so that three B subgroups were graphically identified.

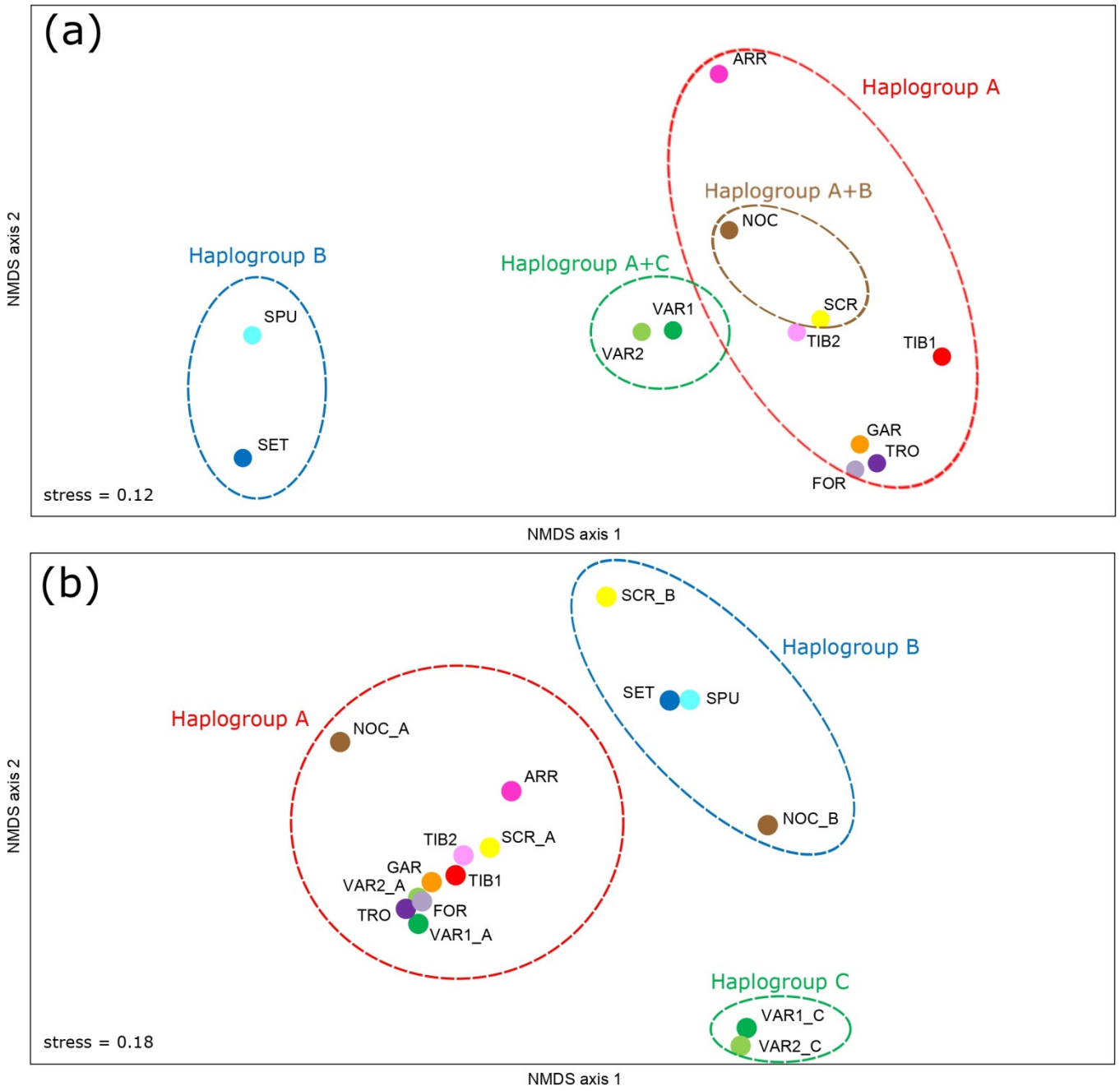



**Figure 4.** Non-metric multidimensional scaling based on CR matrix of ΦST on (**a**) sampling sites, (**b**) splitting individuals by haplogroup when two haplogroups occur in the same site (in VAR, SCR, NOC). Populations colors and abbreviations as in Figure 1.

Among the different hypotheses tested using AMOVA, statistical support was obtained when considering subdivision with three haplogroups, with 95.06% of the variation explained by among-groups differences (Table 3). When a further subdivision within haplogroups, suggested by NMDS, was tested, a similar among-groups percentage was observed (95.18%) but also a smaller value for within-groups variation. On the contrary, a subdivision of samples purely based on ichthyogeographic district origin did not provide statistically supported results.

**Table 3.** AMOVA hierarchical analysis for Italian specimens, examining the partitioning of genetic variance of the mitochondrial control region (CR).

| | Among Groups | | Among Populations within Groups | | Within Populations | |
|---|---|---|---|---|---|---|
| **N. of Groups and Group Composition** | **%var** | **ΦCT** | **%var** | **ΦSC** | **%var** | **ΦST** |
| (1) no structure: 1 group | – | – | 57.68 | – | 42.32 | 0.57682 *** |
| (2) ichthyogeographic districts: 3 groups (PV = TRO; TL = VAR1, VAR2, ARR, TIB1, TIB2; AC = FOR, SET, SPU, SCR, GAR, NOC) | −9.87 | −0.09873 | 66.04 | 0.60108 *** | 43.83 | 0.56169 *** |
| (3) Haplogroups: 3 groups (HpC = VAR1_C, VAR2_C; HpA = VAR1_A, VAR2_A, TRO, FOR, ARR, TIB1, TIB2, SCR_A, GAR, NOC_A; HpB = SET, SPU, SCR_B, NOC_B) | 95.06 | 0.951 *** | 2.07 | 0.418 *** | 2.88 | 0.971 *** |
| (4) NMDS, subgroups within haplogroups: 7 groups (VAR1_C, VAR2_C; VAR1_A, VAR2_A, TRO, FOR, TIB1, TIB2, SCR_A, GAR; ARR; NOC_A; SET, SPU; SCR_B; NOC_B) | 95.18 | 0.952 *** | 0.87 | 0.181 *** | 3.95 | 0.961 *** |

Hypothesized structures: (1) no structure (one group), (2) subdivision by ichthyogeographic district (three groups), (3) subdivision by haplogroup (three groups), (4) clusters identified by NMDS analyses (seven groups). Significance thresholds: *** = $p < 0.001$.

Genetic distances between haplogroups were higher when comparing C to A and B (0.033 and 0.029, respectively, Table S5), than between A and B (0.015). A dating analysis and ancestral areal reconstruction (S-DIVA) based on the CR BI tree revealed a complex phylogeography for *Sarmarutilus* in Italy (Figure 5). Vicariance events were estimated at the basal split of haplogroups; the divergence time of lineage C corresponded to 850–390 thousand years ago (Kya) while between A and B it corresponded to 500–230 Kya. Within haplogroup A, the S-DIVA analysis identified many dispersal events, despite nodes not being supported, from the central to south Tyrrhenian slope and Adriatic slope. Within haplogroup B, S-DIVA estimated a vicariance event between the Fondi plain and the south Tyrrhenian slope.

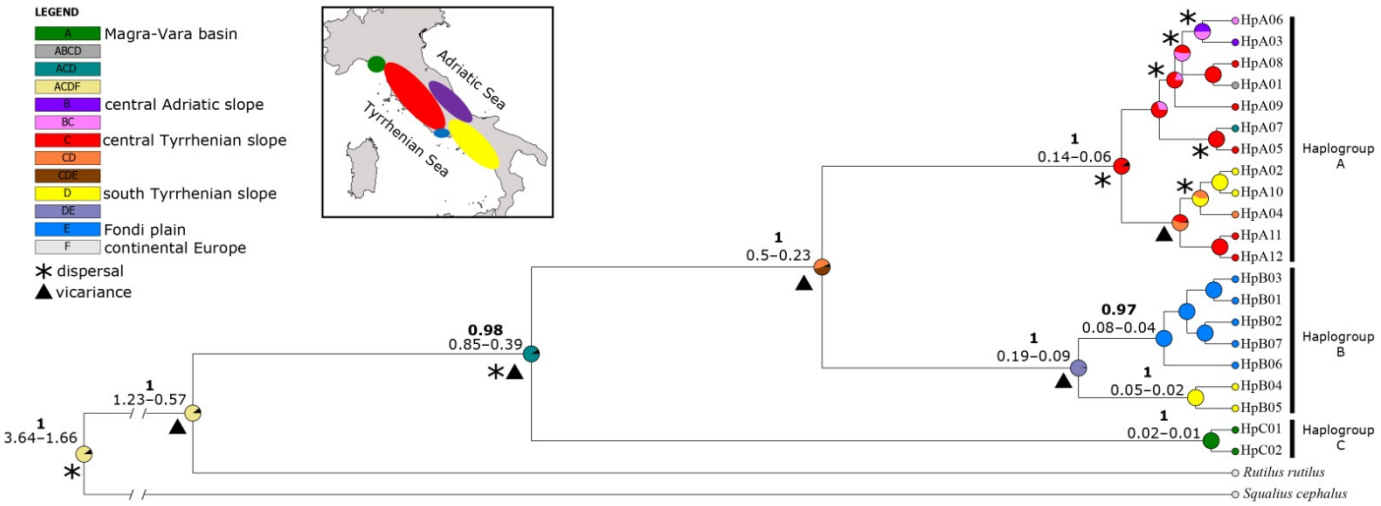



**Figure 5.** Bayesian CR tree with time estimates at supported nodes and ancestral area reconstruction of Sarmarutilus haplogroups and haplotypes. Posterior probability higher than 0.90 in bold; time estimates are in Mya; * denotes dispersal events, triangle vicariance events. Area reconstructions lower than 15% are indicated in black as null.

## 4. Discussion

In this paper, the analysis of mitochondrial DNA in Italian and Albanian roaches, complemented with sequences from representatives of all the currently recognized *Rutilus* and *Leucos* species, confirmed a loose relationship between the two investigated taxa, despite the very similar morphology. Our phylogenetic reconstruction also highlighted the need for an in-depth taxonomical revision of Balkan roach species. Finally, our intraspecific analysis revealed a deep and complex phylogeographic structure for Italian *S. rubilio*, masked by secondary contacts between lineages.

### *4.1. Molecular Identification and Phylogeny*

The COI molecular identification and phylogenetic analysis provide new insights into the relationships among roach species. Our analysis was overall congruent with the COI tree obtained in previous studies [72]; however, due to the integration of a higher number of sequences from different basins and the molecular identification of new lineages, more details emerged. Despite this, some main points remain unresolved.

The Albanian specimens examined in this study (from Skadar and Ohrid lakes, morphologically similar to Italian *S. rubilio*) were clustered separately from Italian samples in the phylogenetic tree. Specifically, specimens from each lake showed one to three site-specific mutations, and haplotypes from the Skadar lake clustered with *L. albus* (typical from this lake) while those from the Ohrid lake grouped with *L. basak* sequences from this site (recorded in GenBank as *R. ohridanus*); furthermore, the Ohrid lake cluster also included *L. aula* and *L. basak* from lake Prespa in Albania (recorded in GenBank as *R.prespensis*). The status of the *Rutilus*/*Leucos* species of the Balkan area is debated and there is no agreement on the validity and distribution of the previously mentioned species. Indeed, based on their morphological and genetic traits, the following three species were identified [37]: *Rutilus albus* in Lake Skadar, *R. ohridanus* endemic in the Ohrid lake, and *R. prespensis* present both in Lake Prespa and Lake Skadar (in this site in sympatry with *R. albus*). Subsequently, the taxonomy of the genus *Rutilus* was revisited [30], wherein the genus *Leucos* was resurrected, and in the Adriatic Balkan area, *R. albus* was renamed *Leucos albus.* Moreover, *R. basak* from the Neretva basin (between Croatia and Bosnia-Herzegovina) was named *Leucos basak*, and the Albanian taxa distributed in the Ohrid lake (previous *R. ohridanus*), Prespa lake (previous *R. prespensis*) and Skadar lake (*R. prespensis* in sympatry with *L. albus*) were considered synonyms of *L. basak* [73]. Our data disagree with this attribution, as the non-monophyletic status of current *L. basak*. specimens from Neretva are highly divergent from those from the Albanian lakes included in this taxon. Indeed, the Neretva and Albanian basins belong to different ichthyogeographic districts [74]. Conversely, specimens from the Ohrid and Prespa lakes are grouped with *L. albus* in the same monophyletic clade according to previous reports [25]. These data allow us to draw some conclusions on the taxonomy of these species. First, Albanian taxa cannot be considered synonyms of *Leucos basak* from the Neretva basin. Second, the choice to synonymize *Rutilus prespensis*, *R. ohridanus* and *R. albus*, considering their close genetic relationship [20], is supported by our data. This seems a wise parsimony choice that fits the need to avoid the taxonomy inflation that characterizes other fish taxa [75], and also considers the low number of mutations between the haplotypes collected in Ohrid or Skadar lakes. Meanwhile, diagnostic mutations are present between fishes of these lakes, and this differentiation should not be disregarded, as they can potentially be considered as geographic populations. Moreover, morphological differentiation is reported between *L. albus* and the sympatric taxa in the Skadar lake and further integrated analyses identifying morphological characteristics, multiple genetic markers and environmental parameters are needed to clarify the basis of such diversity.

Another problematic issue involves *Rutilus virgo*. Our COI phylogenetic analysis clustered this taxon into the *Leucos* genus with strong support, according to previous research [25,76]. However, breeding males of this species show horizontal rows of large tubercles on their scales and head [39] that are typically absent in the *Leucos* species [30]. In addition, some zoogeographic features differentiate *R. virgo* from the *Leucos* species,

including the following: the former inhabits rivers flowing into the Danube River–in Slovakia, and Slovenia, rarely in the Czech Republic [77] while *Leucos* species are typically present in rivers of the Balkans and Italy flowing in the Adriatic Sea. Thus, the relationship between *R. virgo* and the other roach species remains ambiguous. Despite *Leucos* being a well-supported monophyletic clade, its relationship with other *Rutilus* species and *Sarmaratuilus* were not resolved in our analysis; as in previous studies [25,72], we were unable to define whether *Leucos* or *Sarmarutilus* are sister groups of *Rutilus* or, conversely, if they are included in *Rutilus*, so that the latter is not monophyletic. This issue will need to be addressed in the future through analyses based on a combined dataset for all roach species. Finally, our tree topology lends support to the differentiation of the endemic roach from the Volvi lake and Struma river in Greece, as the COI sequences from these sites retrieved from GenBank (there labelled as *R. heckelii* and *R. lacustris*, respectively) are grouped in a well-supported cluster separated from that including *R. lacustris* from Black and Caspian Sea basins (one *R. rutilus* sequences clustering into *R. lacustris* may be due to misidentification as the two species are sympatric in Eastern Europe). Roaches from Volvi lake were recognized as a distinct species, *Rutilus stoumboudae*, based on both their morphological and molecular characteristics [30,78]. However, this putative species is currently considered a synonym of *R. lacustris* [73,79], due to the low genetic divergence from this species. Our data confirm this taxonomic attribution, and as for the case of *R. orhidanus/prespensis*, we suggest considering Volvi roach specimens as a geographic population that has peculiar genetic traits allowing its distinction from *R. lacustris* from the Ponto Caspian area. New studies covering other Greek localities will clarify its distribution in Greece and its conservation status, as the population is probably extinct in the Volvi lake [30].

### *4.2. Divergence between Sarmarutilus Lineages: Cryptic Species?*

The COI phylogenetic tree and both COI and CR haplotype networks revealed three different haplogroups among Italian specimens, all of which were morphologically identified as *Sarmarutilus rubilio*. Two haplogroups (A and B) matched with sequences previously identified by other authors [20], while the third one (C), exclusive to the Magra-Vara basin (VAR sites), was never recorded before and did not show adequate similarity with any sequence stored in international databases (GenBank, BOLD). COI divergence obtained for the lineage C ranges from 3.3 to 2.9% (A–C and B,C haplogroups pairwise comparison, respectively). According to what was observed in 1088 fish species [80], sequences with a divergence of greater than 2% or 3% likely belong to different species (with a probability greater than 95%). This COI divergence threshold was generally confirmed with an analysis implementing several methods for species delimitations in fishes and allowed for the identification of cryptic species [81–83] even in Leuciscidae [84]; furthermore, 2–3% divergence was observed between many *Leucos* and *Rutilus* taxa whose status as good species is not debated. Although our nuclear dataset is limited, a substantial divergence (almost 4%) and a slight difference in length were observed between Cyfun P sequences exclusive to VAR sites and mostly present in specimens belonging to mitochondrial lineage C, and those widespread across *S. rubilio* populations. In other Leuciscidae species, no intraspecific variability in Cyfun P sequence length was observed [42], and a very small divergence (0.3–1.5%) or none at all (if recently split) was observed between species [4,85]. These data are highly indicative of typical Magra-Vara lineage C being a putative new cryptic *Sarmarutilus* species. More in-depth investigations, including a wider nuclear dataset, are necessary to strengthen confidence in our hypothesis. These roaches coexist with lineage A and, as specimens with various combinations of nuclear and mtDNA lineage were found, they can hybridize. This commonly happens in Leuciscidae, between different species and even genera [86–88], and genetic introgression was also reported, not always associated with intermediate morphological characters [4,85,89]. In the absence of morphological diagnostic traits (see Results), a formal description of this putative new species is currently not possible. Moreover, we are carrying out a geometric morphometrics analysis for all

Italian specimens to investigate the morphological differences at a finer scale, and the results will be displayed and discussed in a new paper. Further sampling is also needed to map the distribution of C lineage in the Magra-Vara basin and in nearby small catchments that could also host yet undetected *Sarmarutilus* populations.

### *4.3. Phylogeography of Italian Sarmarutilus*

The presence of multiple haplogroups within *Sarmarutilus* and their coexistence in some sampling sites are congruent with the hypothesis of multiple refugia in Italy during Pleistocene glacial cycles and subsequent secondary contact, as observed in other freshwater taxa [90–94]. Estimated time splits between *Sarmarutilus* haplogroups (850 to 230 Kya) indeed correlate with Pleistocene events. In addition to past climatic fluctuations, freshwater fishes' phylogeographic patterns and distribution in Italy were also driven by mountain ridges and the presence/lack of connections among drainages [14,27,95]; this is especially true for those taxa, such as Leuciscidae, that cannot disperse through the sea due to their little or no tolerance to brackish water [26] and references therein. Indeed, the Apennine chain, dividing peninsular Italy longitudinally from north to south, was found to influence the biogeographic structure even of semiaquatic vertebrate species but not that of the peninsular terrestrial ones (if not marginally) [96].

Despite this, the current distribution of *Sarmarutilus* haplogroups only partially matches the currently recognized ichthyogeographic Italian districts, which are characterized by different hydrogeological histories and evolution and are inhabited by different species [21,97]. See details on each haplogroup below.

#### 4.3.1. Haplogroup C

The presence of the most ancient lineage (haplogroup C) only at the northern border of *Sarmarutilus* range (Magra-Vara basin), permits some inferences on the origin of this genus.

There is no record of roaches westward to this basin, while northward, typical PV district roach species are present, i.e., *Leucos aula* and *Rutilus pigus*, are not strictly related to *Sarmarutilus* (see Figure 2). The Magra-Vara basin has been geographically isolated since the late-Pliocene and early Pleistocene tectonic uplift of the surrounding mountains as follows [98]: (a) from the PV district by the northern Apennines, thus representing the northern border of most of the primary freshwater fish species native to Central Italy [99]; (b) from southern catchments by the Apuan Alps, a geographic barrier extending from the main Apennine to the sea which greatly effects the genetic structure of freshwater species [26,28,100]. In the early Pleistocene, when most of the current major southern basins such as Arno and Tiber were limited in extension due to marine introgression, Magra-Vara still had connections to both basins flowing in the Tyrrhenian Sea and the Adriatic Sea [98]. Thus, it is likely that *Sarmarutilus* moved initially from the Balkans (where it originated, see Introduction) to Magra-Vara through brackish waters [96] and from here it spread to southern catchments. This colonization path is in contrast to the immigration route observed in other Italian vertebrates [96] including other leuciscidae [15,97], i.e., from the Balkans to south Italy and a subsequent northward expansion. The split between *Sarmarutilus* lineage C and *S. rubilio sensu stricto* can be dated back to 850–390 Kya, supporting the hypothesis that the uprising of the Apuan Alps in the early Pleistocene promoted the isolation of lineage C in this basin from *S. rubilio* present in southern catchments.

The presence of the widespread *S. rubilio* lineage A in Magra-Vara today will be discussed later.

A demographic analysis of haplogroup C indicated no deviance from a neutral/stable population and thus the absence of post-glacial population expansion, as conversely observed in this site for *Squalius lucumonis* [28]. Currently, we are not able to decipher whether lineage C demography was unaffected by the alternation of glacial and interglacial periods or if the secondary contact (and partial hybridization) with *S. rubilio* haplogroup A caused the depletion of lineage C haplotypes. Haplogroup A is the most frequent in both

Magra-Vara sites and some vantage selection may exist (see Haplogroup A discussion, here below).

#### 4.3.2. Haplogroup A

This haplogroup is the most widespread, present from the northern to southern distribution limits and in most of the sites and is typical in the TL district, but also present in the AC and the Adriatic slope of Central Italy.

Haplogroup A originated in the TL district after the first split from lineage C and a second one from B, during the Pleistocene, i.e., when it likely remained isolated during the glacial phase somewhere in a refugium and underwent allopatric divergence; after that, dispersal events promoted its geographic spread through the Apennine chain, as suggested by S-DIVA (Figure 5). Indeed, these mountains, despite isolating the Tyrrhenian and Adriatic slopes of Central Italy and being considered the natural border between TL and PV districts and between TL and the northern Adriatic section of AC, could have acted as a semi-permeable barrier. This seems plausible especially for rheophilic species [101], through river connections acting in the upper section of basins [102], and as already reported in Mediterranean trout [103,104]. *Sarmarutilus rubilio* is not a strict rheophilic species; however, it is found in Apennine rivers and could have spread through the hydrographic network, which could explain the presence of HpA01 in the Adriatic slope (TRO and FOR) and the high differentiation observed in those sites that are not linked to the Apennines, i.e., ARR drainage that is fed by the Bracciano volcanic lake [105]. Their existence along the entire Apennine chain of the intermontane lakes up to the Late-Pleistocene [106,107] could also have had an important role in haplogroup A dispersion [108].

The presence of the most common haplotype HpA01 in Magra-Vara, co-existing with lineage C, may be the result of secondary contact. Since *S. rubilio* can adapt to nearly all water habitats, from lentic to lotic, from nearby sea to 1250 m a.s.l. [34], past sea-level changes could have allowed the re-colonization of Magra-Vara from southern catchments through mouth river confluences, a dispersion route observed in freshwater fishes that inhabit coastal streams [109]. Such migration should have occurred quite recently, during the last glacial age (about 22 Kya), when the sea level was lower than today [110,111]. Alternatively and in addition, a human-driven translocation of lineage A in the Magra-Vara basin could not be excluded, but it is unlikely as all allochthonous species present in this basin are those typical of the northern PV district [99], suggesting the translocation of fishes only from geographic areas where *S. rubilio* is absent. Mitochondrial lineage A, besides being the most widespread, is also the most frequent when coexisting with the others (Figure S3); therefore, we cannot exclude a higher fitness associated with lineage A and/or some kind of competition when secondary contact occurs, as suggested in other freshwater fishes when different lineages come in contact [104].

Neutrality tests and mismatch analysis values suggested demographic expansion when considering all individuals belonging to the widespread haplogroup A, although these results were not fully statistically supported. The combination of Hd and $\pi$% values for the main TL district sites (ARR and TIB1-2, Hd > 0.5 and $\pi$ < 0.5%), and those outside TL where only A was found (TRO, FOR and GAR, Hd < 0.5 and $\pi$% < 0.05) provided different estimates, suggesting a population bottleneck followed by rapid population growth for the former and a recent population bottleneck or founder event for the latter [112].

Finally, considering both the hydrogeological history of Italy and genetic data of *S. rubilio*, it is likely that haplogroup A experienced demographic growth after the glacial period within the TL district and a recent natural range expansion through the Central and Southern Apennines; conversely, translocation events occurred southward to the species' native range [113,114], and likely involved individuals from the TL district, thus explaining the presence of only haplogroup A in the southernmost CRA site (Figures 1 and 2).

4.3.3. Haplogroup B

Haplogroup B is present in the area including the Fondi plain (SET, SPU) where the lineage is fixed; in Liri-Garigliano surroundings (SCR) not attributed to either TL or AC districts [21]; and also over the southern border of AC (NOC). In the latter sites, lineage A is also present. The split between haplogroup A and B was estimated at 500–230 Kya, which is congruent with the emergence of the Colli Albani Volcano (around 450–400 Kya) on the Tyrrhenian southern border of the TL district. The volcano reversed the flow of rivers currently belonging to the Liri-Garigliano drainage from the north (toward the Tiber basin, TL district) to the south [115,116], thus likely breaking the connection between Central and Southern Italy and initiating allopatric genetic differentiation. After that, haplogroup B differentiated into subclusters. The Fondi plain was never reached by haplogroup A, likely due to the isolation of local basins from north to east by the Ausoni-Aurunci mountains, whose uplift started in the Pliocene [117], and from west to south by the sea. Conversely, in the nearby Liri-Garigliano area (about only 30 km far from Fondi plain) and southward, the main rivers originated in the Apennines and secondary contact occurred.

Neutrality tests and demographic parameters supported the demographic expansion hypothesis for haplogroup B, with the main contribution provided by Fondi plain populations (SET, SPU). In these sites, several B haplotypes were connected by the haplotype network in a typical star-like shape, due to the few mutations between each other. Demographic expansion can be explained by considering recovery after the glacial period, and thus accounting for individuals spreading from glacial refugia after harsh climate conditions, as already observed in other freshwater fishes [18,28]. Time since expansion for haplogroup B was estimated at 20 and 9 Kya, while when recalculating for SET and SPU it was only was 19 and 8 Kya, which is subsequential to the Last Glacial Maximum (22 Kya). Thus, the Fondi plain could likely represented not only an area of isolation from widespread haplogroup A but also a glacial refugium.

The distribution of lineage B also in the area between the Tyrrhenian northern border of AC and the southern one of TL, i.e., Liri-Garigliano and the Fondi plain (and northward to SIS site, see Figure 1), agrees with those of other freshwater fishes typical of the AC district, such as the barbel *Barbus fucini* [97], and the loach *Cobitis zanandreai* [118]. We suggest that the Tyrrhenian upper border of the AC district should be moved northward, close to the Tiber river mouth, thus including the aforementioned area. On the Adriatic slope, results from the two sites investigated did not allow us to identify the border between the AC and the other districts; moreover, the distribution of other south Italy endemic species in this area, such as *Barbus samniticus* [97], suggests movement of the border on this side further north to the Vomano basin, thus including our FOR site.

*4.4. Implication for Conservation and Management*

From a management perspective, to prevent the local extinction of *Sarmarutilus* populations and preserve their genetic uniqueness, the following three main points emerged from our analysis:

(a). New research is necessary to precisely map the distribution of lineage C. This new putative cryptic species is endemic to the Magra-Vara basin and is found in tributaries not included in protected areas; therefore, it could be particularly exposed to threats. This underlines the importance of also protecting small river course habitats, as they might represent refugia of relict native fish populations [119].

(b). Conservation actions required by the Habitats Directive for the South European roach [35] should consider that haplogroups A and B represent two conservation units, namely Evolutionary Significant Units (ESUs, for a synthesis of this definition, see [120]) and as such should be managed.

(c). The main conservation efforts should be focused on avoiding the introduction of allochthonous species, especially those that can compete or hybridize with *Sarmarutilus*. In addition, the translocation of *Sarmarutilus* populations from different districts and

within the AC district (where subclusters are present) should be avoided, as these roaches represent different genetic entities.

## 5. Conclusions

Our phylogenetic analysis of roach species assigned all specimens from peninsular Italy to the genus *Sarmarutilus*, while Albanian ones were assigned to the *Leucos genus*. We highlighted the need for further in-depth analyses of roaches from Albanian lakes, as their current identification as different species (*L. allbus* and *L. basak*) was not well supported, and *L. basak* was revealed to be a paraphyletic taxon. We also suggest focusing on other Balkanian taxa in future as relationships between species and the genetic distinctiveness of some populations have yet to be fully understood.

A significative intraspecific difference within South European roach was already noticed [20] (Appendix S1), and the need for further analyses and a taxonomical update was suggested. Our phylogeographic analysis confirmed this idea by revealing a strong genetic structure with a strict geographic basis across peninsular Italy, suggesting a long history of isolation. Pleistocene vicariance events, due to the emergence of geographic barriers such as mountainous ridges and volcanoes, promoted allopatric genetic divergence among the three identified haplogroups, of which C could probably be considered as a new putative cryptic species. This latter species, despite the high genetic divergence, lacks morphological/meristic distinctive features; therefore, we are currently working on geometric morphometrics to detect differences in shape which could differentiate the two species. Haplogroups A and B underwent different phylogeographic histories, whereby A originated in Central Italy, subsequently spread across the whole areal and came into secondary contact with the other haplogroups, while B originated in the Tyrrhenian slope of Central-Southern Italy, where it is currently restricted, and likely experienced further allopatric diversification.

New data are needed to thoroughly assess the extent of *S. rubilio* genetic diversity, from uninvestigated areas of native ranges and other nuclear molecular markers.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13061071/s1, Table S1: PCR conditions and primers for each amplified marker [121–123], Table S2: COI Sequences retrieved from GenBank and BOLD, Table S3: COI (a) and CR (b) diagnostic sites identified among the most frequent sequences of the three haplogroups, Table S4: Diagnostic fragment of nuclear marker Cyfun P, Table S5: Inter- and intra-haplogroup genetic mean distances; pairwise ΦST between haplogroups, Table S6: Population ΦST (CR) considering all the populations as they are (a) and splitting individuals by haplogroup for each one (b), Figure S1: COI haplotype network, Figure S2: Specimens with mtDNA and Cyfun P belonging to lineage C (a) and lineage A (b), Figure S3: Spatial haplogroups distribution in *S. rubilio*, Figure S4: mismatch distributions of CR sequences.

**Author Contributions:** Conceptualization, G.P. and A.R.R.; methodology, G.P. and A.R.R.; validation, G.P. and A.R.R.; formal analysis, G.P.; investigation, G.P., M.T. and A.R.; resources, G.P., A.R., L.T. and L.C.; data curation, G.P.; writing—original draft preparation, G.P. and A.R.R.; writing—review and editing, L.T. and A.R.R.; visualization, G.P.; supervision, A.R.R.; project administration, A.R.R.; funding acquisition, A.R.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by Sapienza University of Rome (Università 2020, project n. 2146271 to A.R.R.).

**Institutional Review Board Statement:** We retrieved samples of fish tissues in accordance with the relevant legislation (CEN EN 131 14011/2003-Water quality-Sampling of fish with electricity) and were authorized by Regional Directions responsible for Hunting and Fishing activities and by Directions of Protected areas responsibles for the investigated sites (Prot. n.: Lazio G10101, 25 July 2019; Abruzzo DPD023/171, 12 April 2021; Marche 213, 13 April 2021; Liguria 5166–2021, 30 August 2021; Aurunci protected area 0002963.U, 1 October 2021).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Sequences were deposited in the GenBank database (NCBI) with the accession numbers: OM974277- OM974297 (COI), OM966233-OM966253 (CR), OM966266-OM966281 (Cyfun P), OM966282- OM966297 (S7). All vouchers (AC1687-AC1698) are deposited in the Museum of Comparative Anatomy "Battista Grassi", Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome.

**Acknowledgments:** We thank L. Talarico, M. Graziani and A. Ricci for their help in field and laboratory activities, A. Splendiani for providing very interesting and useful samples, and P. Colangelo for his suggestions. We thank the staff at Aurunci Mountains Regional Park for granting sampling permissions and supervising field activities in the protected area, and the Italian Directions responsible for Hunting and Fishing activities in Liguria, Marche, Lazio and Abruzzo Regions.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Pereira, L.H.G.; Hanner, R.; Foresti, F.; Oliveira, C. Can DNA Barcoding Accurately Discriminate Megadiverse Neotropical Freshwater Fish Fauna? *BMC Genet.* **2013**, *14*, 20. [CrossRef] [PubMed]
2. Perea, S.; Böhme, M.; Zupančič, P.; Freyhof, J.; Šanda, R.; Özulu, M.; Abdoli, A.; Doadrio, I. Phylogenetic Relationships and Biogeographical Patterns in Circum-Mediterranean Subfamily Leuciscinae (Teleostei, Cyprinidae) Inferred from Both Mitochondrial and Nuclear Data. *BMC Evol. Biol.* **2010**, *10*, 265. [CrossRef] [PubMed]
3. Wang, X.Z.; Gan, X.N.; Li, J.B.; Mayden, R.L.; He, S.P. Cyprinid Phylogeny Based on Bayesian and Maximum Likelihood Analyses of Partitioned Data: Implications for Cyprinidae Systematics. *Sci. China Life Sci.* **2012**, *55*, 761–773. [CrossRef]
4. Tancioni, L.; Russo, T.; Cataudella, S.; Milana, V.; Hett, A.K.; Corsi, E.; Rossi, A.R. Testing Species Delimitations in Four Italian Sympatric Leuciscine Fishes in the Tiber River: A Combined Morphological and Molecular Approach. *PLoS ONE* **2013**, *8*, e60392. [CrossRef] [PubMed]
5. Ramirez, J.L.; Carvalho-Costa, L.F.; Venere, P.C.; Carvalho, D.C.; Troy, W.P.; Galetti, P.M. Testing Monophyly of the Freshwater Fish *Leporinus* (Characiformes, Anostomidae) through Molecular Analysis. *J. Fish Biol.* **2016**, *88*, 1204–1214. [CrossRef] [PubMed]
6. Mangit, F.; Yerli, S.V. Systematic Evaluation of the Genus *Alburnus* (Cyprinidae) with Description of a New Species. *Hydrobiologia* **2018**, *807*, 297–312. [CrossRef]
7. Lucentini, L.; Puletti, M.E.; Ricciolini, C.; Gigliarelli, L.; Fontaneto, D.; Lanfaloni, L.; Bilò, F.; Natali, M.; Panara, F. Molecular and Phenotypic Evidence of a New Species of Genus *Esox* (Esocidae, Esociformes, Actinopterygii): The Southern Pike, *Esox flaviae*. *PLoS ONE* **2011**, *6*, e25218. [CrossRef]
8. Palandačić, A.; Bravničar, J.; Zupančič, P.; Šanda, R.; Snoj, A. Molecular Data Suggest a Multispecies Complex of *Phoxinus* (Cyprinidae) in the Western Balkan Peninsula. *Mol. Phylogenet. Evol.* **2015**, *92*, 118–123. [CrossRef]
9. Artaev, O.N.; Ermakov, O.A.; Vekhov, D.A.; Konovalov, A.F.; Levina, M.A.; Pozdeev, I.V.; Ruchin, A.B.; Alyushin, I.V.; Iljin, V.Y.; Levin, B.A. Genetic Screening of Distribution Pattern of Roaches *Rutilus rutilus* and *R. lacustris* (Cyprinidae) in Broad Range of Secondary Contact (Volga Basin). *Inl. Water Biol.* **2021**, *14*, 205–214. [CrossRef]
10. De Santis, V.; Delmastro, G.B.; Vanetti, I.; Britton, J.R.; Zaccara, S. Species Composition of Introduced and Natural Minnow Populations of the *Phoxinus* Cryptic Complex in the Westernmost Part of the Po River Basin (North Italy). *Biol. Invasions* **2021**, *23*, 657–668. [CrossRef]
11. Milana, V.; Šanda, R.; Vukić, J.; Ciccotti, E.; Riccato, F.; Petrosino, G.; Rossi, A.R. Far from Home: Genetic Variability of *Knipowitschia* Sp. from Italy Revealed Unexpected Species in Coastal Lagoons of the Tyrrhenian Coast. *Estuar. Coast. Shelf Sci.* **2021**, *251*, 107260. [CrossRef]
12. Rossi, A.R.; Milana, V.; Pulcini, D.; Cataudella, S.; Martinoli, M.; Tancioni, L. An Integrated Genetic and Morphological Approach to Clarify the Conservation Status of the Threatened Italian Endemic Species *Alburnus albidus* (Cypriniformes: Cyprinidae). *Hydrobiologia* **2016**, *770*, 73–87. [CrossRef]
13. Bernatchez, L. The Evolutionary History of Brown Trout (*Salmo trutta* L.) Inferred from Phylogeographic, Nested Clade, and Mismatch Analyses of Mitochondrial DNA Variation. *Evolution* **2001**, *55*, 351–379. [CrossRef] [PubMed]
14. Buonerba, L.; Zaccara, S.; Delmastro, G.B.; Lorenzoni, M.; Salzburger, W.; Gante, H.F. Intrinsic and Extrinsic Factors Act at Different Spatial and Temporal Scales to Shape Population Structure, Distribution and Speciation in Italian *Barbus* (Osteichthyes: Cyprinidae). *Mol. Phylogenet. Evol.* **2015**, *89*, 115–129. [CrossRef] [PubMed]
15. Buj, I.; Marčić, Z.; Caleta, M.; Sanda, R.; Geiger, M.F.; Freyhof, J.; Machordom, A.; Vukić, J. Ancient Connections among the European Rivers and Watersheds Revealed from the Evolutionary History of the Genus *Telestes* (Actinopterygii; Cypriniformes). *PLoS ONE* **2017**, *12*, e0187366. [CrossRef] [PubMed]
16. Myers, N.; Mittermeler, R.A.; Mittermeler, C.G.; da Fonseca, G.A.B.; Kent, J. Biodiversity Hotspots for Conservation Priorities. *Nature* **2000**, *403*, 853–858. [CrossRef]
17. Tierno de Figueroa, J.M.; López-Rodríguez, M.J.; Fenoglio, S.; Sánchez-Castillo, P.; Fochetti, R. Freshwater Biodiversity in the Rivers of the Mediterranean Basin. *Hydrobiologia* **2013**, *719*, 137–186. [CrossRef]

18. Seifertová, M.; Bryja, J.; Vyskočilová, M.; Martínková, N.; Šimková, A. Multiple Pleistocene Refugia and Post-Glacial Colonization in the European Chub (*Squalius cephalus*) Revealed by Combined Use of Nuclear and Mitochondrial Markers. *J. Biogeogr.* **2012**, *39*, 1024–1040. [CrossRef]
19. Lévêque, C.; Oberdorff, T.; Paugy, D.; Stiassny, M.L.J.; Tedesco, P.A. Global Diversity of Fish (Pisces) in Freshwater. *Hydrobiologia* **2008**, *595*, 545–567. [CrossRef]
20. Geiger, M.F.; Herder, F.; Monaghan, M.T.; Almada, V.; Barbieri, R.; Bariche, M.; Berrebi, P.; Bohlen, J.; Casal-Lopez, M.; Delmastro, G.B.; et al. Spatial Heterogeneity in the Mediterranean Biodiversity Hotspot Affects Barcoding Accuracy of Its Freshwater Fishes. *Mol. Ecol. Resour.* **2014**, *14*, 1210–1221. [CrossRef]
21. Bianco, P.G. An Update on the Status of Native and Exotic Freshwater Fishes of Italy. *J. Appl. Ichthyol.* **2014**, *30*, 62–77. [CrossRef]
22. Bianco, P.G. Factors Affecting the Distribution of Freshwater Fishes Especially in Italy. *Cybium* **1995**, *19*, 241–259.
23. AIIAD. Principi Guida Riguardanti Le Immissioni Di Fauna Ittica Nelle Acque Interne Italiane. 2021. Available online: http://www.aiiad.it/sito/images/docs/sistematica/AIIAD-Principi_guida_immissioni_fauna_ittica_05032021.pdf (accessed on 21 March 2022).
24. Rondinini, C.; Battistoni, A.; Peronace, V.; Teofili, C. Lista Rossa Dei Vertebrati Italiani. Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare, 2013. Available online: http://www.iucn.it/pdf/Comitato_IUCN_Lista_Rossa_dei_vertebrati_italiani.pdf (accessed on 22 March 2022).
25. Schönhuth, S.; Vukić, J.; Šanda, R.; Yang, L.; Mayden, R.L. Phylogenetic Relationships and Classification of the Holarctic Family Leuciscidae (Cypriniformes: Cyprinoidei). *Mol. Phylogenet. Evol.* **2018**, *127*, 781–799. [CrossRef] [PubMed]
26. Marchetto, F.; Zaccara, S.; Muenzel, F.M.; Salzburger, W. Phylogeography of the Italian Vairone (*Telestes muticellus*, Bonaparte 1837) Inferred by Microsatellite Markers: Evolutionary History of a Freshwater Fish Species with a Restricted and Fragmented Distribution. *BMC Evol. Biol.* **2010**, *10*, 111. [CrossRef] [PubMed]
27. Zaccara, S.; Quadroni, S.; de Santis, V.; Vanetti, I.; Carosi, A.; Britton, R.; Lorenzoni, M. Genetic and Morphological Analyses Reveal a Complex Biogeographic Pattern in the Endemic Barbel Populations of the Southern Italian Peninsula. *Ecol. Evol.* **2019**, *9*, 10185–10197. [CrossRef]
28. Rossi, A.R.; Petrosino, G.; Crescenzo, S.; Milana, V.; Talarico, L.; Martinoli, M.; Rakaj, A.; Lorenzoni, M.; Carosi, A.; Ciuffardi, L. Phylogeography and Population Structure of *Squalius lucumonis*: A Baseline for Conservation of an Italian Endangered Freshwater Fish. *J. Nat. Conserv.* **2021**, *64*, 126085. [CrossRef]
29. Nonnis Marzano, F.; Lorenzoni, M.; Tancioni, L. PESCI (Ciclostomi e Osteitti). In *Manuali per il Monitoraggio di Specie e Habitat di Interesse Comunitario (Direttiva 92/43/CEE) in Italia: Specie Animali*; Stoch, P.G.F., Ed.; ISPRA: Rome, Italy, 2016; pp. 129–190.
30. Bianco, P.G.; Ketmaier, V. A Revision of the Rutilus Complex from Mediterranean Europe with Description of a New Genus, *Sarmarutilus*, and a New Species, *Rutilus stoumboudae* (Teleostei: Cyprinidae). *Zootaxa* **2014**, *3841*, 379–402. [CrossRef]
31. Djait, H.; Bahri-Sfar, L.; Laouar, H.; Missaoui, N.; Ben Hassine, O.K. Dietary Comparison of Pike-Perch, *Sander lucioperca* (Linnaeus, 1758) and Catfish, *Silurus glanis* Linnaeus, 1758 in Sidi Salem Dam Reservoir (Tunisia). *Cybium* **2019**, *43*, 61–69. [CrossRef]
32. Keskin, E.; Unal, E.M.; Atar, H.H. Detection of Rare and Invasive Freshwater Fish Species Using EDNA Pyrosequencing: Lake Iznik Ichthyofauna Revised. *Biochem. Syst. Ecol.* **2016**, *67*, 29–36. [CrossRef]
33. Bianco, P.G.; Caputo, V.; Ferrito, V.; Lorenzoni, M.; Nonnis Marzano, F.; Stefani, F.; Sabatini, A.; Tancioni, L. *Rutilus rubilio*. In *Lista Rossa IUCN dei Vertebrati Italiani*; Rondinini, C., Battistoni, A., Peronace, V., Teofili, C., Eds.; Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare, 2013. Available online: http://www.iucn.it/scheda.php?id=-251974278 (accessed on 21 February 2022).
34. Di Tizio, L.; di Felice, P.L. *Rutilus rubilio* (Bonaparte, 1837) (Rovella). In *Manuali per il Monitoraggio di Specie e Habitat di Interesse Comunitario (Direttiva 92/43/CEE) in Italia: Specie Animali*; Stoch, F., Genovesi, P., Eds.; ISPRA: Rome, Italy, 2016; pp. 164–165.
35. EEC. Council Directive 92/43/ECC. *Off. J. Eur. Union* **1992**.
36. Marić, D. *Rutilus albus* Sp. n. (Teleostei: Cyprinidae) from Lake Skadar. *Period. Biol.* **2010**, *112*, 153–158.
37. Milošević, D.; Winkler, K.A.; Marić, D.; Weiss, S. Genotypic and Phenotypic Evaluation of *Rutilus* Spp. from Skadar, Ohrid and Prespa Lakes Supports Revision of Endemic as Well as Taxonomic Status of Several Taxa. *J. Fish Biol.* **2011**, *79*, 1094–1110. [CrossRef] [PubMed]
38. Stoch, F.; Genovesi, P. *Manuali per Il Monitoraggio Di Specie e Habitat Di Interesse Comunitario (Direttiva 92/43/CEE) in Italia: Habitat*; ISPRA: Rome, Italy, 2016.
39. Kottelat, M.; Freyhof, J. *Handbook of European Freshwater Fishes*; Publications Kottelat: Cornol, Switzerland, 2007.
40. Aljanabi, S.M.; Martinez, I. Universal and Rapid Salt-Extraction of High Quality Genomic DNA for PCR-Based Techniques. *Nucleic Acids Res.* **1997**, *25*, 4692–4693. [CrossRef] [PubMed]
41. Hebert, P.D.N.; Ratnasingham, S.; DeWaard, J.R. Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species. *Proc. R. Soc. B Biol. Sci.* **2003**, *270* (Suppl. 1), 96–99. [CrossRef]
42. Lieckfeldt, D.; Hett, A.K.; Ludwig, A.; Freyhof, J. Detection, Characterization and Utility of a New Highly Variable Nuclear Marker Region in Several Species of Cyprinid Fishes (Cyprinidae). *Eur. J. Wildl. Res.* **2006**, *52*, 63–65. [CrossRef]
43. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System: Barcoding. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]
44. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

45. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; Mcgettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
46. Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]
47. Clement, M.; Posada, D.; Crandall, K.A. TCS: A Computer Program to Estimate Gene Genealogies. *Mol. Ecol.* **2000**, *9*, 1657–1659. [CrossRef]
48. Leigh, J.W.; Bryant, D. POPART: Full-Feature Software for Haplotype Network Construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116. [CrossRef]
49. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids Res.* **2016**, *44*, W232–W235. [CrossRef] [PubMed]
50. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
51. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef] [PubMed]
52. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Bio.* **2012**, *61*, 539–542. [CrossRef] [PubMed]
53. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. JModelTest 2: More Models, New Heuristics and Parallel Computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef] [PubMed]
54. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]
55. Excoffier, L.; Laval, G.; Schneider, S. Arlequin (Version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evol. Bioinforma.* **2005**, *1*, 117693430500100. [CrossRef]
56. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: Mew York, NY, USA, 1987.
57. Kimura, M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef]
58. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
59. Adamack, A.T.; Gruber, B. PopGenReport: Simplifying Basic Population Genetic Analyses in R. *Methods Ecol. Evol.* **2014**, *5*, 384–387. [CrossRef]
60. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2020. Available online: https://www.R-project.org/ (accessed on 1 February 2022).
61. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 9.
62. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [CrossRef]
63. Tipton, M.L.; Gignoux-Wolfsohn, S.; Stonebraker, P.; Chernoff, B. Postglacial Recolonization of Eastern Blacknose Dace, *Rhinichthys atratulus* (Teleostei: Cyprinidae), through the Gateway of New England. *Ecol. Evol.* **2011**, *1*, 343–358. [CrossRef] [PubMed]
64. Yu, Y.; Harris, A.J.; He, X. S-DIVA (Statistical Dispersal-Vicariance Analysis): A Tool for Inferring Biogeographic Histories. *Mol. Phylogenet. Evol.* **2010**, *56*, 848–850. [CrossRef] [PubMed]
65. Yu, Y.; Harris, A.J.; Blair, C.; He, X. RASP (Reconstruct Ancestral State in Phylogenies): A Tool for Historical Biogeography. *Mol. Phylogenet. Evol.* **2015**, *87*, 46–49. [CrossRef] [PubMed]
66. Harpending, H.C. Signature of Ancient Population Growth in a Low-Resolution Mitochondrial DNA Mismatch Distribution. *Human Biol.* **1994**, *66*, 591–600.
67. Schneider, S.; Excoffier, L. Estimation of Past Demographic Parameters from the Distribution of Pairwise Differences When the Mutation Rates Vary among Sites: Application to Human Mitochondrial DNA. *Genetics* **1999**, *152*, 1079–1089. [CrossRef]
68. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **1989**, *123*, 585–595. [CrossRef]
69. Fu, Y. Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* **1997**, *147*, 915–925. [CrossRef]
70. Ramos-Onsins, S.E.; Rozas, J. Statistical Properties of New Neutrality Tests against Population Growth. *Mol. Biol. Evol.* **2002**, *19*, 2092–2100. [CrossRef] [PubMed]
71. Rogers, A.R.; Harpending, H. Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Mol. Biol. Evol.* **1992**, *9*, 552–569. [CrossRef] [PubMed]
72. Hashemzadeh Segherloo, I.; Ghojoghi, F.; Tabatabaei, S.N.; Normandeau, E.; Hernandez, C.; Hallerman, E.; Boyle, B.; Bernatchez, L. Population Genomics of the Southern Caspian Sea Vobla Rutilus Lacustris. *Hydrobiologia* **2021**, *848*, 345–361. [CrossRef]
73. Fricke, R.; Eschmeyer, W.N.; van der Laan, R. *Eschmeyer's Catalog of Fishes: Genera, Species, References*; Electronic Version. Available online: http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (accessed on 25 April 2022).

74. Oikonomou, A.; Leprieur, F.; Leonardos, I.D. Biogeography of Freshwater Fishes of the Balkan Peninsula. *Hydrobiologia* **2014**, *738*, 205–220. [CrossRef]
75. Tougard, C.; Justy, F.; Guinand, B.; Douzery, E.J.P.; Berrebi, P. *Salmo macrostigma* (Teleostei, Salmonidae): Nothing More than a Brown Trout (S. Trutta) Lineage? *J. Fish Biol.* **2018**, *93*, 302–310. [CrossRef]
76. Ketmaier, V.; Bianco, P.G.; Durand, J.D. Molecular Systematics, Phylogeny and Biogeography of Roaches (*Rutilus*, Teleostei, Cyprinidae). *Mol. Phylogenet. Evol.* **2008**, *49*, 362–367. [CrossRef]
77. Jurajda, P.; Pavlov, I. Rediscovery of *Rutilus virgo* in the River Dyje, Czech Republic. *Folia Zool.* **2016**, *65*, 98–100. [CrossRef]
78. Tsoumani, M.; Georgiadis, A.; Giantsis, I.A.; Leonardos, I.; Apostolidis, A.P. Phylogenetic Relationships among Southern Balkan *Rutilus* Species Inferred from Cytochrome b Sequence Analysis: Micro-Geographic Resolution and Taxonomic Implications. *Biochem. Syst. Ecol.* **2014**, *54*, 172–178. [CrossRef]
79. Levin, B.A.; Simonov, E.P.; Ermakov, O.A.; Levina, M.A.; Interesova, E.A.; Kovalchuk, O.M.; Malinina, Y.A.; Mamilov, N.S.; Mustafayev, N.J.; Pilin, D.V.; et al. Phylogeny and Phylogeography of the Roaches, Genus *Rutilus* (Cyprinidae), at the Eastern Part of Its Range as Inferred from MtDNA Analysis. *Hydrobiologia* **2017**, *788*, 33–46. [CrossRef]
80. Ward, R.D. DNA Barcode Divergence among Species and Genera of Birds and Fishes. *Mol. Ecol. Resour.* **2009**, *9*, 1077–1085. [CrossRef]
81. Wang, T.; Zhang, Y.P.; Yang, Z.Y.; Liu, Z.; Du, Y.Y. DNA Barcoding Reveals Cryptic Diversity in the Underestimated Genus *Triplophysa* (Cypriniformes: Cobitidae, Nemacheilinae) from the Northeastern Qinghai-Tibet Plateau. *BMC Evol. Biol.* **2020**, *20*, 151. [CrossRef] [PubMed]
82. Pereira, L.H.G.; Castro, J.R.C.; Vargas, P.M.H.; Gomez, J.A.M.; Oliveira, C. The Use of an Integrative Approach to Improve Accuracy of Species Identification and Detection of New Species in Studies of Stream Fish Diversity. *Genetica* **2021**, *149*, 103–116. [CrossRef] [PubMed]
83. Guimarães, K.L.A.; Lima, M.P.; Santana, D.J.; de Souza, M.F.B.; Barbosa, R.S.; Rodrigues, L.R.R. DNA Barcoding and Phylogeography of the *Hoplias malabaricus* Species Complex. *Sci. Rep.* **2022**, *12*, 5288. [CrossRef] [PubMed]
84. Tsoupas, A.; Papavasileiou, S.; Minoudi, S.; Gkagkavouzis, K.; Petriki, O.; Bobori, D.; Sapounidis, A.; Koutrakis, E.; Leonardos, I.; Karaiskou, N.; et al. DNA Barcoding Identification of Greek Freshwater Fishes. *PLoS ONE* **2022**, *17*, e0263118. [CrossRef] [PubMed]
85. Valić, D.; Vardić Smrzlić, I.; Kapetanović, D.; Teskeredžić, Z.; Pleše, B.; Teskeredžić, E. Identification, Phylogenetic Relationships and a New Maximum Size of Two Rudd Populations (*Scardinius*, Cyprinidae) from the Adriatic Sea Drainage, Croatia. *Biologia* **2013**, *68*, 539–545. [CrossRef]
86. Scribner, K.T.; Page, K.S.; Bartron, M.L. Hybridization in Freshwater Fishes: A Review of Case Studies and Cytonuclear Methods of Biological Inference. *Rev. Fish Biol. Fish.* **2000**, *10*, 293–323. [CrossRef]
87. Hayden, B.; Pulcini, D.; Kelly-Quinn, M.; O'Grady, M.; Caffrey, J.; McGrath, A.; Mariani, S. Hybridisation between Two Cyprinid Fishes in a Novel Habitat: Genetics, Morphology and Life-History Traits. *BMC Evol. Biol.* **2010**, *10*, 169. [CrossRef]
88. Sousa-Santos, C.; Matono, P.; Silva, J.D.A.; Ilhéu, M. Evaluation of Potential Hybridization between Native Fishes and the Invasive Bleak, *Alburnus alburnus* (Actinopterygii: Cypriniformes: Cyprinidae). *Acta Ichthyol. Piscat.* **2018**, *48*, 109–122. [CrossRef]
89. Freyhof, J.; Lieckfeldt, D.; Pitra, C.; Ludwig, A. Molecules and Morphology: Evidence for Introgression of Mitochondrial DNA in Dalmatian Cyprinids. *Mol. Phylogenet. Evol.* **2005**, *37*, 347–354. [CrossRef]
90. Carranza, S.; Romano, A.; Arnold, E.N.; Sotgiu, G. Biogeography and Evolution of European Cave Salamanders, *Hydromantes* (Urodela: Plethodontidae), Inferred from MtDNA Sequences. *J. Biogeogr.* **2008**, *35*, 724–738. [CrossRef]
91. Mattoccia, M.; Marta, S.; Romano, A.; Sbordoni, V. Phylogeography of an Italian Endemic Salamander (Genus *Salamandrina*): Glacial Refugia, Postglacial Expansions, and Secondary Contact. *Biol. J. Linn. Soc.* **2011**, *104*, 903–992. [CrossRef]
92. Canestrelli, D.; Salvi, D.; Maura, M.; Bologna, M.A.; Nascetti, G. One Species, Three Pleistocene Evolutionary Histories: Phylogeography of the Italian Crested Newt, *Triturus carnifex*. *PLoS ONE* **2012**, *7*, e41754. [CrossRef] [PubMed]
93. Maura, M.; Salvi, D.; Bologna, M.A.; Nascetti, G.; Canestrelli, D. Northern Richness and Cryptic Refugia: Phylogeography of the Italian Smooth Newt *Lissotriton vulgaris meridionalis*. *Biol. J. Linn. Soc.* **2014**, *113*, 590–603. [CrossRef]
94. Schultze, N.; Spitzweg, C.; Corti, C.; Delaugerre, M.; di Nicola, M.R.; Geniez, P.; Lapini, L.; Liuzzi, C.; Lunghi, E.; Novarini, N.; et al. Mitochondrial Ghost Lineages Blur Phylogeography and Taxonomy of *Natrix helvetica* and *N. natrix* in Italy and Corsica. *Zool. Scr.* **2020**, *49*, 395–411. [CrossRef]
95. Stefani, F.; Galli, P.; Crosa, G.; Zaccara, S.; Calamari, D. Alpine and Apennine Barriers Determining the Differentiation of the Rudd (*Scardinius erythrophthalmus* L.) in the Italian Peninsula. *Ecol. Freshw. Fish* **2004**, *13*, 168–175. [CrossRef]
96. Schmitt, T.; Fritz, U.; Delfino, M.; Ulrich, W.; Habel, J.C. Biogeography of Italy Revisited: Genetic Lineages Confirm Major Phylogeographic Patterns and a Pre-Pleistocene Origin of Its Biota. *Front. Zool.* **2021**, *18*, 34. [CrossRef] [PubMed]
97. Lorenzoni, M.; Carosi, A.; Quadroni, S.; de Santis, V.; Vanetti, I.; Delmastro, G.B.; Zaccara, S. Cryptic Diversity within Endemic Italian Barbels: Revalidation and Description of New Barbus Species (Teleostei: Cyprinidae). *J. Fish Biol.* **2021**, *98*, 1433–1449. [CrossRef]
98. Raggi, G. Neotettonica ed evoluzione paleogeografica plio-pleistocenica del bacino del fiume Magra. *Mem. Della Soc. Geol. Ital.* **1985**, *30*, 35–62.
99. Ciuffardi, L.; Oneto, F.; Raineri, V. L'ittiofauna Delle Acque Interne Della Liguria: Aspetti Filogeografici e Distributivi Rilevanti Ai Fini Dell'applicazione Della Direttiva 2000/60/CE. *Ann. Mus. Civ. Stor. Nat. Giacomo Doria* **2015**, *107*, 213–283.

100. Zaccara, S.; Delmastro, G.B. Tyrrhenian Basins of Ligury as a New Peri-Mediterranean Ichthyogeographic District? Population Structure of *Telestes muticellus* (Osteichthyes, Cyprinidae), a Primary Freshwater Fish. *Hydrobiologia* **2009**, *632*, 285–295. [CrossRef]
101. Bianco, P.G. L'Ittiofauna Continentale Dell'Appennino Umbro-Marchigiano, Barriera Semipermeabile Allo Scambio Di Componenti Primarie Tra Gli Opposti Versanti Dell'Italia Centrale. *Biogeogr.–J. Integr. Biogeogr.* **1994**, *17*, 427–485. [CrossRef]
102. Piacentini, T.; Miccadei, E. The Role of Drainage Systems and Intermontane Basins in the Quaternary Landscape of the Central Apennines Chain (Italy). *Rend. Lincei* **2014**, *25*, 139–150. [CrossRef]
103. Splendiani, A.; Giovannotti, M.; Cerioni, P.N.; Caniglia, M.L.; Caputo, V. Phylogeographic Inferences on the Native Brown Trout MtDNA Variation in Central Italy. *Ital. J. Zool.* **2006**, *73*, 179–189. [CrossRef]
104. Rossi, A.R.; Talarico, L.; Petrosino, G.; Crescenzo, S.; Tancioni, L. Conservation Genetics of Mediterranean Brown Trout in Central Italy (Latium): A Multi-Marker Approach Anna. *Water* **2022**, *14*, 937. [CrossRef]
105. Taviani, S.; Henriksen, H.J. The Application of a Groundwater/Surface-Water Model to Test the Vulnerability of Bracciano Lake (near Rome, Italy) to Climatic and Water-Use Stresses. *Hydrogeol. J.* **2015**, *23*, 1481–1498. [CrossRef]
106. Cavinato, G.P.; Miccadei, E.; Parotto, M. Stato Dell'arte Delle Conoscenze Sulla Geologia Plio-Quaternaria Dell'Italia Centrale (Settore Laziale). *Studi Geol. Camerti Vol. Spec.* **1992**, *1*, 27–31.
107. Aucelli, P.P.C.; D'Argenio, B.; della Seta, M.; Giano, S.I.; Schiattarella, M. Foreword: Intermontane Basins: Quaternary Morphoevolution of Central-Southern Italy. *Rend. Lincei* **2014**, *25*, 107–110. [CrossRef]
108. Bernatchez, L.; Wilson, C.C. Comparative Phylogeography of Nearctic and Palearctic Fishes. *Mol. Ecol.* **1998**, *7*, 431–452. [CrossRef]
109. Tschá, M.K.; Bachmann, L.; Abilhoa, V.; Boeger, W.A. Past Connection and Isolation of Catchments: The Sea-hanges Affect the Distribution and Genetic Variability of Coastal Freshwater Fishes. *Estuar. Coast. Shelf Sci.* **2017**, *190*, 31–39. [CrossRef]
110. Lambeck, K.; Antonioli, F.; Purcell, A.; Silenzi, S. Sea-Level Change along the Italian Coast for the Past 10,000 Yr. *Quat. Sci. Rev.* **2004**, *23*, 1567–1598. [CrossRef]
111. Vai, G.B.; Cantelli, L. *Litho-Palaeoenvironmental Maps of Italy During the Last Two Climatic Extremes*; Museo Geologico Giovanni Capellini: Bologna, Italy, 2004.
112. Grant, W.S.; Bowen, B.W. Shallow Population Histories in Deep Evolutionary Lineages of Marine Fishes: Insights from Sardines and Anchovies and Lessons for Conservation. *J. Hered.* **1998**, *89*, 415–426. [CrossRef]
113. Bianco, P.G. Mediterranean Endemic Freshwater Fishes of Italy. *Biol. Conserv.* **1995**, *72*, 159–170. [CrossRef]
114. Chiesa, S.; Filonzi, L.; Vaghi, M.; Papa, R.; Marzano, F.N. Molecular Barcoding of an Atypical Cyprinid Population Assessed by Cytochrome b Gene Sequencing. *Zoolog. Sci.* **2013**, *30*, 408–413. [CrossRef] [PubMed]
115. Celletti, P.; Giovinazzo, C.; Molinaro, A.; Bidditu, I.; Zarattini, A. A Middle Pleistocene Deposit with Elephas Antiquus Remains near Colleferro (Roma). In *The World of Elephants. Proceedings of the First International Congress. Consiglio Nazionale delle Ricerche*; The World of Elephants—International Congres: Roma, Italy, 2001; pp. 34–37.
116. Bersani, P.; Corda, L. New Data on the Tuscolano-Artemisio Phase of the Alban Hills: Some Insights on Climatic Conditions. *J. Mediterr. Earth Sci.* **2011**, *3*, 25–32. [CrossRef]
117. Aiello, G.; Marsella, E.; Sacchi, M. Quaternary Structural Evolution of Terracina and Gaeta Basins (Eastern Tyrrhenian Margin, Italy). *Rend. Lincei* **2000**, *11*, 41–58. [CrossRef]
118. Bellucci, D.; Novaga, R.; Freyhof, J. New Data on the Distribution of the Volturno Spined Loach *Cobitis zanandreai* (Teleostei: Cobitidae). *J. Appl. Ichthyol.* **2021**, *37*, 885–892. [CrossRef]
119. Carosi, A.; Ghetti, L.; Forconi, A.; Lorenzoni, M. Fish Community of the River Tiber Basin (Umbria-Italy): Temporal Changes and Possible Threats to Native Biodiversity. *Knowl. Manag. Aquat. Ecosyst.* **2015**, *416*, 22. [CrossRef]
120. Allendorf, F.W.; Luikart, G.H.; Aitken, S.N. *Conservation and the Genetics of Populations*; Wiley-Blackwell: Oxford, UK, 2013.
121. Ward, D.R.; Zemlak, T.S.; Innes, B.H.; Last, P.R.; Hebert, P.D.N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B* **2005**, *360*, 1847–1857. [CrossRef]
122. Gilles, A.; Lecointre, G.; Miquelis, A.; Loerstcher, M.; Chappaz, R.; Brun, G. Partial Combination Applied to Phylogeny of European Cyprinids Using the Mitochondrial Control Region. *Mol. Phylogenet. Evol.* **2001**, *19*, 22–33. [CrossRef]
123. Chow, S.; Hazama, K. Universal primer for S7 ribosomal protein gene intron in fish. *Mol. Ecol.* **1998**, *7*, 1255–1256.

**Supplementary Material for**

# Phylogeography of Sarmarutilus rubilio (Cypriniformes: Leuciscidae): complex genetic structure, clues to a new cryptic species and further insights into roaches phylogeny

**Gerardo Petrosino [1,*], Lorenzo Tancioni [2], Martina Turani [1], Arnold Rakaj [2], Luca Ciuffardi [3], and Anna Rita Rossi [1]**

**1 Department of Biology and Biotechnology "C. Darwin", Sapienza University of Rome, 00185 Rome, Italy; gerardo.petrosino@uniroma1.it (G.P.); annarita.rossi@uniroma1.it (A.R.R.)**
**2 Department of Biology, University of Rome "Tor Vergata", 00133 Rome, Italy; tancioni@uniroma2.it (L.T.); arnoldrakaj@gmail.com (A.R.)**
**3 Center for BioNatural Studies srl, 16132 Genova, Italy; info@lucaciuffardi.it**
*** Correspondence: gerardo.petrosino@uniroma1.it**

**Table S1.** PCR conditions and primers for each amplified marker and PCR product length. Total volume reaction = 20 μl: 2 μl of 10x $NH_4$ Buffer, 0.6 μl of 50 mM $MgCl_2$, 0.4 μl of dNTP mix (2.5 μM each dNTP), 0.2 μl of 100 μM of each primer, 0.14 μl of 5 $U\mu l^{-1}$ of BIOTAQ™ DNA Polymerase (Bioline) and 10–100 ng of template DNA.

| Marker | Primer sequence | Reference | PCR conditions | Gene length (bp) |
|---|---|---|---|---|
| COI | FISHF1: 5'TCAACCAACCACAAAGACATTGGCAC3'<br>FishR2 5'-ACTTCAGGGTGACCGAAGAATCAGAA-3' | [121] | 95ºC, 2 min - 30 cycles: 94ºC, 30 s; 54ºC, 30 s; 72º, 1 min - 72ºC, 10 min. | 624 |
| CR | ESTFOR: 5'CATCGGTCTTGTAATCCGAAGAT3'<br>PHE1R: 5'ACATCTTCAGTGTTACGCTT3' | [122] | 95ºC, 2 min - 30 cycles: 94ºC, 30 s; 54ºC, 30 s; 72º, 1 min - 72ºC, 10 min. | 929–930 |
| Cyfun P | Cyp_un FLP1: 5'AAGTGGTGCATCGTGTTGTG3'<br>Cyp_unFL1R: 5'CAGCCTGAACAATCAAAACAG3' | [42] | 94ºC, 3 min - 35 cycles: 94ºC, 30 s; 55ºC, 30 s; 72º, 1 min - 72ºC, 10 min. | 156–218* |
| S7 | S7RPEX1F: 5'TGGCCTCTTCCTTGGCCGTC3'<br>S7RPEX2R: 5'AACTCGTCTGGCTTTTCGCC3' | [123] | 95ºC, 3 min - 30 cycles: 95ºC, 30 s; 54ºC, 1 min; 72º, 2 min - 72ºC, 10 min. | 319–861* |

*total alignment including gaps

## References

121. Lieckfeldt, D.; Hett, A.K.; Ludwig, A.; Freyhof, J. Detection, Characterization and Utility of a New Highly Variable Nuclear Marker Region in Several Species of Cyprinid Fishes (Cyprinidae). *Eur. J. Wildl. Res*. **2006**, *52*, 63–65. https://doi.org/10.1007/s10344-005-0010-6.
122. Ward, D.R.; Zemlak T.S.; Innes B.H.; Last P.R.; Hebert, P.D.N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B*, **2005**, *360*, 1847–1857. https://doi.org/10.1098/rstb.2005.1716.

123. Gilles, A.; Lecointre, G.; Miquelis, A.; Loerstcher, M.; Chappaz, R.; Brun, G. Partial Combination Applied to Phylogeny of European Cyprinids Using the Mitochondrial Control Region. *Mol. Phylogenetics Evol.*, **2001**, *19*, 22–33. https://doi.org/10.1006/mpev.2000.0916.
124. Chow, S.; Hazama, K. Universal primer for S7 ribosomal protein gene intron in fish. *Mol. Ecol.*, **1998**, *7*, 1255–1256.

**Table S2**. COI Sequences retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and Barcode Of Life Data SYSTEM (www.boldsystems.org, in yellow) included in the phylogenetic reconstruction.

| Species label from databases (Genus *Rutilus*) | Accession Number | Current valid species name (Eschmeyer Catalogue of Fishes, https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes) | Country (basin) | ID site (only *S. rubilio*, see Figure 1) |
|---|---|---|---|---|
| *R. albus* | MG806869 | *Leucos albus* | Albania (Drin) | |
| *R. albus* | KJ554332 | *Leucos albus* | Albania (Skadar lake) | |
| *R. albus* | KJ554244 | *Leucos albus* | Albania (Skadar lake) | |
| *R. albus* | KJ554242 | *Leucos albus* | Albania (Skadar lake) | |
| *R. aula* | KJ554532 | *Leucos aula* | Italy (Po) | |
| *R. aula* | KJ554171 | *Leucos aula* | Croatia (Zadar) | |
| *R. aula* | HM560323 | *Leucos aula* | Croatia (Bacisca lake) | |
| *R. aula* | MG806870 | *Leucos aula* | Croatia (Zrmanja) | |
| *R. basak* | MG806871 | *Leucos basak* | Bosnia and Herzegovina (Neretva) | |
| *R. basak* | HM560325 | *Leucos basak* | Bosnia and Herzegovina (Neretva) | |
| *R. basak* | KJ554498 | *Leucos basak* | Bosnia and Herzegovina (Neretva) | |
| *R. basak* | KJ554191 | *Leucos basak* | Bosnia and Herzegovina (Mandek) | |
| *R. frisii* | MG806872 | *Rutilus frisii* | Bulgaria (Rezowska) | |
| *R. frisii* | KJ554245 | *Rutilus frisii* | Bulgaria (Rezowska) | |
| *R. frisii* | MT756353 | *Rutilus frisii* | Iran (Gorgan Gulf Caspian Sea) | |
| *R. frisii* | MT756352 | *Rutilus frisii* | Iran (Gorgan Gulf Caspian Sea) | |
| *R. heckelii* | KJ554478 | *Rutilus lacustris* | Greece (Volvi lake) | |
| *R. heckelii* | KJ554441 | *Rutilus lacustris* | Greece (Volvi lake) | |
| *R. heckelii* | KJ554358 | *Rutilus lacustris* | Greece (Volvi lake) | |
| *R. heckelii* | KJ554261 | *Rutilus lacustris* | Greece (Volvi lake) | |
| *R. kutum* | AQM026-16 | *Rutilus kutum* | Iran (Mazandaran Province) | |
| *R. kutum* | AQM027-16 | *Rutilus kutum* | Iran (Mazandaran Province) | |
| *R. kutum* | AQM028-16 | *Rutilus kutum* | Iran (Mazandaran Province) | |

| *R. kutum* | AQM029-16 | *Rutilus kutum* | Iran (Mazandaran Province) | |
|---|---|---|---|---|
| *R. lacustris* | MT756380 | *Rutilus lacustris* | Iran (Gorgan Gulf Caspian Sea) | |
| *R. lacustris* | HQ561910 | *Rutilus lacustris* | Georgia (Rioni) | |
| *R. lacustris* | HQ561909 | *Rutilus lacustris* | Georgia (Rioni) | |
| *R. lacustris* | MG806873 | *Rutilus lacustris* | Greece (Struma) | |
| *R. meidingeri* | KR477255 | *Rutilus meidingeri* | Austria (Mondsee lake) | |
| *R. meidingeri* | KR477254 | *Rutilus meidingeri* | Austria (Mondsee lake) | |
| *R. meidingeri* | KR477253 | *Rutilus meidingeri* | Austria (Mondsee lake) | |
| *R. meidingeri* | KR477099 | *Rutilus meidingeri* | Austria (Mondsee lake) | |
| *R. ohridanus* | MG806874 | *Leucos basak* | Albania (Ohrid lake) | |
| *R. ohridanus* | KJ554509 | *Leucos basak* | Albania (Ohrid lake) | |
| *R. ohridanus* | KJ554414 | *Leucos basak* | Albania (Ohrid lake) | |
| *R. ohridanus* | KJ554350 | *Leucos basak* | Albania (Ohrid lake) | |
| *R. panosi* | MG806875 | *Leucos panosi* | Greece (Trichonis lake) | |
| *R. panosi* | KJ554528 | *Leucos panosi* | Greece (Trichonis lake) | |
| *R. panosi* | KJ554495 | *Leucos panosi* | Greece (Trichonis lake) | |
| *R. panosi* | KJ554396 | *Leucos panosi* | Greece (Trichonis lake) | |
| *R. pigus* | HM560327 | *Rutilus pigus* | Italy (Adda) | |
| *R. pigus* | HM560326 | *Rutilus pigus* | Italy (Adda) | |
| *R. pigus* | KJ554485 | *Rutilus pigus* | Italy (Como lake) | |
| *R. pigus* | KJ554328 | *Rutilus pigus* | Italy (Como lake) | |
| *R. prespensis* | KJ554482 | *Leucos basak* | Albania (Prespa lake) | |
| *R. prespensis* | KJ554471 | *Leucos basak* | Albania (Prespa lake) | |
| *R. prespensis* | KJ554468 | *Leucos basak* | Albania (Prespa lake) | |
| *R. prespensis* | KJ554445 | *Leucos basak* | Albania (Prespa lake) | |
| *R. rubilio* | KJ554475 | *Sarmarutilus rubilio* | Italy (Arno) | ARN-Cas |
| *R. rubilio* | KJ554114 | *Sarmarutilus rubilio* | Italy (Arno) | ARN-Cas |
| *R. rubilio* | KJ554324 | *Sarmarutilus rubilio* | Italy (Arno) | ARN-Ter |
| *R. rubilio* | KJ554212 | *Sarmarutilus rubilio* | Italy (Arno) | ARN-Ter |
| *R. rubilio* | KJ554352 | *Sarmarutilus rubilio* | Italy (Tiber) | TIB-Ang |

| | | | | |
|---|---|---|---|---|
| *R. rubilio* | KJ554250 | *Sarmarutilus rubilio* | Italy (Tiber) | TIB-Ang |
| *R. rubilio* | KJ554361 | *Sarmarutilus rubilio* | Italy (Tiber) | TIB-Mon |
| *R. rubilio* | KJ554088 | *Sarmarutilus rubilio* | Italy (Tiber) | TIB-Mon |
| *R. rubilio* | KJ554380 | *Sarmarutilus rubilio* | Italy (Sisto) | SIS |
| *R. rubilio* | KJ554226 | *Sarmarutilus rubilio* | Italy (Sisto) | SIS |
| *R. rubilio* | KJ554435 | *Sarmarutilus rubilio* | Italy (Fondi lake) | FON |
| *R. rubilio* | KJ554290 | *Sarmarutilus rubilio* | Italy (Fondi lake) | FON |
| *R. rubilio* | KJ554508 | *Sarmarutilus rubilio* | Italy (Volturno) | VOL |
| *R. rubilio* | KJ554364 | *Sarmarutilus rubilio* | Italy (Volturno) | VOL |
| *R. rubilio* | KJ554200 | *Sarmarutilus rubilio* | Italy (Volturno) | VOL |
| *R. rubilio* | KJ554118 | *Sarmarutilus rubilio* | Italy (Volturno) | VOL |
| *R. rubilio* | KJ554409 | *Sarmarutilus rubilio* | Italy (Crati) | CRA |
| *R. rubilio* | KJ554274 | *Sarmarutilus rubilio* | Italy (Crati) | CRA |
| *R. rubilio* | KJ554129 | *Sarmarutilus rubilio* | Italy (Crati) | CRA |
| *R. rutilus* | KT989765 | *Rutilus rutilus* | Russia (Plescheevo lake) | |
| *R. rutilus* | HQ961042 | *Rutilus rutilus* | Czech Republic (Ohre) | |
| *R. rutilus* | MW473258 | *Rutilus rutilus* | Germany (Rhine) | |
| *R. rutilus* | HM392103 | *Rutilus rutilus* | Germany (Danube) | |
| *R. virgo* | HM392106 | *Rutilus virgo* | Germany (Danube) | |
| *R. virgo* | HM392104 | *Rutilus virgo* | Germany (Vils) | |
| *R. virgo* | HM392102 | *Rutilus virgo* | Germany (Danube) | |
| *R. virgo* | MG806878 | *Rutilus virgo* | Croatia (Sava) | |
| *R. ylikiensis* | MG806879 | *Leucos ylikiensis* | Greece (Yliki lake) | |
| *R. ylikiensis* | KJ554524 | *Leucos ylikiensis* | Greece (Kifissos) | |
| *R. ylikiensis* | KJ554516 | *Leucos ylikiensis* | Greece (Kifissos) | |
| *R. ylikiensis* | KJ554316 | *Leucos ylikiensis* | Greece (Kifissos) | |

**Table S3.** COI (a) and CR (b) diagnostic sites identified among the most frequent sequences of the three haplogroups found in Italian specimens. Deletions are marked with "–".

(a) Diagnostic COI (624 bp) sites

| Position/ Haplotype | 5 4 | 9 6 | 1 4 4 | 1 9 8 | 2 1 6 | 2 1 9 | 2 5 0 | 2 5 2 | 2 5 6 | 2 7 0 | 2 7 6 | 2 7 9 | 2 8 5 | 2 9 4 | 3 0 3 | 3 0 9 | 3 3 0 | 3 8 4 | 4 0 8 | 4 2 9 | 4 3 5 | 4 5 3 | 4 6 2 | 4 7 7 | 4 9 5 | 4 0 4 | 5 2 2 | 5 3 4 | 5 5 5 | 5 6 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HpA01 | G | T | C | G | T | G | T | G | C | C | A | T | C | A | A | G | C | G | C | T | G | A | G | A | C | C | A | A | T | C |
| HpB01 | G | T | C | G | T | G | T | G | C | C | G | T | C | A | A | A | C | G | T | C | G | G | G | A | C | T | G | A | T | T |
| HpC01 | A | C | T | A | C | A | C | A | T | T | A | C | T | G | G | G | T | C | T | T | A | G | A | G | A | C | G | G | C | T |

(b) Diagnostic CR (930 bp) sites

| Position/ Haplotype | 1 4 7 | 3 2 5 | 3 5 5 | 3 8 6 | 4 4 2 | 4 7 1 | 4 7 2 | 4 9 6 | 4 9 3 | 5 1 7 | 5 6 3 | 5 6 5 | 5 7 3 | 5 9 5 | 6 0 3 | 6 1 4 | 6 5 8 | 6 7 0 | 6 7 3 | 7 2 0 | 7 2 4 | 7 8 5 | 8 0 3 | 8 3 8 | 8 3 5 | 8 4 9 | 8 7 4 | 8 7 3 | 8 8 7 | 8 8 9 | 8 9 4 | 9 9 7 | 9 1 6 | 9 3 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HpA01 | C | A | G | T | C | C | G | G | T | G | T | T | A | T | G | A | C | T | T | T | C | A | T | A | G | G | G | A | G | T | G | G | T | A | C |
| HpB01 | G | A | A | T | C | C | G | G | A | G | C | C | A | A | G | A | T | T | T | T | T | T | T | G | A | A | G | A | G | T | A | G | T | A | C |
| HpC01 | C | G | A | C | T | A | T | T | A | T | T | T | G | T | A | T | T | C | C | C | T | T | A | A | A | A | A | G | A | G | A | A | - | G | T |

**Table S4.** Diagnostic fragment of nuclear marker Cyfun P in Italian specimens subset. Individuals with specific VAR sequences are highlighted in green, variable sites/indels in yellow; mtDNA lineage of each specimen is also reported. Missing data are marked with N, deletions with -.

| | CyfunP GenBank Accession Number | MtDNA haplogroup | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VAR1_46 | OM966266 | C | A | A | C | A | A | A | A | A | A | C | A | C | A | R | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR1_49 | OM966267 | C | A | A | C | A | W | A | A | A | A | C | A | C | A | N | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR2_02 | OM966268 | A | A | A | C | A | A | A | A | A | A | C | A | C | A | G | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR2_09 | OM966269 | C | A | A | C | A | A | A | A | A | A | C | A | C | A | N | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR2_11 | OM966270 | C | A | A | C | A | A | A | A | A | A | C | A | N | N | N | N | N | N | - | - | - | N | N | N | N | N | N | N | N | N | N | N | - | - | - | - | - | - |
| VAR2_25 | OM966271 | C | A | A | C | A | A | A | A | A | A | C | A | C | A | N | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR2_32 | OM966272 | C | A | A | C | A | W | A | A | A | A | C | A | C | A | N | A | A | A | - | - | - | A | A | A | T | A | A | A | A | G | G | A | - | - | - | - | - | - |
| VAR1_33 | OM966273 | A | T | A | A | A | A | A | A | A | T | A | A | A | A | N | T | A | A | A | A | T | A | A | A | T | A | A | A | N | N | N | N | N | N | N | N | N | N |
| VAR1_43 | OM966274 | A | T | A | A | A | W | A | A | A | T | A | A | A | A | A | T | A | A | A | A | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N |
| VAR2_04 | OM966275 | C | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | N | N | N | N | N | N | N | N | N | N | N |
| VAR2_05 | OM966276 | C | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | A | A | G | G | A | A | G | T | T | T | C |
| VAR2_06 | OM966277 | C | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | A | A | G | G | A | A | G | T | T | T | C |
| VAR2_28 | OM966278 | A | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | A | A | G | G | A | A | G | T | T | T | C |
| TIB1_26 | OM966279 | A | T | A | A | A | T | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | N | N | N | N | N | N | N | N | N | N | N | N | N | N |
| SCR_01 | OM966280 | A | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | A | A | G | G | A | A | G | T | T | T | C |
| SCR_02 | OM966281 | A | T | A | A | A | A | A | A | A | T | A | A | A | A | A | T | A | A | A | A | T | A | A | A | T | A | A | A | A | G | G | A | A | G | T | T | T | C |

**Table S5.** Inter- (below the diagonal) and intra-haplogroup (diagonal, in grey) genetic mean distances (CR); pairwise ΦST between CR haplogroups (above the diagonal).

| | A | B | C |
|---|---|---|---|
| A | 0.0009 (±0.0005) | 0.94*** | 0.97*** |
| B | 0.0151 (±0.0040) | 0.0013 (±0.0005) | 0.96*** |
| C | 0.0325 (±0.0062) | 0.0291 (±0.0058) | 0.0005 (±0.0005) |

**Table S6.** Population $\Phi_{ST}$ (CR) considering all the populations as they are (a) and splitting individuals by haplogroup for each one (b). Significance thresholds: * = P < 0.05; ** = P < 0.01; *** = P < 0.001.

(a)

| | | Haplogroup A/C | | Haplogroup A | | | | | | Haplogroup A/B | | Haplogroup B | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | VAR1 | VAR2 | TRO | FOR | ARR | TIB1 | TIB2 | GAR | SCR | NOC | SET | SPU |
| HpA/C | VAR1 | - | | | | | | | | | | | |
| | VAR2 | -0.05 | - | | | | | | | | | | |
| HpA | TRO | 0.29 | 0.3 | - | | | | | | | | | |
| | FOR | 0.28 | 0.33 | 0.05 | - | | | | | | | | |
| | ARR | 0.31*** | 0.35 | 0.68*** | 0.64*** | - | | | | | | | |
| | TIB1 | 0.38*** | 0.43*** | 0.13 | 0.13 | 0.45*** | - | | | | | | |
| | TIB2 | 0.06 | 0.16 | 0.20 | 0.19 | 0.47 | 0.13 | - | | | | | |
| | GAR | 0.33 | 0.38 | 0.07 | 0.08 | 0.61*** | 0.15*** | 0.01 | - | | | | |
| HpA/B | SCR | 0.32*** | 0.38** | 0.18 | 0.17 | 0.36*** | 0.21*** | 0.01 | 0.11 | - | | | |
| | NOC | 0.09 | 0.16 | 0.42*** | 0.38*** | 0.31*** | 0.41*** | 0.01 | 0.36*** | 0.12 | - | | |
| HpB | SET | 0.65*** | 0.58*** | 0.98*** | 0.97*** | 0.96*** | 0.95*** | 0.96*** | 0.97*** | 0.90*** | 0.87*** | - | |
| | SPU | 0.51*** | 0.47*** | 0.98*** | 0.96*** | 0.95*** | 0.95*** | 0.95 | 0.97*** | 0.88*** | 0.79*** | 0.12 | - |

(b)

| | | Haplogroup C | | Haplogroup A | | | | | | | | | | Haplogroup B | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | VAR1_C | VAR2_C | VAR1_A | VAR2_A | TRO | FOR | ARR | TIB1 | TIB2 | SCR_A | GAR | NOC_A | SET | SPU | SCR_B | NOC_B |
| HpC | VAR1_C | - | | | | | | | | | | | | | | | |
| | VAR2_C | -0.07 | - | | | | | | | | | | | | | | |
| HpA | VAR1_A | 0.99 | 0.99*** | - | | | | | | | | | | | | | |
| | VAR2_A | 0.99 | 0.99*** | -0.03 | - | | | | | | | | | | | | |
| | TRO | 0.99*** | 0.99*** | -0.04 | 0.01 | - | | | | | | | | | | | |
| | FOR | 0.99 | 0.99*** | 0.02 | 0.04 | 0.05 | - | | | | | | | | | | |
| | ARR | 0.97*** | 0.98*** | 0.63 | 0.63*** | 0.68*** | 0.64*** | - | | | | | | | | | |
| | TIB1 | 0.97*** | 0.98*** | 0.09 | 0.10 | 0.13 | 0.13 | 0.45*** | - | | | | | | | | |
| | TIB2 | 0.97 | 0.98 | 0.22 | 0.18 | 0.20 | 0.19 | 0.47 | 0.13 | - | | | | | | | |
| | SCR_A | 0.97*** | 0.97*** | 0.23 | 0.19 | 0.28** | 0.27** | 0.49*** | 0.28*** | 0.09 | - | | | | | | |
| | GAR | 0.99*** | 0.99** | 0.03 | 0.03 | 0.07 | 0.08* | 0.61*** | 0.15*** | 0.01 | 0.17 | - | | | | | |
| | NOC_A | 0.99 | 0.99*** | 1.00*** | 0.89 | 0.92*** | 0.84*** | 0.64*** | 0.60*** | 0.55 | 0.35 | 0.67 | - | | | | |
| HpB | SET | 0.98*** | 0.98*** | 0.97*** | 0.97*** | 0.98*** | 0.97*** | 0.96*** | 0.95*** | 0.96*** | 0.94*** | 0.97*** | 0.97*** | - | | | |
| | SPU | 0.98 | 0.98*** | 0.98*** | 0.98*** | 0.98*** | 0.98*** | 0.95*** | 0.95*** | 0.95 | 0.93*** | 0.97*** | 0.98 | 0.12 | - | | |
| | SCR_B | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 0.98 | 0.94 | 0.95 | 0.94 | 0.93 | 0.98 | 1.00 | 0.89 | 0.87 | - | |
| | NOC_B | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 0.98 | 0.94 | 0.95 | 0.94 | 0.93 | 0.98 | 1.00 | 0.92 | 0.91 | 1.00 | - |

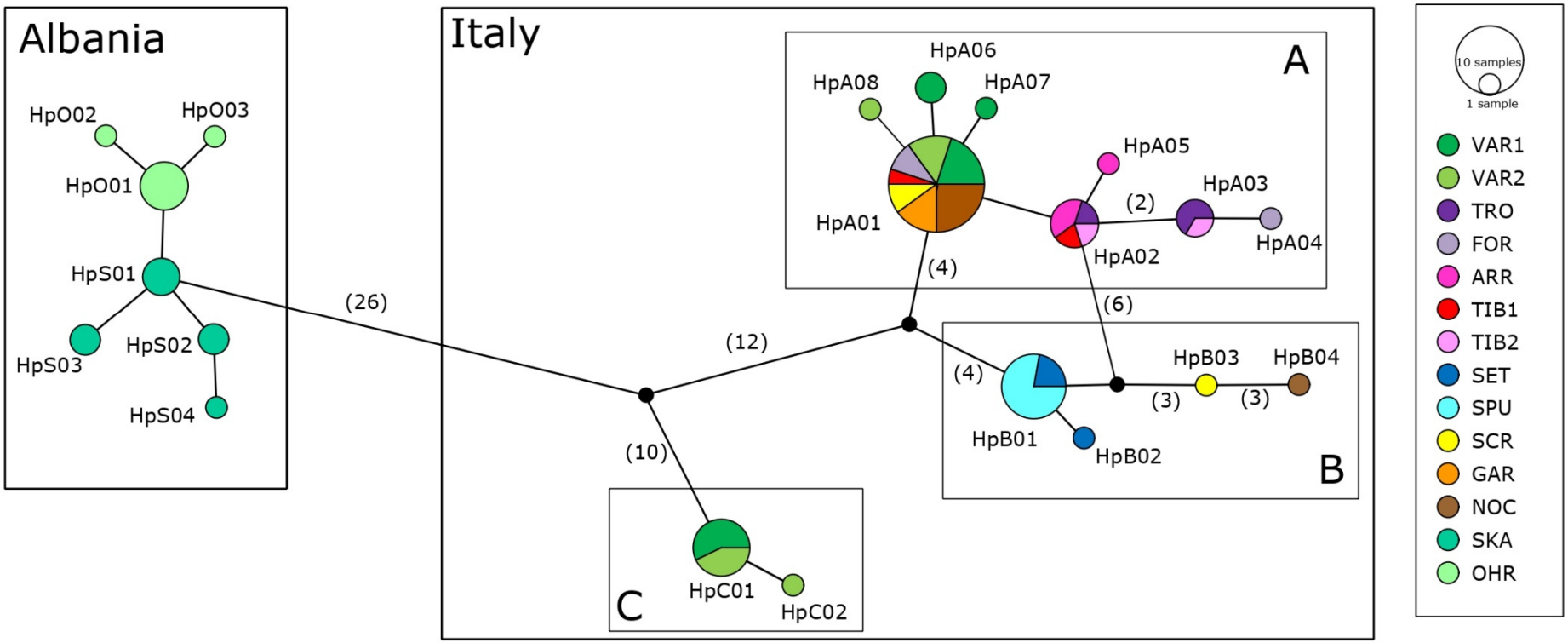



**Figure S1.** Haplotype network based on COI sequences obtained in this study. Each circle corresponds to one haplotype and its dimension is proportional to the haplotype frequency. The number of nucleotide substitutions between haplotypes is indicated in parenthesis. Population abbreviations refer to Table 1. Letters (A, B, C) and boxes refer to *S. rubilio* haplogroups as indicated in Figure 2.

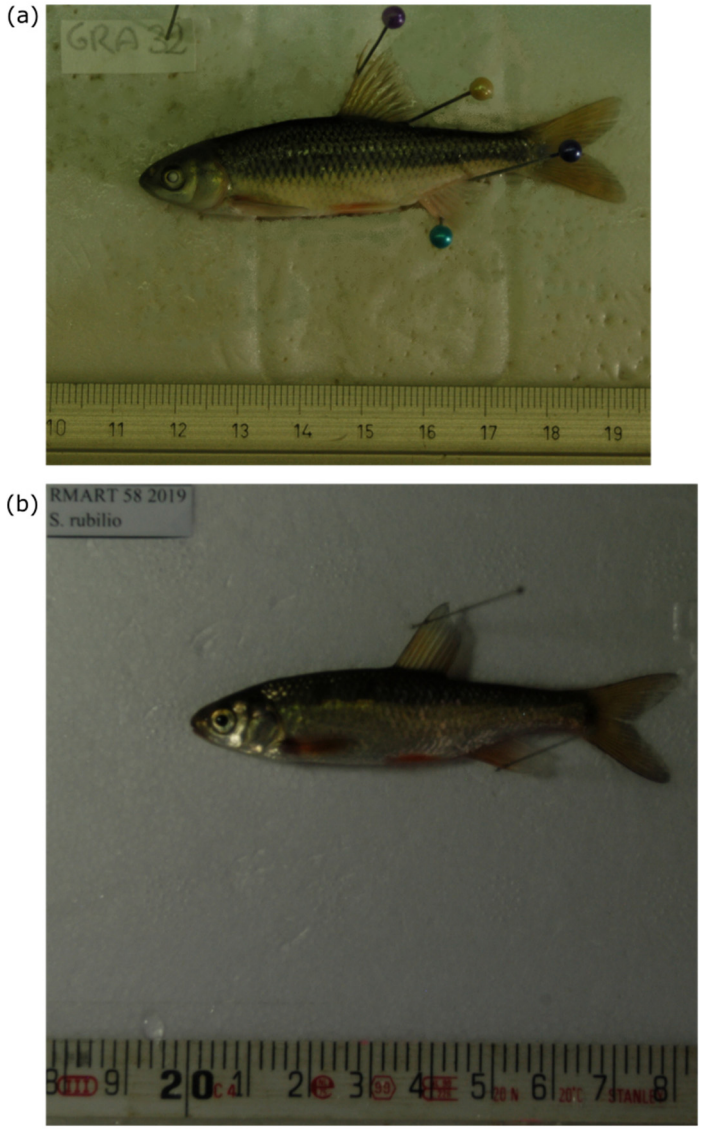



**Figure S2.** Specimens with mtDNA and Cyfun P belonging to lineage C (a) and lineage A (b).

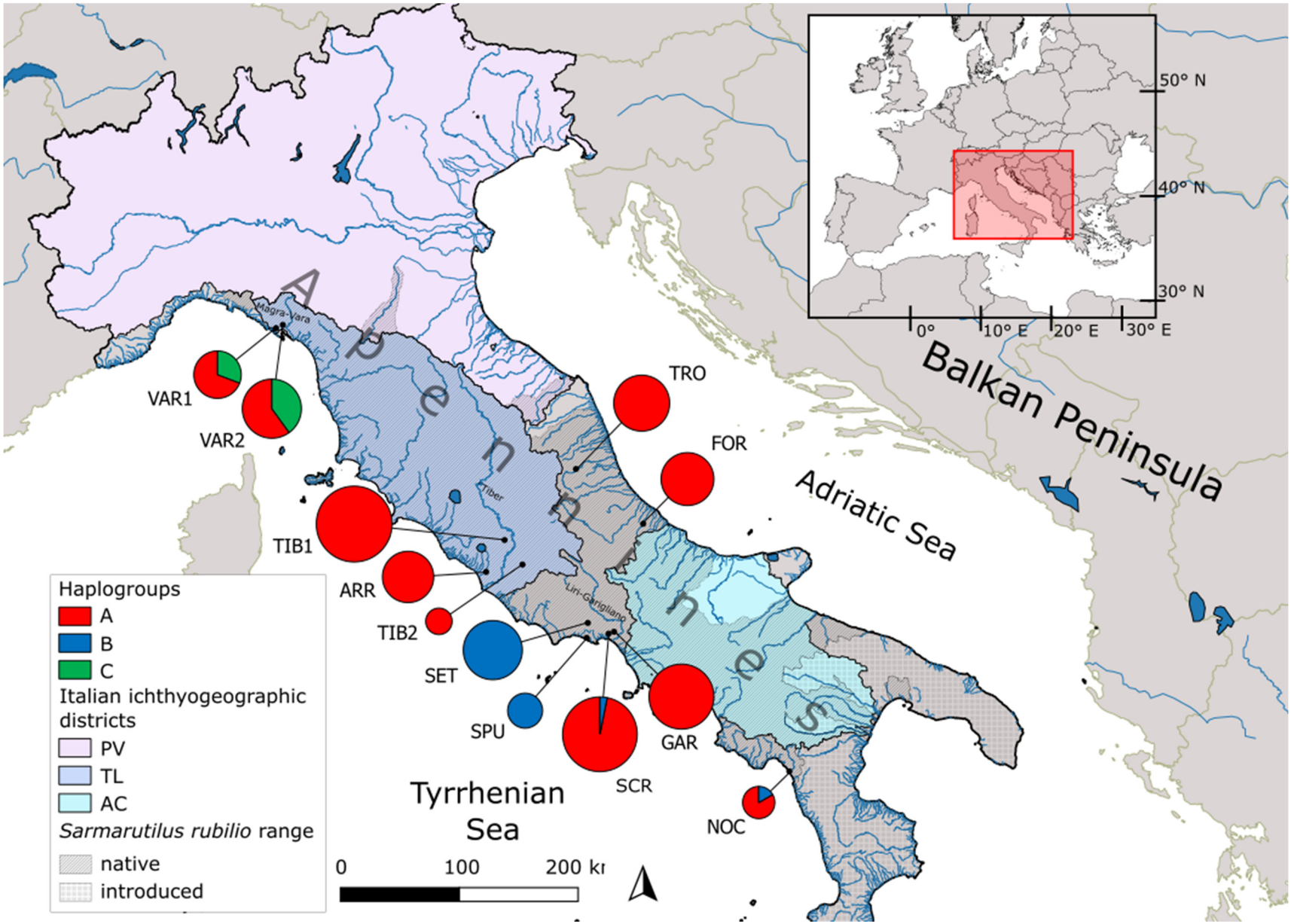



**Figure S3**. Spatial haplogroups distribution in *S. rubilio*, based on 208 Italian CR sequences.

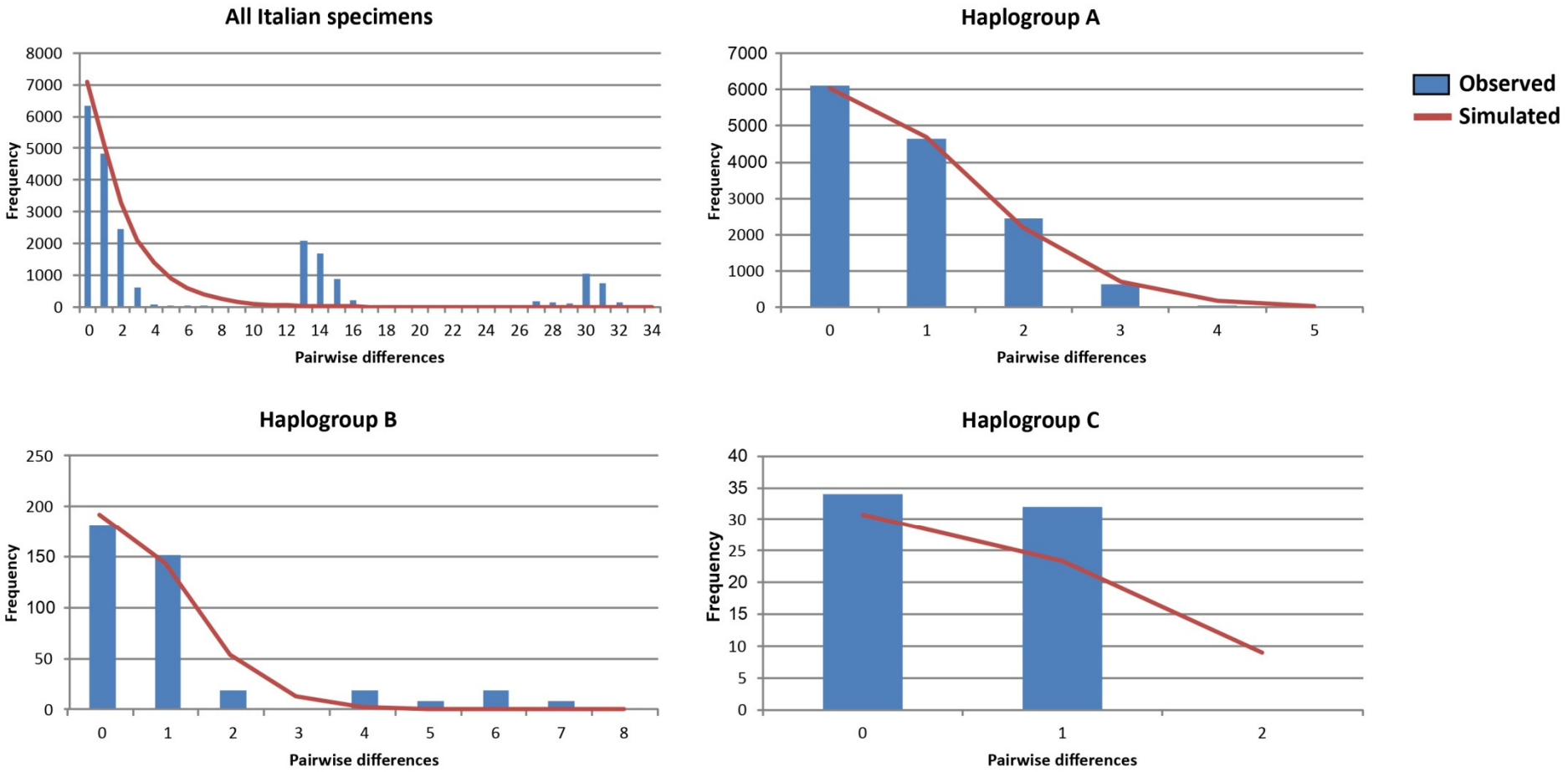



**Figure S4**. Observed mismatch distributions (bars) and expected mismatch distributions under the sudden expansion model (solid line) of CR sequences.

# CHAPTER 2

# Phenotypic plasticity over genetic diversity: ecomorphological patterns revealed in the eurytopic and threatened Italian endemic freshwater fish *Sarmarutilus rubilio* (Bonaparte, 1837)

Submitted to *Biological Journal of the Linnean Society*

# Phenotypic plasticity over genetic diversity: ecomorphological patterns revealed in the eurytopic and threatened Italian endemic freshwater fish *Sarmarutilus rubilio* (Bonaparte, 1837)

GERARDO PETROSINO [1,*], ANNA RITA ROSSI [1], LORENZO TANCIONI [2], FRANCESCO GALLOZZI [1] and PAOLO COLANGELO [3]

[1] Department of Biology and Biotechnology "C. Darwin", Sapienza University of Rome, via Borelli 50, 00161 Rome, Italy

[2] Department of Biology, University of Rome "Tor Vergata", via Cracovia 1, 00133 Rome, Italy

[3]National Research Council, Research Institute on Terrestrial Ecosystems, Via Salaria km 29.300, 00015 Rome, Italy

* Corresponding author: E-mail: gerardo.petrosino@uniroma1.it

**ABSTRACT**

The large number of cryptic species and extensive intraspecific phenotypic plasticity among freshwater fishes pose a challenge to the correct assessment of diversity within these taxa, which is crucial for their conservation. Recently, three different genetic lineages have been identified within the Italian endemic South European roach *Sarmarutilus rubilio*, a threatened species with a broad ecological niche. Using geometric morphometrics we tested the existence of morphological differences associated with genetic common ancestry, i.e. between lineages, and/or environmental parameters. We observed streamlined body shapes in sites scarcely altered by human intervention and with fast water flow, and on the opposite deeper body shapes in canals and one reservoir with slow/still water flow. Our results suggested morphological diversity in the South European roach is mainly the result of phenotypic plasticity in response to different environmental drivers (i.e. lotic vs lentic hydrodynamic patterns) rather than to different genetic backgrounds.

ADDITIONAL KEYWORDS: environmental influence - functional traits - geometric morphometrics - Leuciscidae - mitochondrial lineages - water flow

## INTRODUCTION

Visual identification of freshwater fishes could be a difficult task, as many species show overlapping meristic traits and similar morphological features (Ward, Hanner, & Hebert, 2009; Costa-Silva *et al.*, 2015), and numerous extreme cases of cryptic species (that can be recognized only by molecular approaches) are reported (Palandačić *et al.*, 2017; Arroyave, Martinez, & Stiassny, 2019). On the other hand, several taxa show intraspecific morphological diversity across their geographic range, as a response faced by different populations to changes in environmental features (i.e. phenotypic plasticity), like temperature (Lema *et al.*, 2019), water flow (Kelley *et al.*, 2017), predation (Langerhans & Makowicz, 2009) and foraging (Kristjánsson *et al.*, 2018), sometimes in combination with genetic trait diversity (Collin & Fumagalli, 2011). In the last decade integration of molecular and morphological methods proved to be very useful in solving fish taxonomic issues: the combined approach was successfully used in testing species delimitation between closely related species (Tancioni *et al.*, 2013; Terlecki *et al.*, 2022), providing evidence of new cryptic ones (Shelley *et al.*, 2018; Garcez *et al.*, 2018; Zaccara *et al.*, 2019) and hybridization events (Hayden *et al.*, 2010; Rossi *et al.*, 2016; Konopiński & Amirowicz, 2018), and allowing to distinguish between the genetic and environmental basis of morphological diversity (Berbel-Filho *et al.*, 2016; Ramler *et al.*, 2017; Quadroni *et al.*, 2023)**.**

The South European roach *Sarmarutilus rubilio* (Bonaparte, 1837) is a Leuciscidae fish - previously included in the Cyprinidae, see (Schönhuth *et al.*, 2018)-, endemic to freshwaters of peninsular Italy and introduced in the southern area and Sicily island (Figure 1), and even in Tunisia (Djait *et al.*, 2019) and Turkey (Keskin, Unal, & Atar, 2016); some reports from Albania were recognized as different species (Petrosino *et al.*, 2022). The South European roach is an omnivorous species with a broad ecological niche (Balestrieri *et al.*, 2006) and it could be found in still waters (both natural lakes and artificial dam reservoirs), small rivers, streams and canals, on substrates with rocks, pebbles, sand and gravel, but also muddy and rich in submerged vegetation, mainly in lowland and hilly areas, though some populations were also found at 1250

m above sea level (Bianco *et al.*, 2013; Di Tizio & Di Felice, 2016). No morphological differentiation between sexes is known, except for reddish fins and head tubercles visible in reproductive males (Bianco & Ketmaier, 2014). Despite being a eurytopic species, several local populations disappeared in the last decades, especially in lakes, mainly due to the introduction of allochthonous species. Presently the South European roach is declining (Bianco *et al.*, 2013), and its conservation status was moved from Near Threatened to Vulnerable (Rondinini, Battistoni, & Teofili, 2022). The species is also listed in Annex II of the European Union Habitats Directive, which includes species of community interest whose conservation requires the designation of special areas of conservation (EEC, 1992). Recent investigations about *S. rubilio* genetic diversity (Petrosino *et al.*, 2022) revealed the existence of three mitochondrial lineages, named haplogroup (Hp) A, B and C, likely originated in allopatry during the Pleistocene and subsequently experienced secondary contact. The first lineage is widespread across the species range, the second is localized in South Italy, coexisting with A in some sites and being exclusive in others, and the third was found in a few sites from a single basin at the northern edge of *S. rubilio* native range and always together with A. The genetic divergence observed in the mitochondrial cytochrome oxidase I between C and the other two lineages exceeded the threshold generally accepted to discriminate between fish species (Ward, 2009), thus suggesting it could belong to a new cryptic species; however, no differences in meristic counts supporting the description of a new species were observed between lineage C specimens and the others.

Despite the high genetic and ecological diversity observed between different geographic areas, little is known about morphological variations among *S. rubilio* lineages or populations, which may be a hint for local adaptations (Young, Evans, & Simmons, 2011; COLLIN & FUMAGALLI, 2011) and/or phenotypic plasticity. In the similar species *Rutilus rutilus* - the South European roach was indeed previously classified as *Rutilus rubilio*, (Bianco & Ketmaier, 2014) - which also inhabits an array of different freshwater environments, investigations using the genetic-morphology integrated approach revealed both genetic and phenotypic differentiation between sympatric and allopatric populations, providing evidence for ecological differentiation (Faulks *et al.*, 2015; Rieder *et al.*, 2019). Previous geometric morphometric analyses involving *S. rubilio* didn't find significant morphological differences between nearby populations (Bravi,

Ruffini, & Scalici, 2013) and this was interpreted as the result of gene flow between tributaries in the same basin.

In this study, we analysed *S. rubilio* specimens from thirteen populations sampled in different habitats across the whole species range, to assess the degree of body shape variation and its relationship with the species' genetic diversity, shaped by past geological events, and the current environmental features of the investigated sites. In detail we aimed to test whether differences in body shape can be observed: (1) between the three identified mitochondrial lineages and thus whether specimens differing in their genetic background could be associated with different body shapes and also provide new insights into the status of the putative cryptic species; (2) between different freshwater habitats and thus make inferences about which environmental factors, if any, contribute in shaping fish external morphology and how *S. rutilus* populations respond to the diverse environments. These outcomes will increase our knowledge about the limits of morphological diversity and plasticity in this vulnerable species, and be useful in future to evaluate the effects of anthropogenic habitat alterations and impending climate changes on populations' survival (Berbel-Filho *et al.*, 2016; Andres, Chien, & Knouft, 2019; Potts, Mandrak, & Chapman, 2021).

## MATERIALS AND METHODS

### SAMPLING AND DATA COLLECTION

Overall, 202 South European roaches from 13 sampling sites in lotic (12) and lentic (1) ecosystems were analyzed. We used 146 specimens (from 9 sampling sites) that were already genetically analysed and were representative of the three main mitochondrial genetic lineages identified for *S. rubilio* (Petrosino *et al.*, 2022). In addition, we enriched our sampling set with 56 newly collected specimens, from 4 further sites, whose mitochondrial Control Region (CR) sequences were obtained in this study (Figure 1; Table 1).

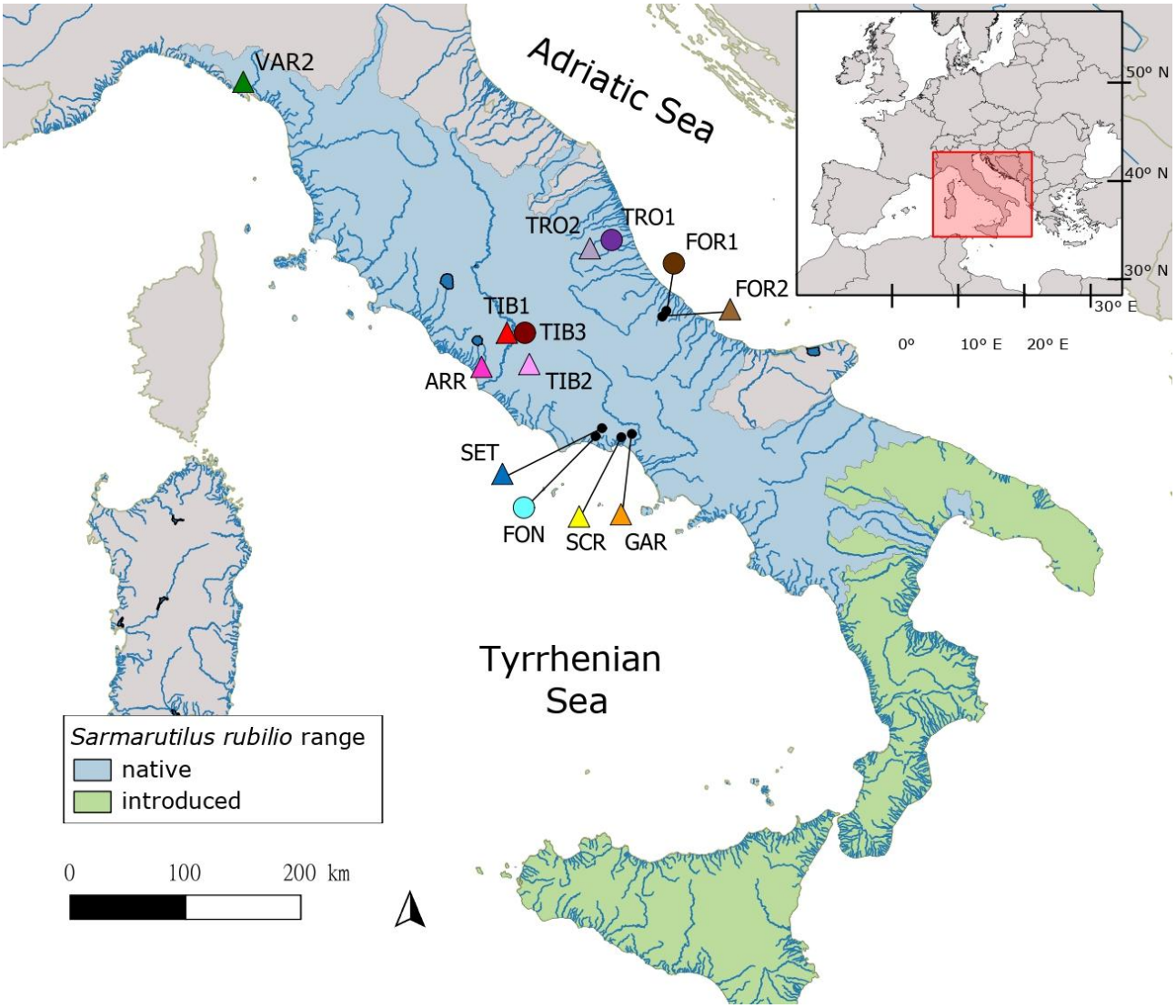



**Figure 1.** Geographic location of the study area (black and red rectangles, top right) and sampling sites: triangle indicates sites investigated by Petrosino *et al.* (2022), circle indicates new sampled sites. For abbreviations refer to Table 1. *S. rubilio* distribution according to IUCN (https://www.iucnredlist.org/species/19786/9014268, accessed on 19 December 2022) is also reported.

**Table 1.** Sampling details for thirteen populations of South European roach. The sampling location, whether lotic (flowing waters) or lentic (still waters) habitat, specimens previously analyzed in Petrosino *et al.* (2022) for the mitochondrial Control Region (CR), geographic coordinates (datum = WGS84) and the sample size (N) are shown for each population. Sampling site IDs were named accordingly to Petrosino et al., 2022, TRO and FOR sites from the previous study were here named TRO2 and FOR2.

| Drainage basin | Sampling site ID | Lotic/lentic waters | Lat (°N) | Lon (°E) | N | CR |
|---|---|---|---|---|---|---|
| Magra-Vara | VAR2 | lotic | 44.191 | 9.790 | 21 | Petrosino et al 2022 |
| Tronto | TRO1 | lotic | 42.857 | 13.701 | 6 | Present study |
| | TRO2 | lotic | 42.802 | 13.465 | 12 | Petrosino et al 2022 |
| Foro | FOR1 | lotic | 42.259 | 14.189 | 7 | Present study |
| | FOR2 | lotic | 42.246 | 14.186 | 9 | Petrosino et al 2022 |
| Arrone | ARR | lotic | 41.914 | 12.265 | 15 | Petrosino et al 2022 |
| Tiber | TIB1 | lotic | 42.173 | 12.545 | 31 | Petrosino et al 2022 |
| | TIB2 | lotic | 41.911 | 12.788 | 6 | Petrosino et al 2022 |
| | TIB3 | lentic | 42.166 | 12.734 | 27 | Present study |
| Fondi | SET | lotic | 41.368 | 13.421 | 20 | Petrosino et al 2022 |
| | FON | lotic | 41.347 | 13.379 | 16 | Present study |
| Santa Croce | SCR | lotic | 41.287 | 13.716 | 8 | Petrosino et al 2022 |
| Liri-Garigliano | GAR | lotic | 41.303 | 13.743 | 24 | Petrosino et al 2022 |

Fish sampling was authorized by Regional Directions responsible for Hunting and Fishing activities and by Directions of Protected areas (Prot. n.: Lazio G10101, 25 July 2019; Abruzzo DPD023/171, 12 April 2021; Marche 213, 13 April 2021; Liguria 5166–2021, 30 August 2021; Aurunci protected area 0002963.U, 01 October 2021). Sampling, morphological identification of specimens and tissue collection for genetic

analysis followed the procedures reported in Petrosino *et al.* (2022). For shape analysis all fishes were photographed in the field in the left lateral view, using a Nikon D100 camera (28–100 mm lens) fixed on a tripod and using a reference scale. Seventeen environmental parameters were recorded at each site using Eutech's multi-parameter CyberScan PCD650 tester, visual estimation and GIS database (for details see Table S1).

DNA EXTRACTION, AMPLIFICATION AND ANALYSIS

DNA was extracted from specimens' fin clips and the mitochondrial control region (CR) was amplified and sequenced following procedures reported in Petrosino *et al.* (2022). Sequences were then compared to those from our previous study (Petrosino *et al.*, 2022) and those available for *S. rubilio* in Genbank using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990), to identify the mitochondrial lineage and the haplotype for each new specimen. We used Arlequin 3.1 (Excoffier, Laval, & Schneider, 2005) to calculate haplotype diversity (Hd) and nucleotide diversity ($\pi$) for both haplogroups and populations and the genetic differentiation between them as pairwise $\Phi_{ST}$ using Kimura 2-parameters distance K2P (Kimura, 1980), testing for significance with 10000 bootstrap replicates and Holm *p*-values correction (Holm, 1979). Finally, we calculated K2P between specimens using Mega 11 (Tamura, Stecher, & Kumar, 2021), to further investigate the correlation between genetics and morphological variation (see below).

MORPHOLOGICAL ANALYSIS

Morphological analysis of individuals' left side was performed by measuring seventeen landmarks (LMs), pointed on digital images using tpsDig2 ver 2.31 (Rohlf, 2017)(Figure 2).

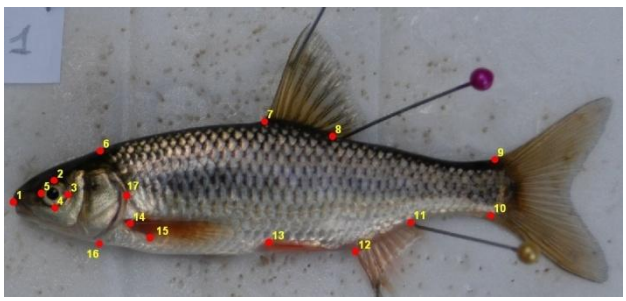



**Figure 2.** Landmarks position for body shape analysis.

Subsequent geometric morphometrics analyses were performed using the packages *geomorph* version 4.0.4 (Adams *et al.*, 2022) and *Morpho* version 2.10 (Schlager, 2017) in R version 4.0.3 (R Core Team, 2020).

A generalized Procrustes superimposition GPA (Rohlf & Slice, 1990) was performed on LMs coordinates to remove non-shape variation and standardize each specimen to unit centroid size (CS). Aligned coordinates were used for subsequent analyses.

We first used a principal component analysis (PCA) on the whole sample to assess the overall morphological variation, highlighting both the site and the mitochondrial lineage for each specimen. Next, to investigate further intergroup differences, we evaluated the level of morphological disparity between sampling sites and between lineages by performing morphological disparity tests, with 10000 iterations and Holm corrections; when lineages (i.e. haplogroups) where considered, variances were calculated using residuals from sampling site mean for each specimen, to remove the effect of sampling location. To check whether differences in sampling size between groups influenced the results of the disparity tests we re-calculated variances using the rarefied_disparity function as implemented in *GeometricMorphometricsMix* package (https://github.com/fruciano/GeometricMorphometricsMix, accessed 10 April 2023), setting the minimum sampling size as 6 (i.e. the minimum number of specimens found among sites) for analysis between sites and 7 (i.e. the minimum number of specimens found among lineages, specifically in C) for analysis between haplogroups.

We performed a Procrustes ANOVA (Goodall, 1991) to investigate the effect on the shape of specimens' mitochondrial lineage, collection site and size (i.e. allometry through CS values), including in our linear model also the interaction among the latter two, setting 10000 iterations for significance testing.

In addition, we explored the correlation between phenotypic and genetic divergence performing a Mantel test (Mantel, 1967) using pairwise Procustes (morphological) and CR Kimura2P (genetic) distances between sites. Indeed, phenotypic divergence may be related to genetic differences due to an overall reduction in gene flow between them, which can be measured using variation at mitochondrial CR as a proxy (Froufe *et al.*, 2003; Verspoor *et al.*, 2010).

To investigate the relationship between morphological and environmental features we focused on comparisons between running water sites, removing SET (the only still water environment) from the

dataset to avoid any bias in the analysis, as its parameters were too different from those measured in the other sites (e.g. higher values of width and depth, no run/riffle/pool structure). We first standardized values for each environmental parameter (scaled to zero mean and unit variance) and subsequently we performed a Partial Least Square (PLS) analysis (Rohlf & Corti, 2000) between shapes of the remaining 182 specimens and the 17 environmental variables collected, with 10000 iterations for significance testing. To assess the contribution of each environmental parameter to the corresponding PLS axis, and thus their putative influence on the observed shape changes, we extrapolated the first singular vector of the environmental parameters block.

## RESULTS

### GENETIC ANALYSIS

We identified 21 Control Region haplotypes (929-930 base pairs) in our set of 202 specimens: 19 were already identified (Petrosino *et al.*, 2022), while 3 were new haplotypes (HpA13-15), all from the TIB3 site and deposited in the GenBank online database (Accession Numbers OQ139458-60) (Figure S1). Sequences from the newly investigated sites (see Figure 1) TRO1, FOR1, and TIB3, all from Central Italy, clustered in haplogroup A, while those from FON in haplogroup B, accordingly to the geographic distribution of mitochondrial lineages previously observed (see Introduction). Haplotype diversity (Hd) ranged between 0.30 (GAR) and 0.75 (SCR) while $\pi$ was generally low in all sites, except for VAR2 ($\pi$ = 1.52) where haplogroup A and C coexists (Table S2a). Molecular indices calculated per haplogroup were similar between HpA and HpB (Hd ≈ 0.5 and $\pi$% ≈ 0.08), while HpC showed only one haplotype (Table S2b). The distance between haplogroup A and B was 0.015, while those between C and A and between C and B were 0.032 and 0.029 respectively. Pairwise $\Phi_{ST}$s were significant and high ($\Phi$ST > 0.5) when sites where haplogroup B was exclusive (SET and FON) were compared to all the others; within haplogroup A, ARR was the most differentiated site (Table S3).

### GEOMETRIC MORPHOMETRICS

Specimens distribution along the first two PCA axes (25.37% and 14.32% of the variance respectively) revealed shapes changes mainly along the PC1, from more elongated bodies, heads and bigger eyes

(negative values) to deeper and shorter ones (positive values), while only a slight difference in caudal peduncle length and body height was observed along PC2 (Figure 3).

When the sampling site partition was examined, Tiber sites (TIB1, TIB2, TIB3) specimens were observed at the negative extreme of the PC1 axis, while GAR and SET were at the positive one. The Procrustes distances confirmed the visual examination with the highest value observed between TIB2 and SET (0.053) and the lowest between TRO2 and FOR2 (0.015). The disparity test revealed Procruste's variance within sites ranged from $0.45x10^{-3}$ (TRO1) to $0.83x10^{-3}$ (FOR2) (Figure S2a), but differences among these values were not significant.

Distribution along the PC1 axis according to mitochondrial lineages showed that almost all specimens belonging to haplogroup C were in the negative half of the plot, while haplogroup B specimens were in the positive one. Procrustes distance between A and B specimens and C and B ones was 0.022 and 0.030 respectively, while between A and C was 0.015 and no shape differences were observed by restricting the analysis to VAR2 site, where they coexist (Figure S3). When the morphological disparity test was performed Procrustes variances correponded to $0.70x10^{-3}$ (HpA), $0.59x10^{-3}$ (HpB) and $0.58x10^{-3}$ (HpC) and differences between haplogroups were not significant (Figure S2b). When variances were re-calculated with the rarefied_disparity function, no difference from previous analysed values was obtained.

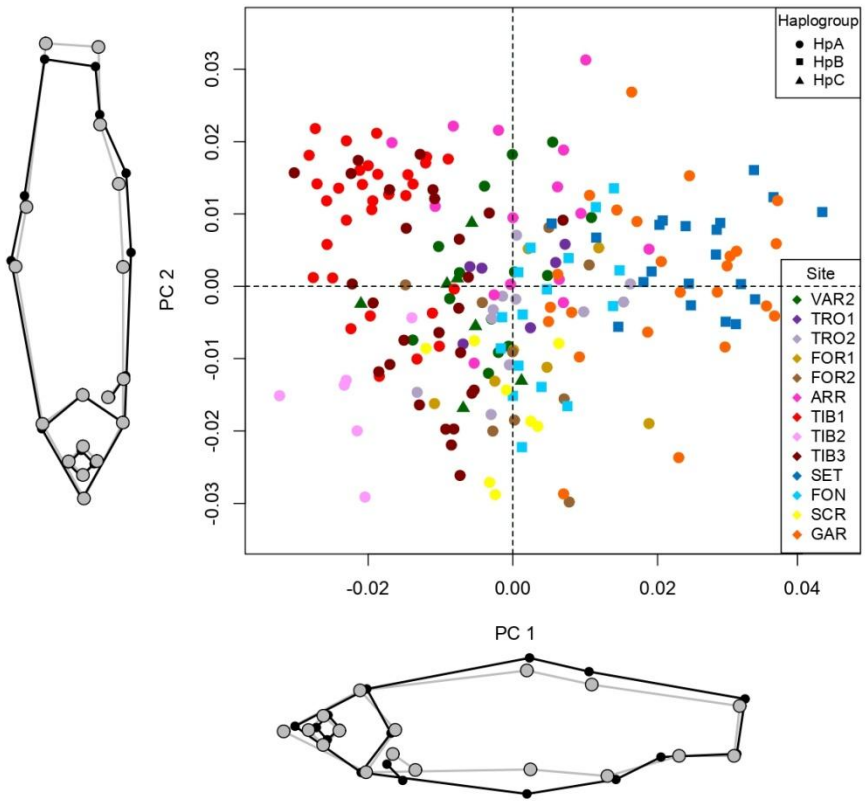



**Figure 3.** Scatterplot for the first two principal components (PC1 and PC2) of principal component analysis (PCA) for body shape. Wireframe graphs indicate the shape changes along each axis, from the most negative value (grey) to the most positive one (black). Colours and abbreviations as in Figure 1.

Procrustes ANOVA investigation revealed that the haplogroup partition explained 7.4% of the shape variance, while 28.5% was influenced by the collection site and only 4.7% by each the fish size and the interaction between size and collection site (Table 2). Collection sites shared mostly the same allometric trajectory, however, for some of them we could observe steeper slopes (Figure 4). In VAR2, where haplogroup A and C coexist, the two corresponding trajectories showed negligible differences. Plotting the minimum and maximum shape scores predicted by the model revealed smaller specimens being more streamlined in shape, and bigger ones more high and oval in profile, with smaller heads and eyes.

When the relationship between shape and genetics was investigated no correlation was found by the Mantel test ($R = 0.10$ , $p > 0.05$ )

**Table 2.** Procrustes ANOVA results, displaying the relative amount of shape variation attributable to mitochondrial lineage (haplogroup), collection site (site), size and interaction among the latter two. SS = sum of squares, MS = mean squares, Rsq = R squared, F = F statistic, *p* = *p*-value.

| **Factors** | **Df** | **SS** | **MS** | **Rsq** | **F** | ***p*** |
|---|---|---|---|---|---|---|
| Haplogroup | 2 | 0.016 | 0.008 | 0.074 | 11.917 | <0.001 |
| Site | 11 | 0.061 | 0.006 | 0.285 | 8.307 | <0.001 |
| Size | 1 | 0.010 | 0.010 | 0.047 | 15.129 | <0.001 |
| Site: size | 12 | 0.010 | <0.001 | 0.047 | 1.251 | 0.038 |
| Residuals | 175 | 0.117 | <0.001 | 0.546 | | |
| Total | 201 | 0.214 | | | | |

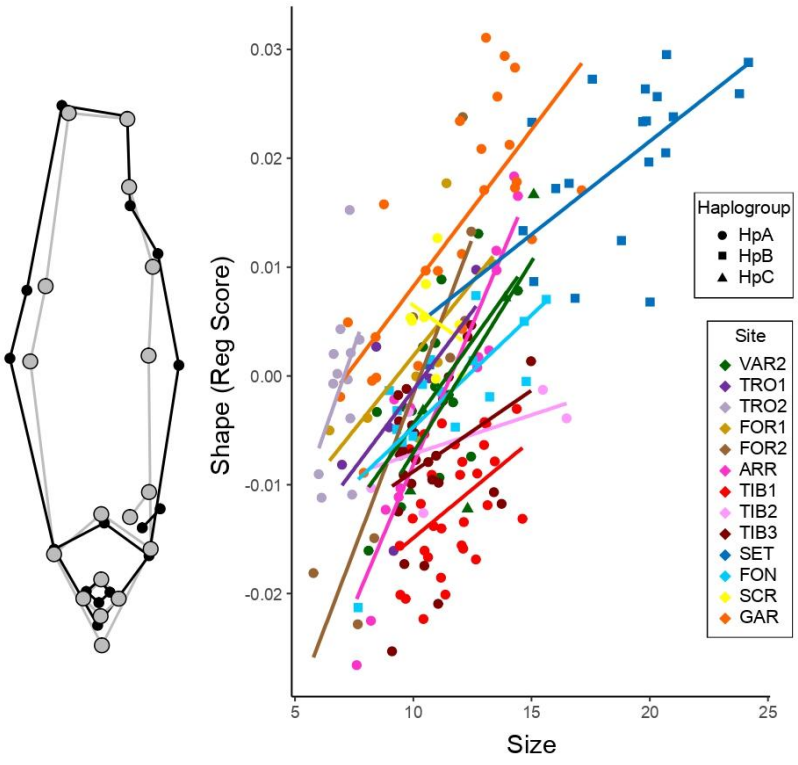



**Figure 4.** Allometry plot: the X-axis represents the centroid size of the body, the Y-axis represents the regression scores of the predicted values in the linear model. Wireframe graph indicates the predicted shape changes along the Y-axis, from the most negative value (grey) to the most positive one (black) and is magnified 1.5 times to appreciate differences. Colours and abbreviations as in Figure 1.

When environmental parameters among running water sites were examined, we observed a low level of correlations between different variables: among 272 comparisons only three provided a correlation higher than the absolute value of 0.7: run *vs* riffle, run *vs* pool and width *vs* elevation. According to this low level of overall correlation among the environmental variables we decided to keep all of them in the analysis. PLS analysis returned significant correlations for almost all combinations of shape and environmental variables vectors; we decided to focus only on the first one as it showed the highest correlation (0.74) and total covariance (64.97%), while the others revealed lower values ($R < 0.56$ and total covariance < 20.49% ). In the PLS plot corresponding to the first vectors pair (Figure 5a) specimens from Tiber basin sites (TIB1, TIB2 and TIB3) and SCR were associated with negative values of the environmental parameters first vector (X-axis), with TIB1 specimens falling in the most extreme ones; conversely, ARR and FON specimens were placed at the positive end; TRO2 and FOR1 overlapped both on the X and the Y-axis, while VAR2 and GRA, despite being associated to the same X values, were well distinguished by shape on the Y-axis. Shapes predicted by the linear correlation moved from slim (negative X-axis values) to oval (positive X-axis values) shapes.

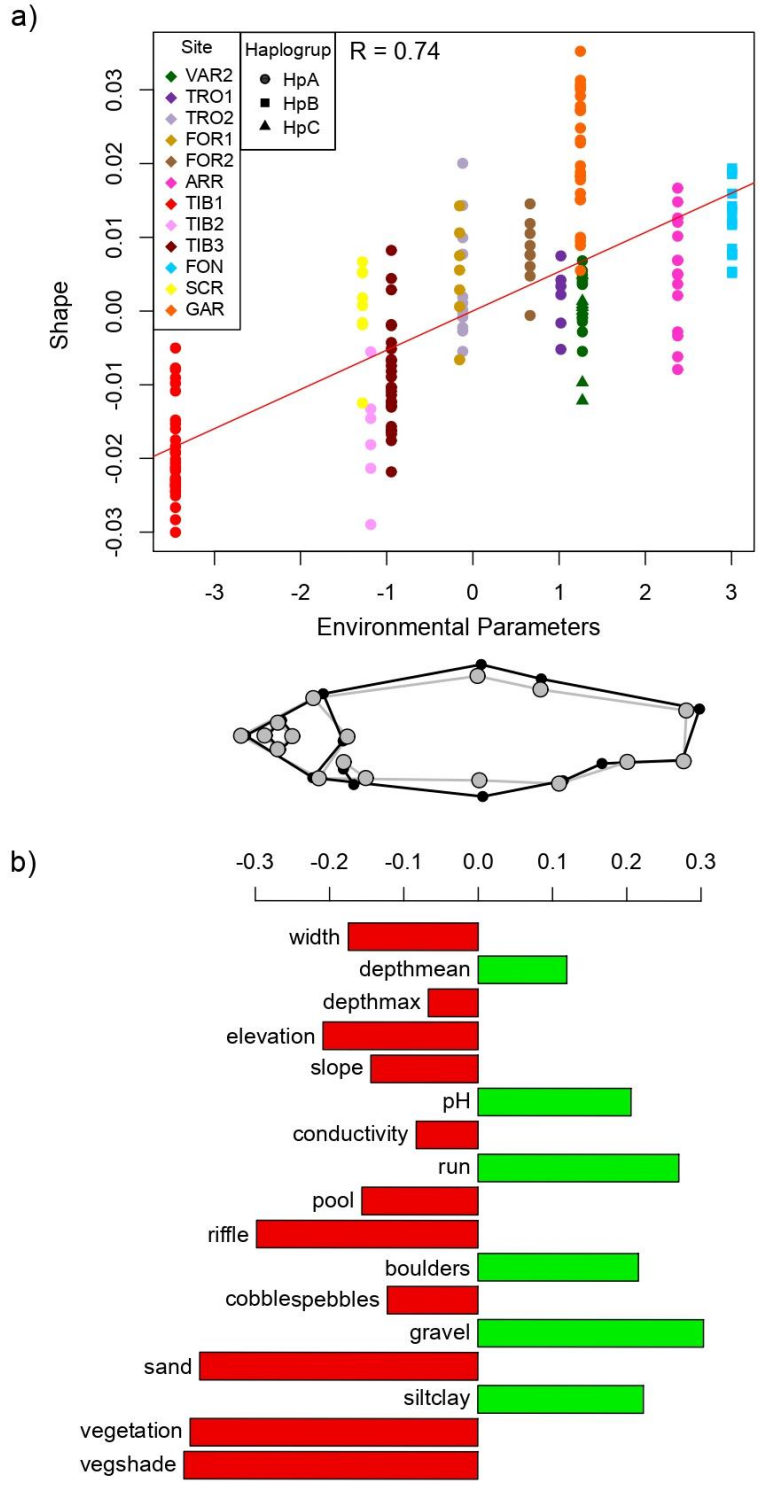



**Figure 5.** a) PLS analysis plot: the X-axis represents the first singular vector of the environmental parameters matrix used for the analysis, while the Y-axis the first singular vector of the landmarks block. Wireframe graph indicates the shape changes predicted by the correlation model (red line), from the most negative value (grey) to the most positive one (black), the graph is magnified 2 times to appreciate changes. Colours and abbreviation as in Figure 1. b) Barplot for the first vector of PLS environmental parameters block. Bars represent the contribution of each environmental parameter to positive (green bars) or negative (red bars) X-axis values in the PLS plot (for parameters explanation see Table S1).

The first singular vector of the environmental parameters block revealed that vegetation both in water and over the banks mostly contributed to negative values observed in the plot, and also sand, percentage of riffles (% of turbulent water flowing) and elevation showed values ≤ -0.2, thus they were associated with streamlined body shapes; on the opposite percentage of gravel and run (% of uniform water flowing) were mostly associated with positive values and thus to deeper fish bodies (Figure 5b).

## DISCUSSION

Intraspecific diversity, i.e. the genomic and phenotypic diversity found within and among populations, is the basis for species adaptation and diversification, thus understanding its magnitude across species and the mechanisms that originate and maintain it are fundamental for the conservation of biodiversity (Caballero & García-Dorado, 2013; Forsman & Wennersten, 2016; Mimura *et al.*, 2017).

In this study genetic analysis of sites not previously investigated for *Sarmarutilus rubilio* provided new mitochondrial haplotypes and confirmed the geographic pattern of diversity previously reported (see Introduction): lineage A is the predominant one across the entire species range, it includes many haplotypes and differing each other only by few mutations; lineage B is fixed in a restricted geographic area between Central and South Italy; lineage C is the most divergent and is present exclusively at the northern border of the species distribution, where it coexists with lineage A. At the same time, our investigations revealed that South European roaches showed significantly diverse body shapes between sampling sites, which differed in environmental characteristics.

The observed morphological diversity could therefore be a consequence of local adaptations fixed through natural selection and thus having a genetic basis (Hendry, Taylor, & McPhail, 2002; Jeukens & Bernatchez, 2012), or the result of phenotypic plasticity in response to different ecological factors (Scharnweber *et al.*, 2013; Franssen, Stewart, & Schaefer, 2013b), or also a combination of both (Young *et al.*, 2011; Franssen, 2011). Our genetic investigations focused on the mitochondrial Control Region, a non-coding marker and thus not directly responsible for phenotypic variations; anyway, the genetic divergence observed in the CR between populations may reflect differences in other genetic regions (e.g. genes and/or regulatory regions) more involved in morpho-functional adaptations. The Procrustes ANOVA found a low but statistically

supported relationship between genetic partition and morphological differences, but a deeper investigation using pairwise Procustes and Kimura2P distances between sites showed no correlation. Morphological variation in haplogroup B was restricted in the positive half of the PCA axis1, i.e. it showed deeper mean body shape than haplogroup A; we currently cannot discriminate whether this is a fixed characteristic of B lineage or exclusive of the investigated restricted area: indeed all specimens showing lineage B were collected from FON and SET sites, which are only 5 km far from each other, along the same coastal plain. Finally we revealed the absence of morphological differences between specimens belonging to lineage C (a putative new cryptic species) and A in VAR2 site where they coexist. This suggests that the strong effect of phenotypic plasticity (see below) may hide genetic-based morphological differences between the two lineages when they are under the same environmental pressures (Franssen *et al.*, 2013a), also considering that putative lineage-fixed characteristics may be the result of allopatric genetic drift and not necessarily of adaptation through natural selection; another explanation is that there may be no difference in body shape between them, as also observed in other freshwater fish species complex (Arroyave, Martinez, & Stiassny, 2019).

This interpretation is supported by the Procrustes ANOVA, which indicated the collection site partition as a more relevant factor than the haplogroup subdivision in explaining the observed body shape differences within *S. rubilio*. Indeed most of the morphological differentiation was observed within haplogroup A, between populations from diverse environments, suggesting that phenotypic plasticity is predominant over lineage characteristics.

Ordination methods placed specimens along a mostly continuous gradient according to sites, with body shape changing from streamlined to oval. The same changing pattern, but with a smaller amount of variation, was observed also within populations and it was the result of allometric growth (i.e. specimens assume more oval body shapes when growing in size) as previously reported for this species (Bravi *et al.*, 2013). Coupling environmental and specimens morphological differences among running water sites revealed some interesting outcomes. Most streamlined fishes were found in sites where the water flow is heterogeneous, due to the presence of pools and riffles; some of these sites are located in hilly environments scarcely altered by human activities, thus explaining why they were also associated with

relatively high elevation, slope and vegetation both in water and on banks. Conversely, fishes with the deepest bodies were found in watercourses at low elevations, which are characterized by slow and homogeneous water flow, and uniform cross-sections, that were highly altered by reclamation or flood protection works. Differences in *S. rubilio* morphology are thus consistent with the "shape-water velocity" patterns observed in other fish species (Jacquemin, Martin, & Pyron, 2013; Haas, Heins, & Blum, 2015). Indeed, streamlined body shapes are suited for swimming in high water flow, as they minimize drag; conversely, deep body shapes are fitted for complicated locomotor patterns where water flow is slower (Langerhans & Reznick, 2010). South European roaches with the deepest bodies were observed in SET site, which is a lentic environment, further suggesting that hydrodynamic pattern and in particular water velocity may be the main driver in determining morphological responses.

Different body shapes in *S. rubilio* were also associated with percentages of sites' streambed sediments. We were not able to identify a clear explanation for this relationship if any exists: indeed the overall granulometry composition of investigated watercourses may not best represent the mesohabitat preferentially used by the South European roaches, which tends to mostly concentrate where the water flow is relatively slower within each stretch that this species inhabits. However substrate-related phenotypic diversification was found also within the leuciscid *Rutilus rutilus* and was suggested to be the result of different feeding strategies (Rieder *et al.*, 2019). Food availability and accessibility were proved to strongly influence the diet of *S. rubilio* between different environments, and this species can turn to suboptimal resources (i.e. vegetation instead of invertebrates) when in sympatry with similar ones and competition for food arises (Balestrieri *et al.*, 2006); noteworthy in SET site, where biggest specimens with the deepest bodies were found, no other fish species that could potentially compete for food was found. Deepening our knowledge about *S. rubilio* mesohabitat use, diet and habits in different environments could thus provide further clues about its morphological response and adaptability.

## CONCLUSIONS

Our study revealed ecomorphological variation between populations of the South European roach and suggested the hydro-morphological features and hydrodynamic pattern may be the main drivers of the observed intraspecific shape diversity through phenotypic plasticity. Indeed, specimens showed alternatively streamlined or oval average body shapes likely to optimize swimming in different freshwater environments. We do not exclude that genetics may also play a role in shaping morphological diversity in the South European roach; however, our results suggested that its contribution could be negligible, and that phenotypic similarity/dissimilarity pattern linked to common genetic ancestry is obscured by the environmental effect, i.e. by the expression of phenotypes locally adapted.

Further research is needed to clarify the role of genetic and environmental parameters in shaping *S. rubilio* morphological differences: we suggest the use of markers more suited to detect a signature of selection (He *et al.*, 2013), the implementation of common garden experiments, i.e. analyzing shape changes in specimens raised under controlled experimental conditions, from early life larval to adult stage (Langerhans *et al.*, 2004; Franssen, 2011; Dunn, O'Brien, & Closs, 2020), and also investigations concerning other phenotypic traits, like skin colour variations, that could be related to both changes in environmental features (Quadroni *et al.*, 2023) and genetic background (Valette *et al.*, 2023).

Increasing our knowledge about *S. rubilio* intraspecific morphological variation will be useful to predict the effects of habitat alterations (e.g. water abstraction and climatic changes) on lineages' and populations' survival and thus for their sustainable management and conservation.

## ACKNOWLEDGMENTS

We thank L. Talarico, M. Graziani and L. Ciuffardi for their help in field activities. We also thank the staff at Aurunci Mountains Regional Park for granting sampling permissions and supervising field activities in the protected area, and the Italian Directions responsible for Hunting and Fishing activities in Liguria, Marche, Lazio and Abruzzo Regions. This work was financially supported by Sapienza University of Rome (Università 2020, project n. 2146271 to A. R. Rossi).

## CONFLICT OF INTEREST

The authors have no conflict of interest to be declared. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## DATA AVAILABILITY

The mitochondrial control-region sequences obtained in our study were deposited in GenBank database (Accession Numbers OQ139458-60). Other data used in the analyses may be obtained from the corresponding author upon request.

## REFERENCES

**Adams DC, Collyer ML, Kaliontzopoulou A & Baken EK**. **2022**. Geomorph: Software for geometric morphometric analyses. R package version 4.0.4.

**Adkison MD**. **1995**. Population differentiation in Pacific salmons: local adaptation genetic drift, or the environment? *Canadian Journal of Fisheries and Aquatic Sciences* **52**: 2762–2777.

**Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ**. **1990**. Basic Local Alignment Search Tool. *Journal of molecular biology* **215**: 403–410.

**Andres KJ, Chien H & Knouft JH**. **2019**. Hydrology induces intraspecific variation in freshwater fish morphology under contemporary and future climate scenarios. *Science of The Total Environment* **671**: 421–430.

**Arroyave J, Martinez CM & Stiassny MLJ**. **2019**. DNA barcoding uncovers extensive cryptic diversity in the African long-fin tetra *Bryconalestes longipinnis* (Alestidae: Characiformes). *Journal of Fish Biology* **95**: 379–392.

**Balestrieri A, Prigioni C, Remonti L, Sgrosso S & Priore G**. **2006**. Feeding ecology of Leuciscus cephalus and *Rutilus rubilio* in southern Italy. *Italian Journal of Zoology* **73**: 129–135.

**Berbel-Filho WM, Martinez PA, Ramos TPA, Torres RA & Lima SMQ**. **2016**. Inter- and intra-basin phenotypic variation in two riverine cichlids from northeastern Brazil: potential eco-evolutionary damages

of São Francisco interbasin water transfer. *Hydrobiologia* **766**: 43–56.

**Bianco PG, Caputo V, Ferrito V, Lorenzoni M, Nonnis Marzano F, Stefani F, Sabatini A & Tancioni L. 2013**. *Rutilus rubilio*. In: Rondinini C, Battistoni A, Peronace V, Teofili C, eds. *Lista Rossa IUCN dei Vertebrati Italiani.* . Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare. http://www.iucn.it/scheda.php?id=-251974278.

**Bianco PG & Ketmaier V**. **2014**. A revision of the *Rutilus complex* from Mediterranean Europe with description of a new genus, *Sarmarutilus*, and a new species, *Rutilus stoumboudae* (Teleostei: Cyprinidae). *Zootaxa* **3841**: 379–402.

**Bravi R, Ruffini M & Scalici M**. **2013**. Morphological variation in riverine cyprinids: a geometric morphometric contribution. *Italian Journal of Zoology* **80**: 536–546.

**Caballero A & García-Dorado A**. **2013**. Allelic Diversity and Its Implications for the Rate of Adaptation. *Genetics* **195**: 1373–1384.

**Collin H & Fumagalli L**. **2011**. Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (*Phoxinus phoxinus*, Cyprinidae). *Molecular Ecology* **20**: 4490–4502.

**Costa-Silva GJ, Rodriguez MS, Roxo FF, Foresti F & Oliveira C**. **2015**. Using Different Methods to Access the Difficult Task of Delimiting Species in a Complex Neotropical Hyperdiverse Group. *PLOS ONE* **10**: e0135075.

**Djait H, Bahri-Sfar L, Laouar H, Missaoui N & Ben Hassine OK**. **2019**. Dietary comparison of pike-perch, *Sander lucioperca* (Linnaeus, 1758) and catfish, *Silurus glanis* Linnaeus, 1758 in Sidi Salem dam reservoir (Tunisia). *Cybium* **43**: 61–69.

**Di Tizio L & Di Felice PL**. **2016**. Rutilus rubilio (Bonaparte, 1837) (Rovella). In: Stoch F, Genovesi P, eds. *Manuali per il monitoraggio di specie e habitat di interesse comunitario (Direttiva 92/43/CEE) in Italia: specie animali*. ISPRA, 164–165.

**Dunn NR, O'Brien LK & Closs GP**. **2020**. Phenotypically Induced Intraspecific Variation in the Morphological Development of Wetland and Stream *Galaxias gollumoides* McDowall and Chadderton. *Diversity* **12**: 220.

**EEC**. **1992**. Council Directive 92/43/ECC. *Official Journal of the European Union* **94**: 40–52.

**Excoffier L, Laval G & Schneider S**. **2005**. Arlequin (version 3.0): An integrated software package for

population genetics data analysis. *Evolutionary Bioinformatics* **1**: 117693430500100.

**Faulks L, Svanbäck R, Eklöv P & Östman Ö**. **2015**. Genetic and morphological divergence along the littoral-pelagic axis in two common and sympatric fishes: Perch, *Perca fluviatilis* (Percidae) and roach, *Rutilus rutilus* (Cyprinidae). *Biological Journal of the Linnean Society* **114**: 929–940.

**Forsman A & Wennersten L**. **2016**. Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography* **39**: 630–648.

**Franssen NR**. **2011**. Anthropogenic habitat alteration induces rapid morphological divergence in a native stream fish. *Evolutionary Applications* **4**: 791–804.

**Franssen NR, Harris J, Clark SR, Schaefer JF & Stewart LK**. **2013a**. Shared and unique morphological responses of stream fishes to anthropogenic habitat alteration. *Proceedings of the Royal Society B: Biological Sciences* **280**: 20122715.

**Franssen NR, Stewart LK & Schaefer JF**. **2013b**. Morphological divergence and flow-induced phenotypic plasticity in a native fish from anthropogenically altered stream habitats. *Ecology and Evolution* **3**: 4648–4657.

**Froufe E, Knizhin I, Koskinen MT, Primmer CR & Weiss S. 2003.** Identification of reproductively isolated lineages of Amur grayling (*Thymallus grubii* Dybowski 1869): concordance between phenotypic and genetic variation. *Molecular Ecology* **12**: 2345-2355.

**Garcez DK, Barbosa C, Loureiro M, Volcan M V., Loebmann D, Quintela FM & Robe LJ**. **2018**. Phylogeography of the critically endangered neotropical annual fish, *Austrolebias wolterstorffi* (Cyprinodontiformes: Aplocheilidae): genetic and morphometric evidence of a new species complex. *Environmental Biology of Fishes* **101**: 1503–1515.

**Goodall C**. **1991**. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**: 285–321.

**Haas TC, Heins DC & Blum MJ**. **2015**. Predictors of body shape among populations of a stream fish (*Cyprinella venusta*, Cypriniformes: Cyprinidae). *Biological Journal of the Linnean Society* **115**: 842–858.

**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**: 169.

**He Y, Li R, Wang J, Blanchet S & Lek S**. **2013**. Morphological Variation Among Wild Populations of Chinese Rare Minnow (*Gobiocypris rarus*): Deciphering the Role of Evolutionary Processes. *Zoological Science* **30**: 475–483.

**Hendry AP, Taylor EB & McPhail JD**. **2002**. Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution* **56**: 1199–1216.

**Holm S**. **1979**. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**: 65–70.

**Jacquemin SJ, Martin E & Pyron M**. **2013**. Morphology of bluntnose minnow *Pimephales notatus* (Cyprinidae) covaries with habitat in a central Indiana watershed. *The American Midland Naturalist* **169**: 137–146.

**Jeukens J & Bernatchez L**. **2012**. Regulatory versus coding signatures of natural selection in a candidate gene involved in the adaptive divergence of whitefish species pairs ( Coregonus spp.). *Ecology and Evolution* **2**: 258–271.

**Kelley JL, Davies PM, Collin SP & Grierson PF**. **2017**. Morphological plasticity in a native freshwater fish from semiarid Australia in response to variable water flows. *Ecology and Evolution* **7**: 6595–6605.

**Keskin E, Unal EM & Atar HH**. **2016**. Detection of rare and invasive freshwater fish species using eDNA pyrosequencing: Lake Iznik ichthyofauna revised. *Biochemical Systematics and Ecology* **67**: 29–36.

**Kimura M**. **1980**. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–120.

**Konopiński MK & Amirowicz A**. **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.

**Kristjánsson BK, Leblanc CAL, Skúlason S, Snorrason SS & Noakes DLG**. **2018**. Phenotypic plasticity in the morphology of small benthic Icelandic Arctic charr (*Salvelinus alpinus*). *Ecology of Freshwater Fish* **27**: 636–645.

**Langerhans RB, Layman CA, Shokrollahi AM & DeWitt TJ**. **2004**. Predator-driven phenotypic diversification

in *Gambusia affinis*. *Evolution* **58**: 2305–2318.

**Langerhans RB & Makowicz AM**. **2009**. Shared and unique features of morphological differentiation between predator regimes in *Gambusia caymanensis*. *Journal of Evolutionary Biology* **22**: 2231–2242.

**Langerhans RB & Reznick DN**. **2010**. Ecology and evolution of swimming performance in fishes: predicting evolution with biomechanics. In: Domenici P, Kapoor BG, eds. *Fish locomotion: an eco-ethological perspective*. Science Publishers Enfield, NH, 248.

**Lema SC, Bock SL, Malley MM & Elkins EA**. **2019**. Warming waters beget smaller fish: evidence for reduced size and altered morphology in a desert fish following anthropogenic temperature change. *Biology Letters* **15**: 20190518.

**Mantel N**. **1967**. The detection of disease clustering and a generalized regression approach. *Cancer research* **27**: 209–20.

**Mimura M, Yahara T, Faith DP, Vázquez-Domínguez E, Colautti RI, Araki H, Javadi F, Núñez-Farfán J, Mori AS, Zhou S, Hollingsworth PM, Neaves LE, Fukano Y, Smith GF, Sato YI, Tachida H & Hendry AP**. **2017**. Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications* **10**: 121–139.

**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.

**Petrosino G, Tancioni L, Turani M, Rakaj A, Ciuffardi L & Rossi AR**. **2022**. Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny. *Genes* **13**: 1071.

**Potts LB, Mandrak NE & Chapman LJ**. **2021**. Coping with climate change: Phenotypic plasticity in an imperilled freshwater fish in response to elevated water temperature. *Aquatic Conservation: Marine and Freshwater Ecosystems* **31**: 2726–2736.

**Quadroni S, De Santis V, Carosi A, Vanetti I, Zaccara S & Lorenzoni M**. **2023**. Past and Present Environmental Factors Differentially Influence Genetic and Morphological Traits of Italian Barbels (Pisces: Cyprinidae). *Water* **15**: 325.

**R Core Team**. **2020**. R: A language and environment for statistical computing.

**Ramler D, Palandačić A, Delmastro GB, Wanzenböck J & Ahnelt H**. **2017**. Morphological divergence of lake and stream *Phoxinus* of Northern Italy and the Danube basin based on geometric morphometric analysis. *Ecology and Evolution* **7**: 572–584.

**Rieder JM, Vonlanthen P, Seehausen O & Lucek K**. **2019**. Allopatric and sympatric diversification within roach (*Rutilus rutilus*) of large pre-alpine lakes. *Journal of Evolutionary Biology* **32**: 1174–1185.

**Rohlf FJ**. **2017**. Program TpsDig, version 2.31. *Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY*.

**Rohlf FJ & Corti M**. **2000**. Use of two-block partial least-squares to study covariation in shape. *Systematic biology* **49**: 740–753.

**Rohlf FJ & Slice D**. **1990**. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic biology* **39**: 40–59.

**Rondinini C, Battistoni A & Teofili C (Eds.)**. **2022**. *Lista rossa IUCN dei vertebrati italiani 2022*. Roma: Comitato Italiano IUCN e Ministero dell'Ambiente e della Sicurezza Energetica.

**Rossi AR, Milana V, Pulcini D, Cataudella S, Martinoli M & Tancioni L**. **2016**. An integrated genetic and morphological approach to clarify the conservation status of the threatened Italian endemic species *Alburnus albidus* (Cypriniformes: Cyprinidae). *Hydrobiologia* **770**: 73–87.

**Scharnweber K, Watanabe K, Syväranta J, Wanke T, Monaghan MT & Mehner T**. **2013**. Effects of predation pressure and resource use on morphological divergence in omnivorous prey fish. *BMC Evolutionary Biology* **13**: 132.

**Schlager S**. **2017**. Morpho and Rvcg--shape analysis in R: R-packages for geometric morphometrics, shape analysis and surface manipulations. In: *Statistical shape and deformation analysis*. Elsevier, 217–256.

**Schönhuth S, Vukić J, Šanda R, Yang L & Mayden RL**. **2018**. Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). *Molecular Phylogenetics and Evolution* **127**: 781–799.

**Shelley JJ, Swearer SE, Adams M, Dempster T, Le Feuvre MC, Hammer MP & Unmack PJ**. **2018**. Cryptic biodiversity in the freshwater fishes of the Kimberley endemism hotspot, northwestern Australia.

*Molecular Phylogenetics and Evolution* **127**: 843–858.

**Tamura K, Stecher G & Kumar S**. **2021**. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**: 3022–3027.

**Tancioni L, Russo T, Cataudella S, Milana V, Hett AK, Corsi E & Rossi AR**. **2013**. Testing Species Delimitations in Four Italian Sympatric Leuciscine Fishes in the Tiber River: A Combined Morphological and Molecular Approach (N Salamin, Ed.). *PLoS ONE* **8**: e60392.

**Terlecki G, dos Santos Rodrigues L, Kikuchi E, Abbatepaulo FV, Bosenbecker C, de Araújo Freire M, da Silva Cortinhas MC, Proietti MC & Cardoso LG**. **2022**. Uncovering *Lampris* species (Actinopterygii, Lampridae) in the southwestern Atlantic Ocean: a molecular and morphometric approach. *Hydrobiologia* **849**: 2745–2759.

**Valette T, Leitwein M, Lascaux J, Desmarais E, Berrebi P & Guinand B**. **2023**. Redundancy analysis, genome-wide association studies and the pigmentation of brown trout (*Salmo trutta* L.). *Journal of Fish Biology* **102**: 96–118.

**Verspoor E, Knox D, Greer R. & Hammar J. 2010.** Mitochondrial DNA variation in Arctic charr (*Salvelinus alpinus* (L.)) morphs from Loch Rannoch, Scotland: evidence for allopatric and peripatric divergence. *Hydrobiologia* **650**: 117–131.

**Ward RD**. **2009**. DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources* **9**: 1077–1085.

**Ward RD, Hanner R & Hebert PDN**. **2009**. The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology* **74**: 329–356.

**Young MJ, Evans JP & Simmons LW**. **2011**. Population genetic structure and a possible role for selection in driving phenotypic divergence in a rainbowfish (Melanotaeniidae). *Biological Journal of the Linnean Society* **102**: 144–160.

**Zaccara S, Quadroni S, De Santis V, Vanetti I, Carosi A, Britton R & Lorenzoni M**. **2019**. Genetic and morphological analyses reveal a complex biogeographic pattern in the endemic barbel populations of the southern Italian peninsula. *Ecology and Evolution* **9**: 10185–10197.

## SUPPLEMENTARY MATERIALS

**Table S1.** Seventeen environmental variables recorded for each sample site. Source data are reported in the first row in grey, and details for variables measurements are reported in rows below the table.

| | | Field measurements | | | GIS data* | | Field measure-ments with multiparameter tester** | | Visual estimation on the field | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Chemical-Physical variables | | Riffle-Pool-Run Sequence | | | Riverbed Granulometry | | | | | Vegetation Cover | |
| **Site ID** | **Flowing/still waters** | **River width (m)** | **Mean depth (m)** | **Max depth (m)** | **Elevation (m)** | **Slope (%)** | **pH** | **Water con-ductivity (µS/cm)** | **Run (%)** | **Pool (%)** | **Riffle (%)** | **Boulders (%)** | **Cobbles and pebbles (%)** | **Gravel (%)** | **Sand (%)** | **Silt and clay (%)** | **Vegetation in water (%)** | **Riparian vegetation shade (%)** |
| VAR2 | flowing | 4.0 | 0.3 | 1.3 | 68 | 2.0 | 7.9 | 560 | 70 | 20 | 10 | 10 | 20 | 60 | 10 | 0 | 15 | 20 |
| TRO1 | flowing | 1.5 | 0.3 | 1.5 | 65 | 3.3 | 8.2 | 590 | 50 | 40 | 10 | 0 | 10 | 80 | 10 | 0 | 0 | 60 |
| TRO2 | flowing | 6.5 | 0.8 | 1.3 | 262 | 8.0 | 7.9 | 610 | 70 | 20 | 10 | 70 | 10 | 20 | 0 | 0 | 40 | 20 |
| FOR1 | flowing | 3.0 | 0.3 | 2 | 193 | 1.0 | 7.9 | 410 | 40 | 40 | 20 | 30 | 40 | 20 | 10 | 0 | 0 | 40 |
| FOR2 | flowing | 4.0 | 0.3 | 0.6 | 227 | 3.3 | 8.7 | 459 | 50 | 10 | 40 | 10 | 30 | 50 | 10 | 0 | 0 | 30 |
| ARR | flowing | 1.0 | 0.2 | 0.2 | 20 | 1.1 | 8.1 | 774 | 90 | 10 | 0 | 0 | 0 | 40 | 40 | 20 | 10 | 10 |
| TIB1 | flowing | 3.0 | 0.3 | 0.8 | 107 | 1.7 | 7.5 | 587 | 35 | 30 | 35 | 0 | 0 | 10 | 90 | 0 | 40 | 80 |
| TIB2 | flowing | 3.0 | 0.35 | 1.8 | 102 | 7.3 | 7.8 | 499 | 40 | 30 | 25 | 0 | 20 | 35 | 35 | 0 | 25 | 25 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TIB3 | flowing | 4.0 | 0.4 | 1.0 | 178 | 3.7 | 7.8 | 533 | 60 | 20 | 20 | 0 | 80 | 10 | 10 | 0 | 20 | 35 |
| SET | still | 50.0 | 2.5 | 6.0 | 11 | 0.1 | 7.9 | 780 | - | - | - | 0 | 0 | 0 | 90 | 10 | 90 | 20 |
| FON | flowing | 2.0 | 0.9 | 1.2 | 3 | 1.0 | 7.8 | 390 | 100 | 0 | 0 | 0 | 0 | 10 | 20 | 80 | 30 | 0 |
| SCR | flowing | 3.0 | 0.8 | 1.5 | 33 | 2.3 | 7.0 | 530 | 70 | 20 | 10 | 0 | 10 | 20 | 60 | 10 | 40 | 60 |
| GAR | flowing | 2.0 | 0.3 | 0.7 | 36 | 1.3 | 7.8 | 561 | 50 | 30 | 20 | 20 | 10 | 50 | 15 | 5 | 0 | 30 |

*extrapolated by Google Earth (http://www.google.com/earth/). Slope was calculated using the 'Show Ruler' tool, tracing a distance of 100 m upstream of the fish sampling site and measuring the elevation (m) at the sampling site (s) and 100 m upstream (u), which served as inputs for solving the equation: slope = [(u – s)/100 m]. Site elevation was calculated as mean between elevation at the site and that measured 100 m upstream

**Eutech's multi-parameter CyberScan PCD650 tester

**river width:** mean distance between two point where waters surface come into contact with the banks, perpendicular to water flow

**mean/max depth:** depth measured from the water surface to the bottom of the river/lake

**elevation:** mean meters above the sea level measured at the sampling site

**slope:** ratio between the vertical change and the horizontal change between the initial and final points of the sampling transect (moving against flow)

**pH:** ph of the water in the sampling site

**conductivity:** measure of conductance (ion concentration) in microsiemens per cm (µS/cm)

**runs:** portion of the river in which water flows fast with no turbulence

**pool:** deep portion of the river in which water flows slowly

**riffle:** shallow portion of the river in which water flows turbulent

**boulders**: riverbed grains which size > 256 mm

**cobbles and pebbles**: riverbed grains which size between 32 and 256 mm

**gravel**: riverbed grains which size between 2 and 32 mm

**sand**: riverbed grains which size between 0.0625 and 2 mm

**Silt and clay**: riverbed grains which size < 0.0625 mm

**vegetation in water**: river/lake surface covered by in water vegetation

**riparian vegetation shade**: vertically projected shape on water surface from surrounding vegetation

**Table S2.** Summary of CR genetic diversity in specimens analysed for morphological differences, according to (a) populations and (b) haplogroups. The sample size (N), the number of haplotypes (Hp), the haplotype diversity (Hd) and the nucleotide diversity (π) are shown for each sampled population. Standard errors (s.e.) are given in parentheses for Hd and π.

(a)

| Site | N | Hp | Hd (±s.e.) | π % (±s.e.) |
|---|---|---|---|---|
| VAR2 | 21 | 3 | 0.58 (±0.08) | 1.52 (±0.80) |
| TRO1 | 6 | 2 | 0.33 (±0.22) | 0.04 (±0.05) |
| TRO2 | 12 | 1 | - | - |
| FOR1 | 7 | 1 | - | - |
| FOR2 | 9 | 2 | 0.39 (±0.16) | 0.04 (±0.05) |
| ARR | 15 | 4 | 0.60 (±0.11) | 0.09 (±0.08) |
| TIB1 | 31 | 4 | 0.60 (±0.07) | 0.08 (±0.07) |
| TIB2 | 6 | 2 | 0.33 (±0.22) | 0.04 (±0.05) |
| TIB3 | 27 | 6 | 0.71 (±0.07) | 0.10 (±0.08) |
| SET | 20 | 5 | 0.44 (± 0.13) | 0.08 (±0.07) |
| FON | 16 | 4 | 0.64 (± 0.08) | 0.05 (±0.05) |
| SCR | 8 | 4 | 0.75 (±0.14) | 0.15 (±0.11) |
| GAR | 24 | 3 | 0.30 (±0.11) | 0.04 (±0.04) |
| | 202 | 21 | 0.72 (±0.03) | 0.71 (±0.37) |

(b)

| Haplogroup | N | Hp | Hd (±s.e.) | π % (±s.e.) |
|---|---|---|---|---|
| A | 159 | 15 | 0.57 (±0.05) | 0.09 (±0.07) |
| B | 36 | 5 | 0.60 (±0.07) | 0.07 (±0.06) |
| C | 7 | 1 | - | - |
| Tot. | 202 | 21 | 0.72 (±0.03) | 0.71 (±0.37) |

**Table S3.** Populations pairwise $\Phi_{ST}$s based on CR sequences of specimen analysed for morphological differences. Haplogroups found in this study for each population are indicated in the first row. Significance thresholds: * = P < 0.05; ** = P < 0.01; *** = P < 0.001.

| | Haplogroup A/C | Haplogroup A | | | | | | | | | | Haplogroup B | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VAR2 | TRO1 | TRO2 | FOR1 | FOR2 | ARR | TIB1 | TIB2 | TIB3 | SCR | GAR | SET | FON |
| VAR2 | | | | | | | | | | | | | |
| TRO1 | 0.15 | | | | | | | | | | | | |
| TRO2 | 0.23 | 0.12 | | | | | | | | | | | |
| FOR1 | 0.17 | 0.03 | 0.00 | | | | | | | | | | |
| FOR2 | 0.20 | 0.07 | 0.17 | 0.09 | | | | | | | | | |
| ARR | 0.29*** | 0.56*** | 0.66*** | 0.60*** | 0.58*** | | | | | | | | |
| TIB1 | 0.34*** | 0.07 | 0.10 | 0.05 | 0.12 | 0.46 | | | | | | | |
| TIB2 | 0.15 | 0.00 | 0.12 | 0.03 | 0.07 | 0.53*** | 0.07 | | | | | | |
| TIB3 | 0.30 | 0.09 | 0.13 | 0.09 | 0.14 | 0.48*** | 0.19*** | -0.04 | | | | | |
| SCR | 0.19 | 0.24 | 0.40 | 0.30 | 0.30 | 0.49*** | 0.31*** | 0.16 | 0.15 | | | | |
| GAR | 0.30 | 0.05 | 0.05 | 0.01 | 0.12 | 0.61*** | 0.14* | -0.11 | 0.04 | 0.27 | | | |
| SET | 0.60*** | 0.96*** | 0.97*** | 0.96*** | 0.96*** | 0.95*** | 0.95*** | 0.96*** | 0.94*** | 0.94*** | 0.96*** | | |
| FON | 0.58*** | 0.97*** | 0.98*** | 0.98*** | 0.97*** | 0.96*** | 0.96*** | 0.97*** | 0.95*** | 0.95*** | 0.97*** | -0.04 | |

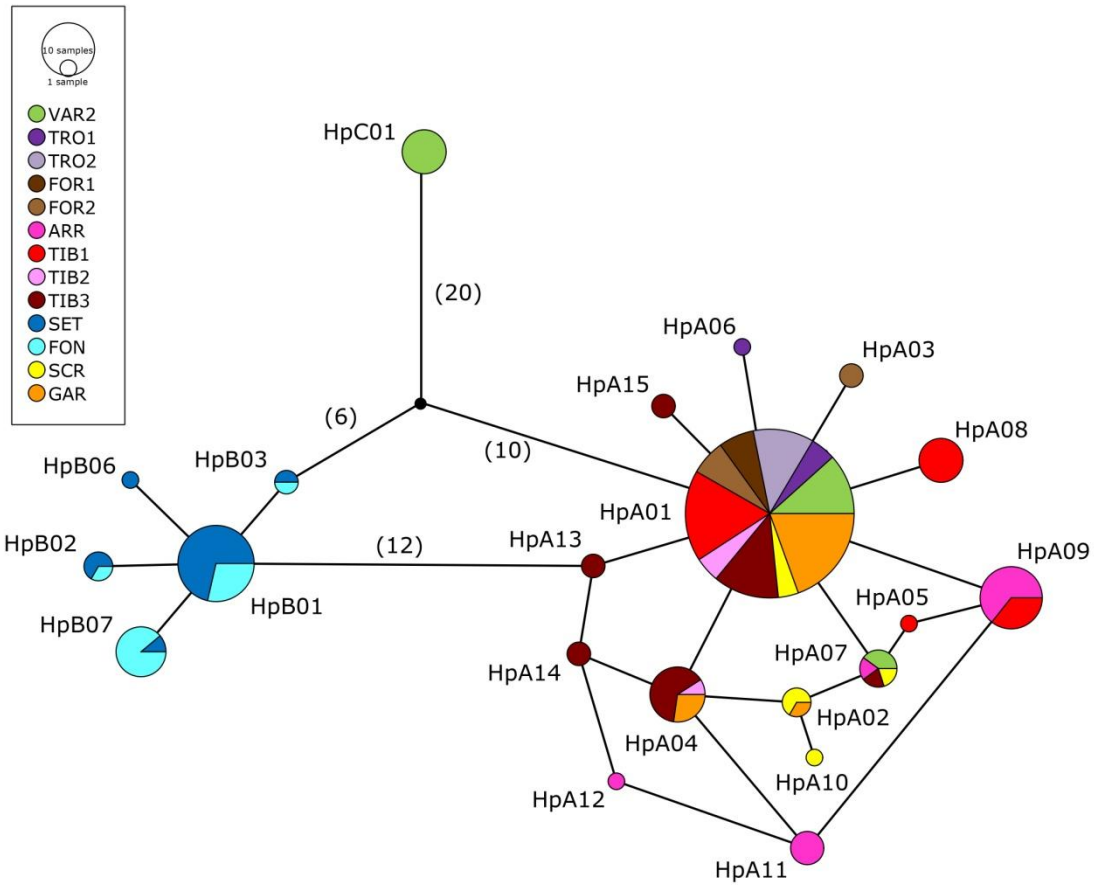



**Figure S1.** Haplotype network based on CR sequences of specimen analysed for morphological differences. Each circle corresponds to one haplotype and its dimension is proportional to the haplotype frequency. The number of nucleotide substitutions between haplotypes is indicated in parenthesis. Haplotype names and numeration follow those in Petrosino *et al.* (2022), with capital letters identifying the lineage (A, B, C). Population abbreviations refer to Table 1.

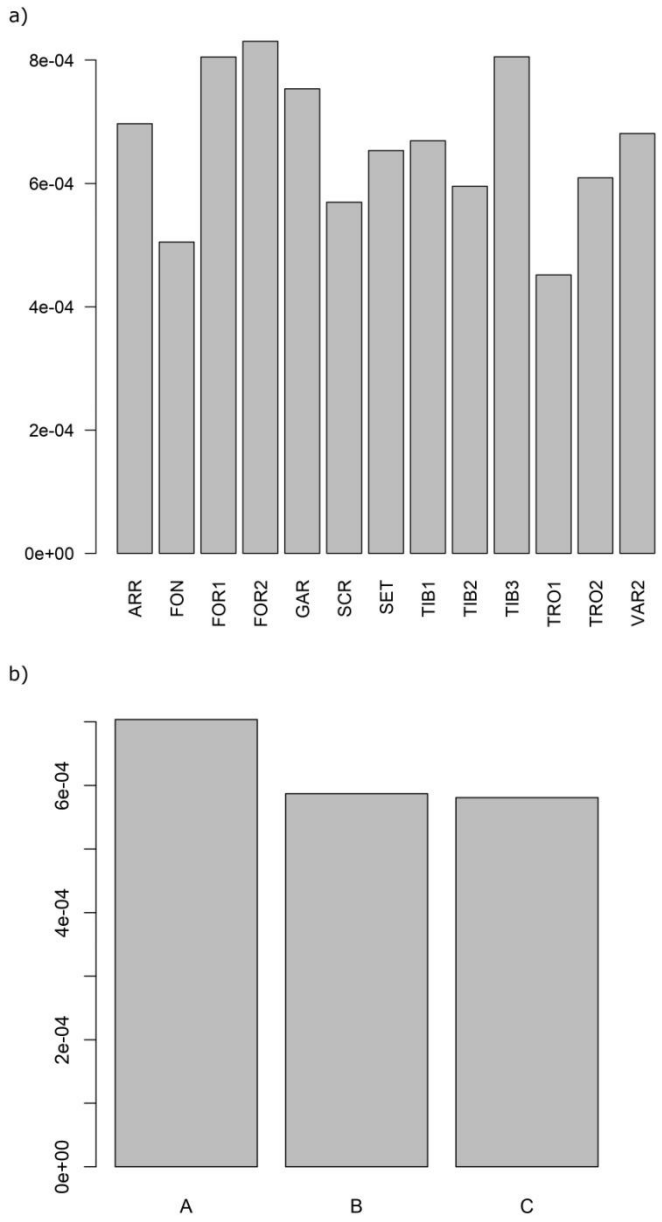



**Figure S2.** Morphological variance barplot (a) per population and (b) per haplogroup.

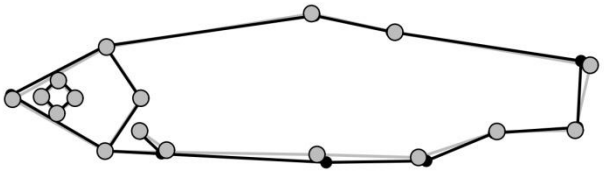


**Figure S3.** Mean body shape for lineage C (gray) and A (black) specimens in VAR2 site. The graph is magnified 2 times.

# CHAPTER 3

# Microsatellite polymorphism in *Sarmarutilus rubilio*: insights into the complex phylogeographic history of an Italian endemic freshwater fish

# Microsatellite polymorphism in *Sarmarutilus rubilio*: insights into the complex phylogeographic history of an Italian endemic freshwater fish


## ABSTRACT

In freshwater fish species microsatellite loci are very useful in disentangling genetic structure. Due to their high mutation rate, they can provide insights about recent isolation or contemporary gene flow between populations and complement mitochondrial investigations, which generally better depict historical processes. In this study, we analysed genetic variation at nine microsatellite loci in the South European roach, an Italian endemic freshwater fish species, in which previous investigations revealed a complex phylogeography and the presence of three different mitochondrial lineages, named A, B and C, the latter suggested to be a putative cryptic species. We observed an overall genetic differentiation between populations belonging to different basins, due to the lack of recent or contemporary connections between them preventing gene flow. We also observed alleles diagnostic for lineage B, but none for C, due to its complete admixing with A in the only basin in which it is present. Moreover, distinct allele frequencies of populations where two lineages coexist suggested that past secondary contact events responsible for their sympatry were followed by isolation of these sites, which lasts up to the present day, thus excluding fish translocations in the last centuries as the main cause of this distribution. Anyway, translocations could explain the only case of high inter-basin similarity observed, between the Tyrrhenian (Tiber River) and Adriatic (Tronto River) slopes of Italy


## INTRODUCTION

Genetic diversity in freshwater fishes is strongly dependent on geological events, which determined past isolation and connection between river basins, and thus populations' genetic divergence, migration and secondary contact events (Won, Jeon, & Suk, 2020; Perea *et al.*, 2021; MacGuigan, Orr, & Near, 2023); in addition, current characteristics of hydrographic networks, e.g. direction and intensity of water flow, coupled with species-specific features (like habitat preferences and dispersal ability) and stochastic factors can determine the amount of genetic differentiation also within rivers (Faulks, Gilligan, & Beheregaray,

2010; Braga-Silva & Galetti, 2016; Pérez-Rodríguez *et al.*, 2021). Microsatellites due to their high variation in the number of repeats (i.e. number of alleles) have been widely used in these taxa as molecular markers, often or in combination with mitochondrial ones, to assess the amount and distribution of genetic variation and disentangle relationships between populations across and within basins, thus making inferences about the drivers of the observed diversity or lack of it (Marchetto *et al.*, 2010; Bezault *et al.*, 2011; Wetjen *et al.*, 2020; Wang *et al.*, 2022).

Microsatellite allele frequencies were also useful to evaluate the effects on genetic diversity of the most common threats to the conservation of freshwater fish species, such as human mediate translocations of specimens from different basins (Launey *et al.*, 2006; Lopes-Cunha *et al.*, 2012), hybridization with introduced allochthonous species or lineages (Meraner *et al.*, 2013; Rossi *et al.*, 2022), and anthropic rivers alteration and fragmentation (i.e. construction of dams, weirs and locks) (Fluker, Kuhajda, & Harris, 2014; Gouskov *et al.*, 2016).

These markers, therefore, represent a powerful tool not only for making inferences about the historical and demographic events that shaped freshwater fishes' genetic structure, but also for identifying populations with unique characteristics which need protection and management interventions, and finally to prevent the loss of genetic variability (Angienda *et al.*, 2011; Kaus *et al.*, 2019; Finger *et al.*, 2022).

The South European roach *Sarmarutilus rubilio* (Bonaparte, 1837) is a small-medium-sized freshwater fish species with a broad ecological niche, belonging to the Leuciscidae family (Schönhuth *et al.*, 2018). It is endemic to the Italian peninsula (Bianco & Ketmaier, 2014), where it inhabits the basins on both the Thyrrenian and Adriatic slopes of Central and South Italy and was also introduced in the southernmost ones and Sicily (Bianco *et al.*, 2013) (Figure 1). The basins in which this species is native were alternatively connected and isolated from each other since Miocene and until the last glacial age (Bianco, 1995), and therefore *S. rubilio* can be a good model for testing the relationship between genetic diversity, past geological events and current isolation of hydrographic networks.

Recent phylogeographic investigations (Petrosino *et al.*, 2022, 2023), based on maternally inherited mitochondrial DNA (the barcoding fragment of cytochrome oxidase I, COI and the non-coding control region, CR) revealed the existence of three different genetic lineages, named haplogroup (Hp) A, B and C;

they show distinctive geographic distributions and likely originated in allopatry, partially according to the boundaries of the ichthyogeographic districts identified in the Italian peninsula (Bianco, 2014). HpA is found to be widespread across the species range, both in the Tyrrhenian and Adriatic slopes of Central Italy; HpB is typical of Fondi Plain, a small area between Central and South Italy, and coexists with HpA in southernmost sites; HpC, the most divergent ones, is found in only two sites from the same basin at the northern edge of *S. rubilio* native range (Magra-Vara basin), always together with HpA. However, the outcomes of these researches revealed the need for further investigations. In detail open questions remained on whether (a) the coexistence of different lineages in the southernmost and northernmost areas is the result of secondary contact that occurred due to past temporary connection between basins, or the consequence of human-mediated translocation of specimens in the last centuries; (b) further genetic subdivision, not detected by mitochondrial analysis, actually exists within haplogroup A according to geographic barriers between basins (e.g. the Apennine Mountains between Tyrrhenian and Adriatic slopes) or lack of gene flow within them. To answer these questions and deeper assess population diversity and genetic structure based on mitochondrial data (Petrosino *et al.*, 2022), we decided to genotype the same *S. rubilio* populations at microsatellite loci (STR); indeed these markers are characterized by a mutation rate higher than that of mtDNA, and thus are more suitable to unravel contemporary population structure and gene flow, while mitochondrial markers generally depict historical processes (Sala-Bozano, Ketmaier, & Mariani, 2009). In detail we aimed to: a) test the existence of distinctive alleles typical of each mitochondrial lineage and identify nuclear patterns for the putative cryptic species; b) compare STRs genetic structure with that inferred by mtDNA, to provide further insights about this species' phylogeography, such as recent basins isolation or gene flow between populations. Finally, Considering that the South European roach conservation status is currently classified as Vulnerable by the International Union for the Conservation of Nature (Rondinini, Battistoni, & Teofili, 2022), these new outcomes will be precious to identifying target populations for conservation interventions due to their genetic distinctiveness, and guide future management actions as required by the by European Habitats Directive for this species (EEC, 1992).

## MATERIALS AND METHODS

STUDY AREA, SAMPLE COLLECTION, AND DNA EXTRACTION

Our analysis was carried out on a subset of 196 specimens from 11 sites, covering most of the species native range (Figure 1; Table 1), among those collected by Petrosino *et al.* (2022) and Petrosino *et al.* (2023), which were already analysed for the mitochondrial control region (CR) and geometric morphometrics. Fishes were caught by electrofishing, anaesthetized and released after taking a small fin clip, that was fixed in 90 % ethanol (for sampling authorizations see the two cited papers). DNA was extracted from fins using the salting out procedure reported by (Aljanabi & Martinez, 1997)

.

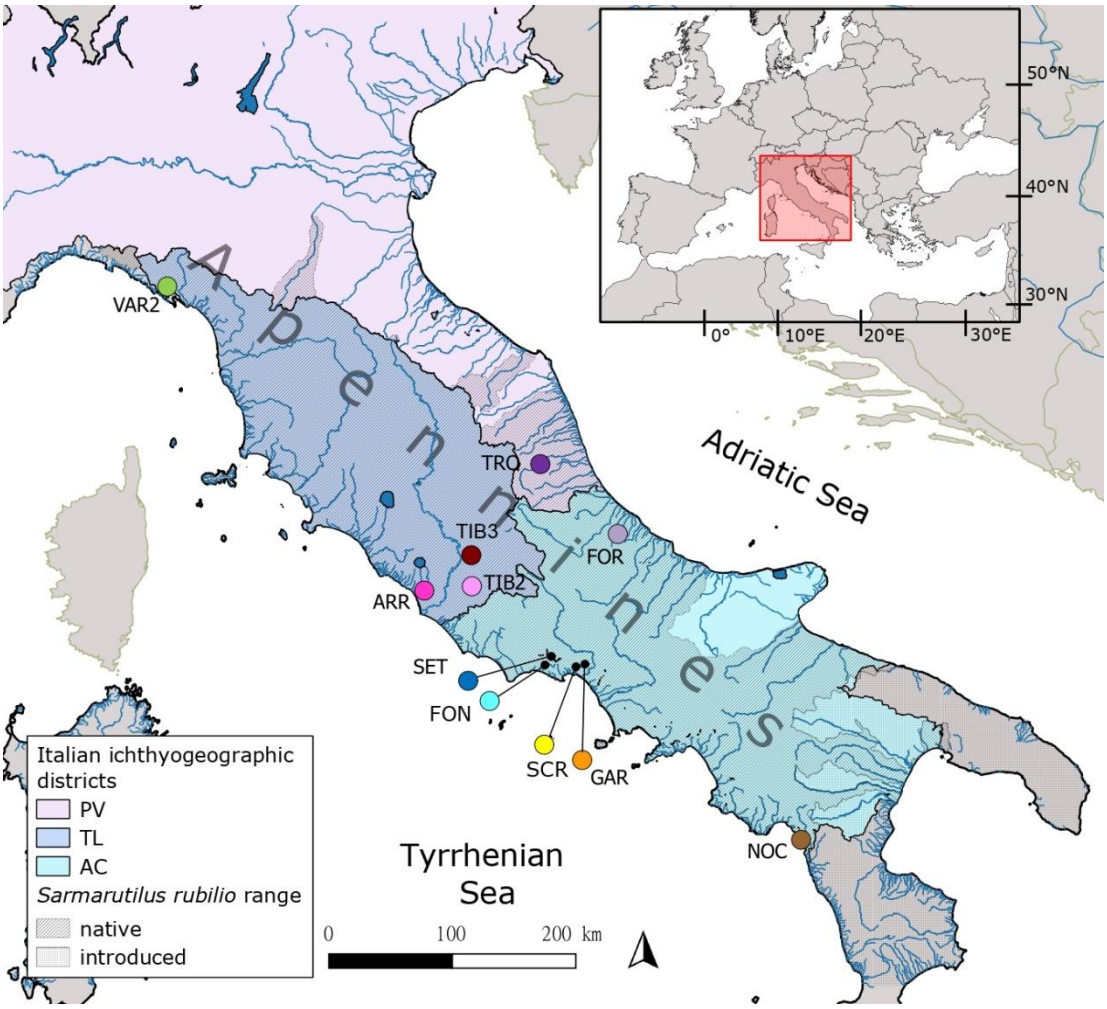



**Figure 1.** Geographic location of the study area (black and red rectangles, top right), along with *S. rubilio* distribution according to IUCN (https://www.iucnredlist.org/species/19786/9014268, accessed on 11 April 2022), and sampling site for populations analysed for microsatellite loci. For site, abbreviations refer to Table 1. Modified from Petrosino *et al.* (2022); the extent of ichthyogeographic districts was changed following the inferences from *S. rubilio* mitochondrial analysis discussed within the reference paper

.

**Table 1.** Summary of genetic variation for the nine microsatellites, along with sampling details for eleven populations of South European roach from nine drainage basins in Italy. The sampling location (Site) and geographic coordinates (datum = WGS84), the sample size (N), the mean number of alleles and percentage of private alleles (A), the allelic richness (A rich), the observed (Ho) and expected (He) heterozygosity are shown for each population (Pop). Standard errors (s.e.) are given in parentheses for Hd, π, Ho and He. Sampling sites (Pop ID) were named accordingly to Petrosino *et al.* (2022)and Petrosino *et al.* (2023); mitochondrial haplogroups identified in those studies were indicated (Hap).

| | | | | | | **Microsatellites** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Drainage basin** | **Site** | **Pop ID** | **Lat (°N)** | **Lon (°E)** | **N** | **A %private** | **A rich** | **Ho** (± s.e.) | **He** (± s.e.) | **Hap** |
| Magra-Vara | Graveglia | VAR2 | 44.191 | 9.790 | 20 | 9.33<br>7.14% | 6.12 | 0.68<br>(0.01) | 0.70<br>(0.07) | A and C |
| Tronto | Tronto | TRO | 42.802 | 13.465 | 15 | 5.11<br>2.17% | 4.33 | 0.65<br>(0.09) | 0.62<br>(0.07) | A |
| Foro | Foro | FOR | 42.246 | 14.186 | 17 | 5.67<br>1.94% | 4.42 | 0.55<br>(0.10) | 0.59<br>(0.09) | A |
| Arrone | Arrone | ARR | 41.914 | 12.265 | 20 | 5.67<br>1.94% | 4.29 | 0.58<br>(0.08) | 0.57<br>(0.08) | A |
| Tiber | Fosso San Vittorino | TIB2 | 41.911 | 12.788 | 18 | 7.44<br>9.01% | 5.28 | 0.67<br>(0.08) | 0.67<br>(0.06) | A |
| | Fosso Corese | TIB3 | 42.170 | 12.745 | 25 | 6.78<br>6.49% | 4.55 | 0.57<br>(0.08) | 0.57<br>(0.08) | A |
| Fondi | Settecannelle | SET | 41.368 | 13.421 | 20 | 5.33<br>2.06% | 4.30 | 0.54<br>(0.11) | 0.55<br>(0.12) | B |
| | Canale San Magno | FON | 41.347 | 13.379 | 15 | 6,67<br>11,69% | 4.76 | 0.61<br>(0.11) | 0.58<br>(0.10) | B |
| Santa Croce | Santa Croce | SCR | 41.287 | 13.716 | 22 | 6,67<br>4.95% | 4.74 | 0.59<br>(0.09) | 0.61<br>(0.08) | A and B |
| Liri-Garigliano | Ausentello | GAR | 41.303 | 13.743 | 18 | 5.78<br>3.81% | 4.27 | 0.67<br>(0.09) | 0.59<br>(0.06) | A |
| Noce | Canale Pamafi | NOC | 39.934 | 15.752 | 6 | 3.22<br>0% | NA | 0.54<br>(0.11) | 0.47<br>(0.09) | A and B |

PCR AMPLIFICATION AND GENOTYPING

In the literature, there were no STRs specifically designed for *S. rubilio* or tested for cross-species amplifications in this species. So we initially tested 12 polymorphic microsatellite markers among a subset of those previously developed for other leuciscid species. We obtained successful amplifications and genetic data for 9 of them, namely: Sluc5 and Sluc13 (Gigliarelli *et al.*, 2012), LC27 (Seifertová *et al.*, 2012), LleA-150, Lsou05, BL1-T2, Ca1, Ca3 and N7K4 (Dubut *et al.*, 2010). Other 4 new loci were tested to replace those for which amplification failed and they all returned satisfactory preliminary results; we are currently amplifying and genotyping these new STRs to implement our data set to 13 microsatellite loci. For each locus, we designed an optimized protocol, and the 5' end of one of the two primers was labelled with a fluorescent dye (Table S1, S2). After microsatellites amplification was completed, we sent amplicons to Macrogen company (dna.macrogen.com) for length polymorphism screening, which returned results as FSA format files (ABI 3730xl System genetic analyzer with GeneScan-400HD as internal size standard was used by the external service). Subsequently, we determined allele sizes for each specimen using the Peak Scanner Software 2.0 (Applied Biosystems).

GENETIC DIVERSITY

Genetic diversity parameters, like the number of alleles, private alleles, observed and expected heterozygosity, for each population were calculated with GenAlEx 6.5 (Peakall & Smouse, 2006); allelic richness was calculated using the *PopGenReport* package version 3.0.7 (Adamack & Gruber, 2014) in R version 4.0.3 (R Core Team, 2020), setting a random sample of 15 individuals per population to account for different sample size. The frequency of null alleles was estimated with FreeNA (Chapuis & Estoup, 2007) for each locus and population according to Dempster's EM algorithm. GenePop 4.7.5 (Rousset, 2008) was used for evaluating deviations from Hardy-Weinberg equilibrium and linkage disequilibrium, adjusting the *p*-value with Holm correction (Holm, 1979). Signatures of recent population bottlenecks were tested in BOTTLENECK 1.2 (Piry, Luikart, & Cornuet, 1999), setting the microsatellite mutation model (TPM: 95 % single-step mutations and 5 % multistep mutations, variance among multiple steps = 12) with 10,000

replications and one-tailed Wilcoxon signed-rank test for statistical significance. We also carried out an outlier analysis to identify the presence of loci under selection using BayeScan 2.1 (Foll & Gaggiotti, 2008).

GENETIC STRUCTURE

Populations' genetic differentiation was estimated by calculating pairwise Nei's $F_{ST}$ in Arlequin 3.5 (Excoffier & Lischer, 2010) and assessing significance with 10000 permutations with Holm correction. In addition, we re-calculated pairwise $F_{ST}$s splitting VAR, SCR and NOC specimens into two groups each (VAR2-HpC, N = 8; VAR2-HpA, N = 12; SCR-HpA, N = 21; SCR-HpB, N = 1; NOC-HpA, N = 5; NOC-HpB, N = 1), according to their mitochondrial lineage previously identified by (Petrosino *et al.*, 2022), to test if different lineages within the same population showed different alleles. Genetic relationships among populations were visualized with a non-metric multidimensional scaling (NMDS) implemented in PAST 3.26 (Hammer, Harper, & Ryan, 2001), based on the pairwise $F_{ST}$s matrix. $F_{ST}$ overall populations for each locus was estimated using Arlequin, with 10000 permutations to assess the contribution of each locus to the between-populations differentiation.

To make inferences about the number of genetic groups we performed a multi-locus Bayesian analysis of population structure in STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000). We performed five repeated runs of the admixture model with correlated allele frequencies for each K value in the range 1–14 (i.e. the maximum number of populations if we split VAR2, SCR and NOC into two groups each, see above) and no population origin priors, setting 500,000 iterations after a burn-in period of 200,000 iterations; the results of replicated runs were combined with CLUMPAK (Kopelman *et al.*, 2015). To find the number of groups (K) that best explain the genetic partition in our data we evaluated different methods using Structure Selector (Li & Liu, 2018): the LnP method (Pritchard *et al.*, 2000) based on the highest log probability of K; the ΔK method (Evanno, Regnaut, & Goudet, 2005) based on the highest rate of change in LnP between successive K; MedMedK, MedMeaK, MaxMedK and MaxMeaK estimators proposed by Puechmaille (2016).

To test various hypotheses of (hierarchical) spatial patterns of genetic structure we also performed AMOVAs (Analysis of Molecular Variance) using Arlequin and considering different options: (1) absence of genetic structure (i.e., no grouping); (2) nine groups corresponding to basins; (3) four groups according to combinations of mitochondrial lineages found in populations (see Hap column in Table 1). To further

explore the relationship between the genetic structure revealed by the STR analysis and that identified by the mtDNA, we tested the correlation between pairwise $F_{ST}$s and $\Phi_{ST}$ (the latter recalculated for the same populations investigated here, using CR sequences from Petrosino *et al.*, 2022 and Petrosino *et al.*, 2023) using the Mantel test.

## RESULTS

### GENETIC DIVERSITY

Presently we have complete data on 9 loci and the amplification of the remaining 4 loci is still going on. The number of alleles at each locus ranged from 5 to 39 (Table S3). The highest mean number of alleles across all loci was observed in VAR2 (9.33), which also showed the highest allelic richness (6.12); conversely, if we exclude the NOC site due to its few specimens, the lowest values were observed in TRO and SET (5.11 and 5.33 respectively) while allelic richness was similar across all populations but VAR2 and TIB2 (Table 1). Significant deviation from Hardy-Weinberg equilibrium was observed only in VAR2 at the Sluc5 locus (Table S4), likely due to the presence of null alleles (30% of frequency); other null alleles were identified in single loci/sites (Sluc5 in SCR, Ca1 in TRO and FOR, and LC27 in FOR), with frequencies between 11 and 16%)(Table S5). We excluded physical linkage between loci as linkage disequilibrium was significant ($p < 0.05$) in only 1 out of 396 comparisons (Ca3-Lsou05 in SCR). Bottleneck analysis revealed population size reduction only in SET ($p = 0.02$). Finally, BayeScan analysis revealed the Sluc5 locus to be under selection ($q$-value < 0.05), however, we decided to retain this locus in the subsequent analysis as we observed alleles strongly associated with mitochondrial lineages and which highly contributed to populations differentiation (see below).

### GENETIC STRUCTURE

Most Pairwise $F_{ST}$s were significant, except for those involving NOC (due to its few specimens) and within basins (SET and FON, TIB2 and TIB3). We also observed a lack of differentiation between TIB2 and some populations from other basins (VAR2, TRO and ARR) (Table 2). The highest values of $F_{ST}$ were observed when comparing FON and SET to all other populations. When we split VAR2, SCR and NOC specimens

according to mitochondrial lineages we didn't observe differentiation in allelic frequencies within these populations and overall $F_{ST}$s were identical to those observed without the split.

**Table 2.** Pairwise $F_{ST}$ between populations. Significance thresholds: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

| | | VAR2 | TRO | FOR | ARR | TIB2 | TIB3 | SET | FON | SCR | GAR | NOC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Central Italy | VAR2 | - | | | | | | | | | | |
| | TRO | 0.04* | - | | | | | | | | | |
| | FOR | 0.09*** | 0.04 | - | | | | | | | | |
| | ARR | 0.08*** | 0.03 | 0.11** | - | | | | | | | |
| | TIB2 | 0.02 | -0.01 | 0.06** | 0.02 | - | | | | | | |
| | TIB3 | 0.04*** | 0.01 | 0.11*** | 0.05*** | 0.01 | - | | | | | |
| South Italy | SET | 0.27*** | 0.33*** | 0.28*** | 0.35*** | 0.30*** | 0.36*** | - | | | | |
| | FON | 0.25*** | 0.32*** | 0.27*** | 0.33*** | 0.27*** | 0.35*** | 0.00 | - | | | |
| | SCR | 0.07*** | 0.06*** | 0.10*** | 0.09*** | 0.05*** | 0.06*** | 0.34*** | 0.31*** | - | | |
| | GAR | 0.12*** | 0.10*** | 0.13*** | 0.15*** | 0.08*** | 0.13*** | 0.36*** | 0.33*** | 0.05** | - | |
| | NOC | -0.04 | -0.03 | -0.01 | -0.04 | -0.15 | -0.05 | 0.20*** | 0.16*** | -0.08 | -0.05 | - |

The NMDS analysis based on the $F_{ST}$s values provided ordination stress (0.25) slightly exceeding the commonly accepted limit of 0.2. Considering that it should not be intended as a strict cutoff (Dexter *et al.*, 2018) we assumed the NMDS adequately represented the distances between populations: we observed high divergence of SET and FON sites from all the others and, among the latter, a further divergence of SCR and GAR from the remaining populations (Figure 2). When NMDS was repeated excluding SET and FON, we did not observe clear ordination patterns.

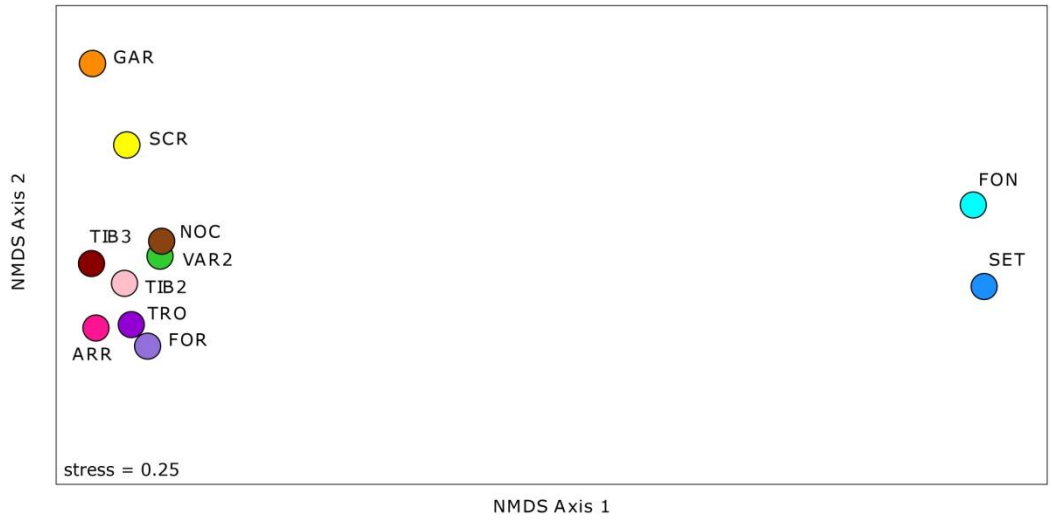



**Figure 2.** Non-metric multidimensional scaling based on $F_{ST}$ matrix between populations. Colours and abbreviations as in Figure 1.

When $F_{ST}$ contribution per locus was examined Sluc5 and Ca1 showed the highest values (0.47 and 0.29 respectively) (Table S6). We found that these results were linked to the frequencies of Sluc5 alleles 227 and 230 and Ca1 alleles 101 and 107: these were fixed in SET and FON populations (Table S3, Figure S1), in which only mitochondrial haplogroup B was found (see Table 1), and were also observed in SCR, GAR and NOC.

STRUCTURE analysis supported the presence of different population clusters (K): 2 (according to Evanno's method, Figure 2b), 5 (according to Puechmaille's methods Figure S2c,d) and 6 (according to Puechmaille's and Pritchard's methods, Figure S2a,e,f). When K=2 (as well as in the other options) SET and FON always formed one homogenous cluster, distinct from all other populations, according to the NMDS analysis (Figure 3a). When K = 5, VAR2 and South Italy populations (SCR, GAR and PAM) were considered as different groups each despite few specimens showed not clear assignment (Figure 3b); finally, for K = 6, ARR and FON sites are better defined as two different groups from TRO, TIB2 and TIB3 clusters, whose individuals showed a similar amount of uncertainty in their classification (Figure 3c).

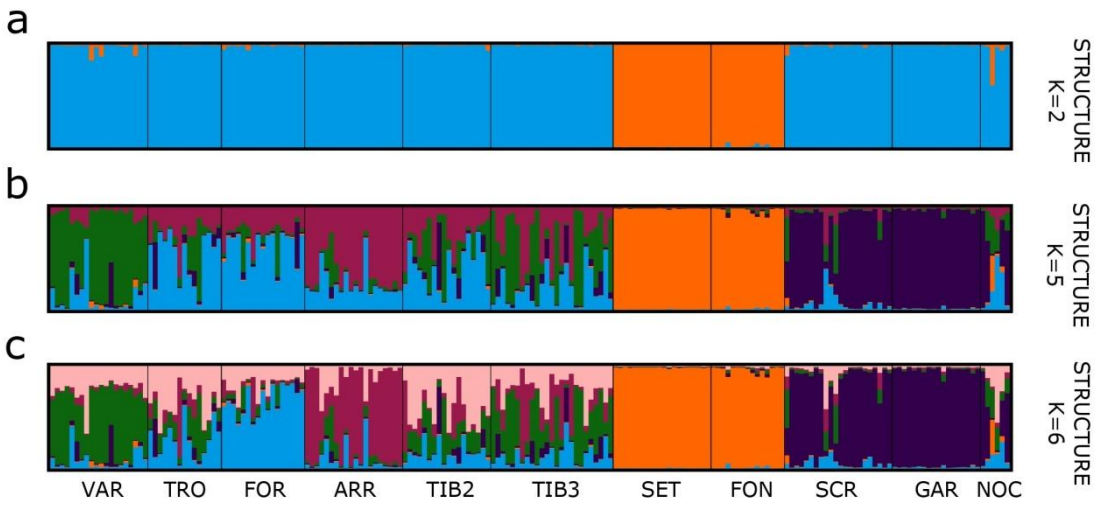



**Figure 3.** Assignment graphs based on the individual genotypes in STRUCTURE for K = 2 (a), K = 5 (b) and K = 6 (c). Each colour represents an inferred genetic cluster, and each vertical line represents a single individual. Different colours in the vertical lines show the proportion of assignment of a single individual to each cluster. Populations abbreviations refer to Table 1.

The AMOVA analysis provided significant results for the hypothesis of genetic variance partition according to haplogroups combinations, showing the maximum variance among groups (16.81%) and the minimum within groups (4.02% ) (Table 3). Conversely the hypothesis of genetic variance partition according to drainages (nine groups) was rejected. A significant correlation (0.82, $p < 0.01$) was observed between STR $F_{ST}$ and mitochondrial CR $\Phi_{ST}$ (Table S7) .

**Table 3.** AMOVA hierarchical analysis examining the partitioning of genetic variance of 9 microsatellite loci according to different hypothesized structures: no structure (one group), drainages (nine groups), haplogroups combinations (four groups, see Table 1). Significance thresholds: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

| Number of groups (group composition) | Hierarchical level | Variation% | F-statistic |
|---|---|---|---|
| One group | Among populations | 16.18 | 0.162*** |
| | Within populations | 83.82 | |
| Four groups: haplogroup combinations VAR2/(TRO,FOR,ARR,TIB2,TIB3,GAR)/(SET,FON)/(SCR,NOC) | Among groups | 16.81 | 0.168** |
| | Within groups | 4.02 | 0.048*** |
| | Within populations | 79.17 | 0.208*** |
| Nine groups: Drainages VAR2/TRO/ARR/FOR/(TIB2,TIB3)/(SET,FON)/SCR/GAR/NOC | Among groups | 16.11 | 0.161 |
| | Within groups | 0.66 | 0.007 |
| | Within populations | 83.23 | 0.167*** |

## DISCUSSION

The distribution of intraspecific genetic diversity among populations is linked to historical dispersal (affected by geological and climatic events) and current migration (ecological scale dispersal) (Miller-Sims *et al.*, 2008).

Microsatellite loci, due to their fast evolution rate, better assess genetic variation between populations at short temporal (i.e. recently diverged populations) and spatial scale (e.g. within or between adjacent basins) compared to other molecular markers, and thus can complement mtDNA data (Bartáková, Bryja, & Reichard, 2018). In this study, their application on the South European roach gave us further insights into the relationships among populations, previously inferred from the mitochondrial analysis alone (Petrosino *et al.*, 2022).

### GENETIC DIVERSITY

We tested nine microsatellite markers developed for other species and they all were polymorphic in *S. rubilio*: excluding NOC (due to the few specimens analyzed), we found private alleles in each population, which were all characterized by high genetic diversity (Ho and He > 0.5), according to the mean value of heterozygosity (0.46) reported in freshwater fishes (DeWoody & Avise, 2000). We observed an overall

significant differentiation between populations (47 out of 55 pairwise $F_{ST}$ were statistically supported) with contribution from all loci, but mainly Sluc5 and Ca1, with the former being putatively under selection. These two loci showed alleles highly associated with mitochondrial lineage B or A: fixed alleles in SET and FON populations, where all specimens belong to haplogroup B, were found also in SCR and NOC, the ones in which B was found too (see Table 1 and Figure S1), and they were never observed in Central Italy populations, which is likely the origin area of haplogroup A (Petrosino *et al.*, 2022); conversely no association was observed between specific alleles and lineage C.

GENETIC STRUCTURE

The analyses of genetic differentiation between populations based on STRs were mostly concordant with that defined by the mtDNA investigation (as revealed by AMOVA and Mantel test), supporting high divergence of SET and FON from all the other populations. However STRs genetic structure showed some discrepancies from that inferred from CR: more than four genetic groups (i.e. the number of haplogroup combinations) were supported, pointing out the divergence of SCR and GAR in South Italy from the remaining populations and further differences between central Italy populations, that were not detected by the mtDNA phylogeographic analysis (Petrosino *et al.*, 2022).

Among the three identified mitochondrial lineages, C was strictly localized in the basin corresponding to VAR2, where it coexisted with A, and was the most differentiated. According to STRs analysis, we didn't find differences within VAR2 (between A and C lineage specimens) nor strong divergence of VAR2 from other populations from Central Italy ($F_{ST} < 0.1$). However, this site could be distinguished from other populations when five or six groups were defined, due to its specific allele frequencies. *S. rubilio* mtDNA data suggested that the presence of two lineages in VAR2 could be explained considering that lineage A reached this site, where the local population was characterized by the native lineage C when a connection between this basin (Magra-Vara) and those of the southern area was established during glacial periods (Petrosino *et al.*, 2022). Now microsatellite data allowed us to better define this hypothesis: indeed, the past secondary contact and hence the gene flow between the two allopatric lineages (A and C) was confirmed by the microsatellite high heterozygosity and allelic richness observed in VAR2, and we can infer that this site subsequently

experienced new isolation, that lasts up to the present day. Indeed, although VAR2 haplotypes of lineage A correspond to those present in other populations, distinctive microsatellite allele frequencies in this site suggest the independent evolution of this population after the secondary contact; furthermore according to the high number of private alleles, gene flow between VAR and southern sites likely did not occur in recent times (confirming the glacial period time scale hypothesis), similarly to what has already been reported for another leuciscid species from Italy, *Squalius lucumonis* (Rossi *et al.*, 2021). Finally, the long-term coexistence of two highly divergent lineages determined a full and distinctive intermixing of genotypes (also indicating that likely the reproductive barriers were not stabilized) and as a result, presently it is not possible to identify distinct alleles assemblages diagnostic for each of the two lineages in VAR2, despite specimens homozygotic for specific C lineage nuclear sequences were identified with Cyfun P nuclear marker (Petrosino *et al.*, 2022). The genotypes' intermixing could also had a role in determining the lack of morphological differences observed between the two lineages in VAR2 (Petrosino *et al.*, 2023), as observed in hybridization events between other Leuciscidae species (Hayden *et al.*, 2010; Valić *et al.*, 2013; Tancioni *et al.*, 2013). Finally, microsatellite data of VAR2 allowed us to exclude that the mitochondrial admixture observed in this site is due to a recent human mediate translocation of fishes, as we didn't find specimens with alleles' combination identical to that of any other basin, as conversely observed when fishes are manipulated by human activities (Rossi *et al.*, 2022; Kitada, 2022).

Microsatellites allowed a clear distinction between populations where haplogroup B is fixed (SET and FON in Fondi plain) and all the others ($F_{ST} > 0.2$, K = 2); so despite the divergence of mitochondrial lineage B from A is more recent than that of lineage C (Petrosino *et al.*, 2022), the lack of secondary contact between A and B in the Fondi plain preserved the original combination of alleles typical of the latter. Indeed the alleles present here are likely diagnostic for lineage B (see below in the Discussion) however, considering that SET and FON belong to the same hydrographic network which is isolated from the surrounding basins, we cannot exclude that genetic drift following isolation (Nguyen & Sunnucks, 2012) may have contributed further to their high divergence from any other site.

In South Italy SCR and GAR populations sites clustered together and were differentiated from the populations from Central Italy (TRO, FOR, ARR, TIB2 and TIB3). Like in VAR2 the distinctive signature of SCR

and GAR could be due to secondary contact, between Central Italy native lineage A and South Italy native B (Petrosino *et al.*, 2022), so the admixture of alleles typical of the two genetic groups may have produced distinguishable allelic frequencies identified by STRUCTURE analysis. Indeed HpB haplotypes were found in SCR, although only at low frequencies - only one out of thirty-two, see Petrosino *et al.* (2022) - and we cannot exclude they could have been overlooked in the mitochondrial analysis of the geographically close GAR population, due to not extensive sampling. To support this hypothesis we found that alleles fixed in lineage B (observed in SET and FON) were also observed in SCR and GAR, but not in the populations from Central Italy characterized by lineage A. Analysis of the additionally tested loci (see Materials and Methods) will help in clarifying the relationship between South Italy populations and the others and why specimens with mitochondrial haplogroup B are very few (or absent) compared to A ones in this area. Finally, consideration on the southernmost site (NOC) could not be done, because of the few specimens analysed.

Further genetic divergence was observed between populations of Central Italy, all characterized by the presence of lineage A only. According to K = 6, both ARR (Arrone basin) and FOR (Foro basin) clustered in a different group, distinguishable from that of TRO, TIB2 and TIB3 (Tronto and Tiber basins), despite few specimens misclassification. This structure is consistent with the hypothesis of lack of gene flow between different yet geographically close basins (Arrone, Foro and Tiber) and conversely of current gene flow within the same basin (i.e. between TIB2 and TIB3 within Tiber), as commonly observed within freshwater fish species (Coleman *et al.*, 2010; Wetjen *et al.*, 2020; Amoutchi *et al.*, 2023); noteworthy lack of genetic differentiation was observed between Tronto and Tiber basins. ARR differentiation from Tiber populations (TIB2, TIB3) was consistent with the slight divergence between them revealed by the mtDNA (Petrosino *et al.*, 2022, 2023). As TRO population is on the Adriatic slope, presently isolated from the TIB sites (Tiber River, on the Tyrrhenian slope) we cannot exclude the translocations of individuals from the Tiber basin in this site.

Finally, the lack of microsatellites differentiation between Tiber basin populations (TIB2 and TIB3) seems consistent with current gene flow among them, or until very recent times, considering that several anthropic barriers (i.e. dams and weirs) were built since 1950 in the investigated area. This is consistent with the ecological features of the South European roach, a eurytopic species (Bianco *et al.*, 2013; Di Tizio &

Di Felice, 2016) that supposedly can disperse through different freshwater environments within the same watercourse, allowing gene flow between populations. Indeed other species characterized by high specific habitat preferences and fragmented distribution provided different results for the same sites within the Tiber River (Rossi *et al.*, 2021).

Analysis of more loci and one additional population from the Tiber basin is currently ongoing. Future results will clarify whether there is a real lack of differentiation at the intrabasin (TIB2 vs TIB3) and interbasin (TRO vs TIB2/3) scale and if this is related to effective gene flow between populations and/or human translocations.

## CONCLUSIONS AND MANAGEMENT IMPLICATIONS

Microsatellite analysis in *S. rubilio* allowed us to further deepen the phylogeographic structure revealed by the mitochondrial investigation and provided useful data for species management.

First, VAR2 allelic frequencies (a) supported the distinction of this population from the others, and thus its long-term isolation, but (b) demonstrated the complete admixing of the two distinct lineages (A and C) present in this site. This, coupled with the lack of alleles diagnostic for the highly divergent lineage C and the lack of morphological differences with A, does not support the hypothesis that the former currently corresponds to a cryptic species.

Second, the complete genetic isolation of populations from a small area between Central and South Italy, i.e. SET and FON in the Fondi plain suggested by mtDNA data was confirmed: these sites showed typical microsatellite alleles associated with lineage B and strong divergence from other sites.

Thus our investigations underlined the genetic uniqueness of two distinct geographic areas: the Magra-Vara basin (VAR2) and Fondi plain (SET and FON), that should be considered as distinct Management Units – Mus, see Moritz (1994) and Allendorf (2017) - and prioritized in conservation actions required by the by European Habitats Directive for *S. rubilio* (EEC, 1992). Indeed considering that they have limited geographic extension, corresponding populations may be more susceptible to habitat alterations due to climate change or human activities, or to other threats such as the introduction of allochthonous species.

Third, distinctive allele frequencies in SCR/GAR sites allowed us to exclude recent human mediate translocation to explain the coexistence of different lineages in this area. However the same cannot be inferred in other geographic areas: further investigation is needed to clarify the relationship between the Tiber basin and those from the Adriatic slope (e.g. TRO and FOR), to define where and how past geological events and/or translocations allowed the South European roach to cross the Apennine Mountains barrier among them (Bianco, 1994).

Finally, as a general comment on species management, we recommend the combined use of both *in situ* interventions, i.e. sustainable conservation of stream flows, habitat restoration, pollution reduction and "genetic sanctuary" creation (Arthington *et al.*, 2016), and *ex-situ* ones, i.e. release of hatchery-reared juveniles, with the necessity to use breeders collected from the same or the nearby areas of intervention to preserve population genetic identity (Sousa-Santos, Gil, & Almada, 2014).

## REFERENCES

**Adamack AT & Gruber B**. **2014**. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* **5**: 384–387.

**Aljanabi SM & Martinez I**. **1997**. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research* **25**: 4692–4693.

**Allendorf FW**. **2017**. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* **26**: 420–430.

**Amoutchi AI, Kersten P, Vogt A, Kohlmann K, Kouamelan EP & Mehner T**. **2023**. Population genetics of the African snakehead fish *Parachanna obscura* along West Africa's water networks: Implications for sustainable management and conservation. *Ecology and Evolution* **13**.

**Angienda PO, Lee HJ, Elmer KR, Abila R, Waindi EN & Meyer A**. **2011**. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conservation Genetics* **12**: 243–255.

**Arthington AH, Dulvy NK, Gladstone W & Winfield IJ**. **2016**. Fish conservation in freshwater and marine realms: status, threats and management. *Aquatic Conservation: Marine and Freshwater Ecosystems* **26**: 838–857.

**Bartáková V, Bryja J & Reichard M**. **2018**. Fine-scale genetic structure of the European bitterling at the intersection of three major European watersheds. *BMC Evolutionary Biology* **18**: 105.

**Bezault E, Balaresque P, Toguyeni A, Fermon Y, Araki H, Baroiller JF & Rognon X**. **2011**. Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* **12**: 102.

**Bianco PG**. **1994**. L'Ittiofauna continentale dell'Appennino umbro-marchigiano, barriera semipermeabile allo scambio di componenti primarie tra gli opposti versanti dell'Italia centrale. *Biogeographia – The Journal of Integrative Biogeography* **17**.

**Bianco PG**. **1995**. Factors Affecting the Distribution of Freshwater Fishes Especially in Italy. *Cybium*.

**Bianco PG, Caputo V, Ferrito V, Lorenzoni M, Nonnis Marzano F, Stefani F, Sabatini A & Tancioni L. 2013**. *Rutilus rubilio*. In: Rondinini C, Battistoni A, Peronace V, Teofili C, eds. *Lista Rossa IUCN dei Vertebrati Italiani.* . Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare. http://www.iucn.it/scheda.php?id=-251974278.

**Bianco PG**. **2014**. An update on the status of native and exotic freshwater fishes of Italy. *Journal of Applied Ichthyology* **30**: 62–77.

**Bianco PG & Ketmaier V**. **2014**. A revision of the *Rutilus* complex from Mediterranean Europe with description of a new genus, *Sarmarutilus*, and a new species, *Rutilus stoumboudae* (Teleostei: Cyprinidae). *Zootaxa* **3841**: 379–402.

**Braga-Silva A & Galetti PM**. **2016**. Evidence of isolation by time in freshwater migratory fish *Prochilodus costatus* (Characiformes, Prochilodontidae). *Hydrobiologia* **765**: 159–167.

**Chapuis MP & Estoup A**. **2007**. Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution* **24**: 621–631.

**Coleman RA, Pettigrove V, Raadik TA, Hoffmann AA, Miller AD & Carew ME**. **2010**. Microsatellite markers and mtDNA data indicate two distinct groups in dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae), a threatened freshwater fish from south-eastern Australia. *Conservation Genetics* **11**: 1911–1928.

**DeWoody J & Avise JC**. **2000**. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* **56**: 461–473.

**Dexter E, Rollwagen-Bollens G & Bollens SM. 2018.** The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography Methods* **16**: 434-443.

**Di Tizio L & Di Felice PL**. **2016**. *Rutilus rubilio* (Bonaparte, 1837) (Rovella). In: Stoch F, Genovesi P, eds. *Manuali per il monitoraggio di specie e habitat di interesse comunitario (Direttiva 92/43/CEE) in Italia: specie animali*. ISPRA, 164–165.

**Dubut V, Sinama M, Martin JF, Meglécz E, Fernandez J, Chappaz R, Gilles A & Costedoat C**. **2010**. Cross-species amplification of 41 microsatellites in European cyprinids: A tool for evolutionary, population genetics and hybridization studies. *BMC Research Notes* **3**: 135.

**EEC**. **1992**. Council Directive 92/43/ECC. *Official Journal of the European Union* **94**: 40–52.

**Evanno G, Regnaut S & Goudet J**. **2005**. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

**Excoffier L & Lischer HEL**. **2010**. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**: 564–567.

**Faulks LK, Gilligan DM & Beheregaray LB**. **2010**. Islands of water in a sea of dry land: hydrological regime predicts genetic diversity and dispersal in a widespread fish from Australia's arid zone, the golden perch

(*Macquaria ambigua*). *Molecular Ecology* **19**: 4723–4737.

**Finger AJ, Benjamin A, Crookshanks C, Campbell MA & Sağlam İK**. **2022**. Broad- and fine-scale structure across the distribution of the relict dace (*Relictus solitarius*) in the Great Basin desert, USA. *Conservation Science and Practice* **4**.

**Fluker BL, Kuhajda BR & Harris PM**. **2014**. The effects of riverine impoundment on genetic structure and gene flow in two stream fishes in the Mobile River basin. *Freshwater Biology* **59**: 526–543.

**Foll M & Gaggiotti O**. **2008**. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–993.

**Gigliarelli L, Puletti ME, Giannetto D, Franchi E, Lanfaloni L, Panara F, Lorenzoni M & Lucentini L**. **2012**. Isolation of microsatellite markers in *Squalius lucumonis* (Bianco, 1983) and cross-species amplification within the family Cyprinidae and other freshwater fish species. *Italian Journal of Zoology* **79**: 169–174.

**Gouskov A, Reyes M, Wirthner-Bitterlin L & Vorburger C**. **2016**. Fish population genetic structure shaped by hydroelectric power plants in the upper Rhine catchment. *Evolutionary Applications* **9**: 394–408.

**Hammer Ø, Harper DAT & Ryan PD**. **2001**. PAST : Paleontological Statistics Software Package for Education and Data Analysis PAST : PALEONTOLOGICAL STATISTICS SOFTWARE PACKAGE FOR EDUCATION AND DATA ANALYSIS. *Palaeontologia Electronica* **4**: 1–9.

**Hayden B, Pulcini D, Kelly-Quinn, 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**: 169.

**Holm S**. **1979**. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**: 65–70.

**Kaus A, Michalski S, Hänfling B, Karthe D, Borchardt D & Durka W**. **2019**. Fish conservation in the land of steppe and sky: Evolutionarily significant units of threatened salmonid species in Mongolia mirror major

river basins. *Ecology and Evolution* **9**: 3416–3433.

**Kitada S**. **2022**. Long-term translocation explains population genetic structure of a recreationally fished iconic species in Japan: Combining current knowledge with reanalysis. *Aquaculture, Fish and Fisheries* **2**: 130–145.

**Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA & Mayrose I**. **2015**. Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources* **15**: 1179–1191.

**Launey S, Morin J, Minery S & Laroche J**. **2006**. Microsatellite genetic variation reveals extensive introgression between wild and introduced stocks, and a new evolutionary unit in French pike *Esox lucius* L. *Journal of fish biology* **68**: 193–216.

**Li YL & Liu JX**. **2018**. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* **18**: 176–177.

**Lopes-Cunha M, Aboim MA, Mesquita N, Alves MJ, Doadrio I & Coelho MM**. **2012**. Population genetic structure in the Iberian cyprinid fish *Iberochondrostoma lemmingii* (Steindachner, 1866): Disentangling species fragmentation and colonization processes. *Biological Journal of the Linnean Society* **105**: 559–572.

**MacGuigan DJ, Orr OD & Near TJ**. **2023**. Phylogeography, hybridization, and species discovery in the *Etheostoma nigrum* complex (Percidae: Etheostoma: Boleosoma). *Molecular Phylogenetics and Evolution* **178**: 107645.

**Marchetto F, Zaccara S, Muenzel FM & Salzburger W**. **2010**. Phylogeography of the Italian vairone (*Telestes muticellus*, Bonaparte 1837) inferred by microsatellite markers: evolutionary history of a freshwater fish species with a restricted and fragmented distribution. *BMC Evolutionary Biology* **10**: 111.

**Meraner A, Venturi A, Ficetola GF, Rossi S, Candiotto A & Gandolfi A**. **2013**. Massive invasion of exotic *Barbus barbus* and introgressive hybridization with endemic *Barbus plebejus* in Northern Italy: where, how and why? *Molecular Ecology* **22**: 5295–5312.

**Miller-sims VC, Gerlach G, Kingsford MJ & Atema J**. **2008**. Dispersal in the spiny damselfish, *Acanthochromis polyacanthus* , a coral reef fish species without a larval pelagic stage. *Molecular Ecology* **17**: 5036–5048.

**Moritz C**. **1994**. Defining evolutionarily significant units for conservation. *Trends in ecology & evolution* **9**: 373–375.

**Nguyen TTT & Sunnucks P**. **2012**. Strong population genetic structure and its management implications in the mud carp *Cirrhinus molitorella*, an indigenous freshwater species subject to an aquaculture and culture-based fishery. *Journal of Fish Biology* **80**: 651–668.

**Peakall ROD & Smouse PE**. **2006**. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology notes* **6**: 288–295.

**Perea S, Sousa-Santos C, Robalo J & Doadrio I**. **2021**. Historical biogeography of the Iberian Peninsula: multilocus phylogeny and ancestral area reconstruction for the freshwater fish genus *Squalius* (Actinopterygii, Leuciscidae). *Journal of Zoological Systematics and Evolutionary Research* **59**: 858–886.

**Pérez-Rodríguez R, Esquivel-Bobadilla S, Orozco-Ruíz AM, Olivas-Hernández JL & García-De León FJ**. **2021**. Genetic structure and historical and contemporary gene flow of *Astyanax mexicanus* in the Gulf of Mexico slope: a microsatellite-based analysis. *PeerJ* **9**: e10784.

**Petrosino G, Tancioni L, Turani M, Rakaj A, Ciuffardi L & Rossi AR**. **2022**. Phylogeography of *Sarmarutilus rubilio* (Cypriniformes: Leuciscidae): Complex Genetic Structure, Clues to a New Cryptic Species and Further Insights into Roaches Phylogeny. *Genes* **13**: 1071.

**Petrosino G, Rossi AR, Tancioni L, Gallozzi F & Colangelo P**. **2023**. Phenotypic plasticity over genetic diversity: ecomorphological patterns revealed in the eurytopic and threatened Italian endemic freshwater fish Sarmarutilus rubilio (Bonaparte, 1837). *submitted to Biological Journal of the Linnean Society*.

**Piry S, Luikart G & Cornuet JM**. **1999**. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90**: 502–503.

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

**Puechmaille SJ**. **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.

**R Core Team**. **2020**. R: A language and environment for statistical computing.

**Rondinini C, Battistoni A & Teofili C (Eds.)**. **2022**. *Lista rossa IUCN dei vertebrati italiani 2022*. Roma: Comitato Italiano IUCN e Ministero dell'Ambiente e della Sicurezza Energetica.

**Rossi AR, Petrosino G, Crescenzo S, Milana V, Talarico L, Martinoli M, Rakaj A, Lorenzoni M, Carosi A, Ciuffardi L & Tancioni L**. **2021**. Phylogeography and population structure of *Squalius lucumonis*: A baseline for conservation of an Italian endangered freshwater fish. *Journal for Nature Conservation* **64**: 126085.

**Rossi AR, Talarico L, Petrosino G, Crescenzo S & Tancioni L**. **2022**. Conservation Genetics of Mediterranean Brown Trout in Central Italy (Latium): A Multi-Marker Approach. *Water* **14**: 937.

**Rousset F**. **2008**. A complete re-implementation of the GENEPOP software for software for teaching and research. *Molecular Ecolology Resources* **8**: 103–106.

**Sala-bozano M, Ketmaier V & Mariani S**. **2009**. Contrasting signals from multiple markers illuminate population connectivity in a marine fish. *Molecular Ecology* **18**: 4811–4826.

**Schönhuth S, Vukić J, Šanda R, Yang L & Mayden RL**. **2018**. Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). *Molecular Phylogenetics and Evolution* **127**: 781–799.

**Seifertová M, Bryja J, Vyskočilová M, Martínková N & Šimková A**. **2012**. Multiple Pleistocene refugia and post-glacial colonization in the European chub (*Squalius cephalus*) revealed by combined use of nuclear and mitochondrial markers. *Journal of Biogeography* **39**: 1024–1040.

**Sousa-Santos C, Gil F & Almada VC**. **2014**. Ex situ reproduction of Portuguese endangered cyprinids in the context of their conservation. *Ichthyological Research* **61**: 193–198.

**Tancioni L, Russo T, Cataudella S, Milana V, Hett AK, Corsi E & Rossi AR. 2013.** Testing Species Delimitations in Four Italian Sympatric Leuciscine Fishes in the Tiber River: A Combined Morphological and Molecular Approach. *PLoS ONE* **8**: e60392.

**Valić D, Vardić Smrzlić I, Kapetanović D, Teskeredžić Z, Pleše B &Teskeredžić E. 2013.** Identification, Phylogenetic Relationships and a New Maximum Size of Two Rudd Populations (*Scardinius*, Cyprinidae) from the Adriatic Sea Drainage, Croatia. *Biologia* **68**: 539–545.

**Wang J, Zhang W, Wu J, Li C, Ju YM, Lin HD & Zhao J**. **2022**. Multilocus Phylogeography and Population Genetic Analyses of *Opsariichthys hainanensis* Reveal Pleistocene Isolation Followed by High Gene Flow around the Gulf of Tonkin. *Genes* **13**: 1908.

**Wetjen M, Schmidt T, Schrimpf A & Schulz R**. **2020**. Genetic diversity and population structure of burbot *Lota lota* in Germany: Implications for conservation and management. *Fisheries Management and Ecology* **27**: 170–184.

**Won H, Jeon HB & Suk HY**. **2020**. Evidence of an ancient connectivity and biogeodispersal of a bitterling species, *Rhodeus notatus*, across the Korean Peninsula. *Scientific Reports* **10**: 1011.

## SUPPLEMENTARY MATERIALS

**Table S1** Name and characteristics of 13 microsatellite loci successfully amplified in *Sarmarutilus rubilio*. The 9 loci analysed in the present work are highlited in grey boxes.

| Locus | Repeat motif | Size range observed in *Sarmarutilus rubilio* | Sequence (5'→ 3') | Annealing temperature (°C) | Dye | Reference |
|---|---|---|---|---|---|---|
| Sluc5 | (TTC)10(TTA)5 | 227-239 | F: GAGAAAGAGAGACCAATCCATAGTT<br>R: CCAAGCAAGCATCAAACCTG | 57 | Hex | Gigliarelli et al., 2012 |
| Sluc13 | (CA)35 | 202-264 | F: CACCCAGGCAATAAACAAGG<br>R: GGGTTAAGGGTCGGTTTAGG | 57 | Fam | Gigliarelli et al., 2012 |
| BL1-T2 | (TATC)9 | 243-295 | F: TGTTTGCTCAGACAAAACTTTCA<br>R: ACGGGTCTCAGATGATGCTC | 60 | Atto550 | Dubut et al., 2010 |
| Ca1 | (CA)24 | 97-111 | F: AAGACGATGCTGGATGTTTAC<br>R: CTATAGCTTATCCCGGCAGTA | 60 | Hex | Dubut et al., 2010 |
| Ca3 | (TAGA)14 | 228-416 | F: GGACAGTGAGGGACGCAGAC<br>R: TCTAGCCCCCAAATTTTACGG | 60 | Fam | Dubut et al., 2010 |
| N7K4 | (TG)16 | 156-172 | F:ACGAGCATCAGTATCCAGAGACAC<br>R:CATGTTTCCACATCTGAGCTAAAA | 63 | Atto550 | Dubut et al., 2010 |
| LC27 | (CT)22(CACT)3(CT)2 | 146-154 | F: TCCAGTTCTTCCTTCCTAATT<br>R: GCGGAGGGAGAGTATGTCAA | 60 | Fam | Vyskočilová et al., 2007 |
| Lsou05 | (CA)17 | 176-238 | F:CTGAAGAAGACCCTGGTTCG<br>R:CCCACATCTGCTGACTCTGAC | 60 | Hex | Dubut et al., 2010 |
| LleA-150 | (GC)4(GT)22 | 176-246 | F:AAAGTGTAAATCCAGATGTTTAAGT<br>R:AAAGGATAATTTTCAGAGTAACGAG | 60 | Atto550 | Dubut et al., 2010 |
| Sluc11 | (CA)14 | 236-268 (preliminary data) | F: CACACTGGCACCTCTGACAT<br>R: AGCCCCGTCAACAAACTGT | 63 | Hex | Gigliarelli et al., 2012 |
| LleC-090 | (TC)15GG(TC)3 | 241-267 (preliminary data) | F:TCAGACACAACTAACCGACC<br>R:GGCGCTGTCCAGAACTGA | 57 | Fam | Dubut et al., 2010 |
| Rser10 | (GT)12 | 179-194 (preliminary data) | F:TGCGTAATCGTGAAGCGGTG HEX 39<br>R:GCCACTAAAGCGCAGAAGCC | 57 | Hex | Dubut et al., 2010 |
| CtoG-075 | (GC)4 | 206-222 (preliminary data) | F:TCATTTGGATAACAATCCATCATCAC<br>R:ACTATGTTAGCATCCACACC | 57 | Fam | Dubut et al., 2010 |

**Table S2.** PCR conditions (a) and reaction mix (b) applied for the amplification of microsatellite loci.

(a)

| | Temperature | Time | Number of cycles |
|---|---|---|---|
| **Denaturation** | 94°C | 3 minutes | 1 |
| **Denaturation** | 95°C | 30 seconds | 30 |
| **Annealing** | see Table S1 | 45 seconds | |
| **Extension** | 72°C | 1 minute | |
| **Extension** | 72°C | 5 minutes | 1 |

(b)

| Component | Stock solution concentration | Volume in a single raction tube (µl) |
|---|---|---|
| PCR-grade water | -- | 7.73 |
| NH4 Buffer | 10x | 1 |
| MgCl2 | 50 mM | 0.3 |
| dNTPs mix | 2,5 µM each dNTP | 0.2 |
| Primer forward | 100 µM | 0.1 |
| Primer reverse | 100 µM | 0.1 |
| BIOTAQTM DNA Polymerase (Bioline) | 5 $U\mu l^{-1}$ | 0.07 |
| DNA | 10-100 ng | 0.5 |
| | **Total** | 10 |

**Table S3.** Allele frequencies and number of individuals examined (N) in each population at the nine completed loci (see Table S1).

| Locus | Allele/n | VAR2 | TRO | FOR | ARR | TIB2 | TIB3 | SET | FON | SCR | GAR | NOC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sluc5** | **N** | 20 | 15 | 12 | 18 | 18 | 23 | 20 | 15 | 22 | 18 | 6 |
| | **227** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.033 | 0.000 | 0.000 | 0.000 |
| | **230** | 0.150 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.967 | 0.045 | 0.000 | 0.333 |
| | **233** | 0.375 | 0.867 | 0.917 | 0.944 | 0.667 | 0.848 | 0.000 | 0.000 | 0.591 | 0.583 | 0.667 |
| | **236** | 0.350 | 0.133 | 0.000 | 0.000 | 0.250 | 0.152 | 0.000 | 0.000 | 0.114 | 0.028 | 0.000 |
| | **239** | 0.100 | 0.000 | 0.083 | 0.056 | 0.083 | 0.000 | 0.000 | 0.000 | 0.250 | 0.389 | 0.000 |
| **Sluc13** | **N** | 20 | 15 | 12 | 17 | 18 | 23 | 20 | 15 | 22 | 18 | 6 |
| | **202** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.033 | 0.000 | 0.000 | 0.000 |
| | **206** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.075 | 0.100 | 0.068 | 0.139 | 0.000 |
| | **208** | 0.025 | 0.000 | 0.000 | 0.118 | 0.111 | 0.087 | 0.575 | 0.500 | 0.000 | 0.000 | 0.000 |
| | **210** | 0.100 | 0.000 | 0.000 | 0.029 | 0.056 | 0.022 | 0.000 | 0.000 | 0.068 | 0.111 | 0.083 |
| | **214** | 0.125 | 0.033 | 0.000 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.000 |
| | **216** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.045 | 0.222 | 0.000 |
| | **218** | 0.000 | 0.000 | 0.208 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.136 | 0.083 | 0.000 |
| | **220** | 0.125 | 0.467 | 0.250 | 0.500 | 0.389 | 0.543 | 0.350 | 0.367 | 0.477 | 0.250 | 0.500 |
| | **222** | 0.050 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **224** | 0.150 | 0.067 | 0.000 | 0.059 | 0.000 | 0.043 | 0.000 | 0.000 | 0.091 | 0.083 | 0.083 |
| | **226** | 0.050 | 0.100 | 0.000 | 0.000 | 0.056 | 0.109 | 0.000 | 0.000 | 0.000 | 0.000 | 0.083 |
| | **228** | 0.025 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **230** | 0.025 | 0.067 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.083 |
| | **232** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **234** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 |
| | **236** | 0.125 | 0.000 | 0.000 | 0.000 | 0.028 | 0.022 | 0.000 | 0.000 | 0.045 | 0.000 | 0.167 |
| | **238** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **240** | 0.050 | 0.067 | 0.167 | 0.118 | 0.028 | 0.087 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **242** | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **244** | 0.000 | 0.000 | 0.000 | 0.088 | 0.000 | 0.000 | 0.000 | 0.000 | 0.023 | 0.000 | 0.000 |
| | **246** | 0.050 | 0.033 | 0.000 | 0.029 | 0.139 | 0.000 | 0.000 | 0.000 | 0.045 | 0.000 | 0.000 |
| | **252** | 0.000 | 0.167 | 0.167 | 0.059 | 0.000 | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **254** | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **258** | 0.050 | 0.000 | 0.042 | 0.000 | 0.000 | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **260** | 0.000 | 0.000 | 0.125 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **264** | 0.025 | 0.000 | 0.042 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **BL1-T2** | **N** | 20 | 15 | 12 | 17 | 18 | 23 | 20 | 14 | 22 | 18 | 6 |
| | **243** | 0.775 | 0.667 | 0.417 | 0.765 | 0.611 | 0.848 | 0.000 | 0.000 | 0.750 | 0.500 | 0.667 |
| | **247** | 0.125 | 0.000 | 0.125 | 0.176 | 0.056 | 0.022 | 0.275 | 0.250 | 0.023 | 0.000 | 0.000 |
| | **251** | 0.025 | 0.033 | 0.375 | 0.000 | 0.111 | 0.022 | 0.475 | 0.464 | 0.159 | 0.194 | 0.167 |
| | **255** | 0.000 | 0.133 | 0.083 | 0.000 | 0.028 | 0.043 | 0.100 | 0.179 | 0.000 | 0.028 | 0.167 |
| | **259** | 0.000 | 0.000 | 0.000 | 0.000 | 0.111 | 0.065 | 0.000 | 0.000 | 0.023 | 0.028 | 0.000 |
| | **263** | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 |
| | **267** | 0.000 | 0.067 | 0.000 | 0.059 | 0.000 | 0.000 | 0.000 | 0.036 | 0.045 | 0.167 | 0.000 |
| | **271** | 0.050 | 0.100 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **279** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 | 0.071 | 0.000 | 0.000 | 0.000 |
| | **283** | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **287** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | **295** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.000 |
| **Ca1** | **N** | 20 | 14 | 12 | 18 | 18 | 24 | 20 | 15 | 22 | 18 | 6 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **97** | 0.700 | 0.500 | 0.208 | 0.667 | 0.667 | 0.708 | 0.000 | 0.000 | 0.568 | 0.722 | 0.667 |
| | | **99** | 0.150 | 0.393 | 0.708 | 0.139 | 0.083 | 0.271 | 0.000 | 0.000 | 0.273 | 0.250 | 0.167 |
| | | **101** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.675 | 0.667 | 0.159 | 0.028 | 0.083 |
| | | **103** | 0.000 | 0.107 | 0.000 | 0.194 | 0.194 | 0.021 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **105** | 0.125 | 0.000 | 0.083 | 0.000 | 0.028 | 0.000 | 0.025 | 0.033 | 0.000 | 0.000 | 0.000 |
| | | **107** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.300 | 0.300 | 0.000 | 0.000 | 0.000 |
| | | **111** | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.083 |
| **Ca3** | **N** | | 17 | 8 | 13 | 17 | 17 | 18 | 17 | 14 | 19 | 16 | 0 |
| | | **228** | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.036 | 0.079 | 0.094 | 0.000 |
| | | **232** | 0.029 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 0.059 | 0.036 | 0.000 | 0.000 | 0.000 |
| | | **236** | 0.000 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 0.118 | 0.143 | 0.000 | 0.000 | 0.000 |
| | | **240** | 0.059 | 0.000 | 0.000 | 0.000 | 0.118 | 0.000 | 0.118 | 0.393 | 0.000 | 0.000 | 0.000 |
| | | **244** | 0.000 | 0.000 | 0.038 | 0.000 | 0.029 | 0.000 | 0.235 | 0.107 | 0.000 | 0.000 | 0.000 |
| | | **248** | 0.029 | 0.000 | 0.077 | 0.000 | 0.000 | 0.083 | 0.088 | 0.036 | 0.000 | 0.000 | 0.000 |
| | | **252** | 0.029 | 0.000 | 0.077 | 0.000 | 0.088 | 0.056 | 0.000 | 0.036 | 0.026 | 0.000 | 0.000 |
| | | **256** | 0.088 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 0.118 | 0.071 | 0.105 | 0.031 | 0.000 |
| | | **260** | 0.029 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 0.118 | 0.071 | 0.000 | 0.000 | 0.000 |
| | | **264** | 0.029 | 0.000 | 0.115 | 0.147 | 0.000 | 0.056 | 0.088 | 0.000 | 0.000 | 0.031 | 0.000 |
| | | **268** | 0.000 | 0.000 | 0.038 | 0.029 | 0.029 | 0.000 | 0.000 | 0.000 | 0.211 | 0.031 | 0.000 |
| | | **272** | 0.029 | 0.000 | 0.000 | 0.088 | 0.029 | 0.028 | 0.000 | 0.036 | 0.026 | 0.063 | 0.000 |
| | | **276** | 0.029 | 0.125 | 0.192 | 0.000 | 0.029 | 0.139 | 0.059 | 0.036 | 0.079 | 0.219 | 0.000 |
| | | **280** | 0.029 | 0.125 | 0.000 | 0.000 | 0.088 | 0.083 | 0.000 | 0.000 | 0.158 | 0.000 | 0.000 |
| | | **284** | 0.088 | 0.000 | 0.000 | 0.118 | 0.000 | 0.000 | 0.000 | 0.000 | 0.026 | 0.031 | 0.000 |
| | | **288** | 0.029 | 0.063 | 0.000 | 0.029 | 0.000 | 0.194 | 0.000 | 0.000 | 0.079 | 0.094 | 0.000 |
| | | **292** | 0.000 | 0.250 | 0.000 | 0.000 | 0.088 | 0.083 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **296** | 0.029 | 0.000 | 0.000 | 0.029 | 0.059 | 0.028 | 0.000 | 0.000 | 0.000 | 0.031 | 0.000 |
| | | **300** | 0.029 | 0.063 | 0.038 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.079 | 0.000 | 0.000 |
| | | **304** | 0.029 | 0.063 | 0.154 | 0.088 | 0.147 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **308** | 0.059 | 0.063 | 0.000 | 0.029 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **312** | 0.029 | 0.063 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.026 | 0.313 | 0.000 |
| | | **316** | 0.088 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **320** | 0.000 | 0.063 | 0.038 | 0.000 | 0.059 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **324** | 0.059 | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **328** | 0.029 | 0.000 | 0.077 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **332** | 0.059 | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **336** | 0.029 | 0.000 | 0.000 | 0.088 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **340** | 0.029 | 0.063 | 0.000 | 0.029 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **344** | 0.000 | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **348** | 0.000 | 0.000 | 0.000 | 0.206 | 0.029 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **352** | 0.000 | 0.000 | 0.000 | 0.059 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.063 | 0.000 |
| | | **360** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.053 | 0.000 | 0.000 |
| | | **364** | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **368** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.053 | 0.000 | 0.000 |
| | | **380** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **384** | 0.000 | 0.000 | 0.000 | 0.000 | 0.059 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **388** | 0.000 | 0.063 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **416** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **N7K4** | **N** | | 17 | 8 | 14 | 17 | 17 | 18 | 19 | 14 | 20 | 17 | 0 |
| | | **156** | 0.029 | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 | 0.000 | 0.025 | 0.059 | 0.000 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **158** | 0.059 | 0.000 | 0.000 | 0.000 | 0.059 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **160** | 0.324 | 0.438 | 0.393 | 0.412 | 0.441 | 0.417 | 0.000 | 0.143 | 0.900 | 0.853 | 0.000 |
| | | **162** | 0.235 | 0.250 | 0.464 | 0.471 | 0.206 | 0.139 | 0.289 | 0.250 | 0.025 | 0.000 | 0.000 |
| | | **164** | 0.235 | 0.313 | 0.143 | 0.059 | 0.235 | 0.417 | 0.132 | 0.107 | 0.050 | 0.059 | 0.000 |
| | | **166** | 0.088 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.553 | 0.429 | 0.000 | 0.000 | 0.000 |
| | | **168** | 0.029 | 0.000 | 0.000 | 0.000 | 0.029 | 0.000 | 0.026 | 0.036 | 0.000 | 0.029 | 0.000 |
| | | **172** | 0.000 | 0.000 | 0.000 | 0.059 | 0.000 | 0.000 | 0.000 | 0.036 | 0.000 | 0.000 | 0.000 |
| **LC27** | **N** | | 20 | 11 | 15 | 19 | 13 | 21 | 20 | 14 | 17 | 18 | 6 |
| | | **146** | 0.750 | 0.682 | 0.867 | 0.684 | 0.731 | 0.738 | 1.000 | 0.893 | 0.676 | 0.556 | 0.417 |
| | | **148** | 0.100 | 0.000 | 0.000 | 0.000 | 0.038 | 0.048 | 0.000 | 0.071 | 0.029 | 0.139 | 0.167 |
| | | **152** | 0.100 | 0.318 | 0.133 | 0.316 | 0.231 | 0.214 | 0.000 | 0.036 | 0.294 | 0.306 | 0.417 |
| | | **154** | 0.050 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Lsou05** | **N** | | 20 | 11 | 15 | 19 | 13 | 20 | 20 | 13 | 16 | 17 | 6 |
| | | **176** | 0.125 | 0.045 | 0.067 | 0.500 | 0.192 | 0.125 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **180** | 0.150 | 0.364 | 0.533 | 0.184 | 0.385 | 0.375 | 0.100 | 0.077 | 0.250 | 0.471 | 0.167 |
| | | **184** | 0.125 | 0.091 | 0.033 | 0.132 | 0.192 | 0.025 | 0.000 | 0.000 | 0.125 | 0.118 | 0.167 |
| | | **186** | 0.100 | 0.000 | 0.033 | 0.132 | 0.000 | 0.225 | 0.000 | 0.000 | 0.281 | 0.118 | 0.000 |
| | | **188** | 0.375 | 0.091 | 0.033 | 0.000 | 0.038 | 0.150 | 0.075 | 0.115 | 0.000 | 0.000 | 0.000 |
| | | **190** | 0.075 | 0.364 | 0.200 | 0.000 | 0.192 | 0.075 | 0.100 | 0.077 | 0.031 | 0.147 | 0.500 |
| | | **192** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.025 | 0.000 | 0.038 | 0.031 | 0.118 | 0.000 |
| | | **194** | 0.050 | 0.045 | 0.100 | 0.053 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **196** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.000 |
| | | **198** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 | 0.346 | 0.000 | 0.000 | 0.083 |
| | | **200** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.075 | 0.038 | 0.000 | 0.000 | 0.000 |
| | | **202** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.038 | 0.000 | 0.029 | 0.000 |
| | | **204** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | 0.000 | 0.094 | 0.000 | 0.000 |
| | | **206** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 |
| | | **208** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | 0.038 | 0.031 | 0.000 | 0.000 |
| | | **210** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | 0.038 | 0.031 | 0.000 | 0.083 |
| | | **212** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | 0.000 | 0.094 | 0.000 | 0.000 |
| | | **214** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.000 |
| | | **238** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.031 | 0.000 | 0.000 |
| **LleA** | **N** | | 20 | 11 | 16 | 19 | 13 | 21 | 20 | 14 | 17 | 18 | 6 |
| | | **176** | 0.250 | 0.273 | 0.156 | 0.158 | 0.231 | 0.310 | 0.000 | 0.000 | 0.235 | 0.056 | 0.000 |
| | | **178** | 0.100 | 0.000 | 0.375 | 0.211 | 0.000 | 0.190 | 0.000 | 0.000 | 0.088 | 0.028 | 0.167 |
| | | **180** | 0.075 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **182** | 0.050 | 0.091 | 0.000 | 0.211 | 0.154 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **186** | 0.150 | 0.136 | 0.000 | 0.079 | 0.077 | 0.071 | 0.000 | 0.000 | 0.059 | 0.028 | 0.000 |
| | | **188** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **190** | 0.075 | 0.000 | 0.031 | 0.000 | 0.077 | 0.048 | 0.000 | 0.000 | 0.206 | 0.694 | 0.417 |
| | | **192** | 0.000 | 0.227 | 0.188 | 0.026 | 0.077 | 0.024 | 0.000 | 0.000 | 0.029 | 0.000 | 0.000 |
| | | **194** | 0.100 | 0.000 | 0.000 | 0.132 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.083 |
| | | **196** | 0.000 | 0.000 | 0.000 | 0.026 | 0.000 | 0.048 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **198** | 0.025 | 0.227 | 0.031 | 0.105 | 0.231 | 0.024 | 0.000 | 0.000 | 0.088 | 0.000 | 0.000 |
| | | **200** | 0.050 | 0.000 | 0.000 | 0.000 | 0.038 | 0.095 | 0.000 | 0.000 | 0.206 | 0.056 | 0.000 |
| | | **202** | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 | 0.048 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **204** | 0.025 | 0.000 | 0.000 | 0.026 | 0.000 | 0.119 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **206** | 0.050 | 0.000 | 0.031 | 0.000 | 0.000 | 0.024 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | **208** | 0.000 | 0.000 | 0.031 | 0.026 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.083 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **210** | 0.000 | 0.000 | 0.000 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **212** | 0.000 | 0.045 | 0.063 | 0.000 | 0.000 | 0.000 | 0.100 | 0.179 | 0.000 | 0.000 | 0.000 |
| **214** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.025 | 0.000 | 0.000 | 0.000 | 0.000 |
| **216** | 0.000 | 0.000 | 0.094 | 0.000 | 0.000 | 0.000 | 0.025 | 0.036 | 0.059 | 0.139 | 0.000 |
| **218** | 0.000 | 0.000 | 0.000 | 0.000 | 0.077 | 0.000 | 0.025 | 0.000 | 0.029 | 0.000 | 0.167 |
| **220** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.050 | 0.036 | 0.000 | 0.000 | 0.000 |
| **222** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.071 | 0.000 | 0.000 | 0.000 |
| **224** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.050 | 0.000 | 0.000 | 0.000 | 0.083 |
| **226** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.100 | 0.214 | 0.000 | 0.000 | 0.000 |
| **228** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.025 | 0.036 | 0.000 | 0.000 | 0.000 |
| **230** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.036 | 0.000 | 0.000 | 0.000 |
| **232** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.075 | 0.143 | 0.000 | 0.000 | 0.000 |
| **236** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | 0.071 | 0.000 | 0.000 | 0.000 |
| **238** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.100 | 0.036 | 0.000 | 0.000 | 0.000 |
| **240** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.075 | 0.036 | 0.000 | 0.000 | 0.000 |
| **242** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.175 | 0.036 | 0.000 | 0.000 | 0.000 |
| **244** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.050 | 0.036 | 0.000 | 0.000 | 0.000 |
| **246** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.036 | 0.000 | 0.000 | 0.000 |

**Table S4**. Hardy-Weinberg equilibrium *p*-tests per population and 9 microsatellite loci. Significant *p*-values ($p < 0.05$) after Holm correction are given in bold.

| | Sluc5 | Sluc13 | BL1-T2 | Ca1 | Ca3 | N7K4 | LC27 | Lsou05 | LleA-150 |
|---|---|---|---|---|---|---|---|---|---|
| **VAR2** | **0.00** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **TRO** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **FOR** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **ARR** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **TIB2** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **TIB3** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **SET** | NA | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | NA | 1.00 | 1.00 |
| **FON** | NA | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **SCR** | 0.07 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.59 |
| **GAR** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| **PAM** | 1.00 | 1.00 | 1.00 | 1.00 | NA | NA | 1.00 | 1.00 | 1.00 |

**Table S5**. Estimated frequencies of null alleles for 9 microsatellite loci and 11 populations according to Dempster's EM algorithm.

Frequencies higher than 10% are indicated in the grey boxes

| | Sluc5 | Sluc13 | BL1-T2 | Ca1 | Ca3 | N7K4 | LC27 | Lsou05 | LleA-150 |
|---|---|---|---|---|---|---|---|---|---|
| **VAR2** | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.02 |
| **TRO** | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **FOR** | 0.00 | 0.07 | 0.00 | 0.15 | 0.00 | 0.03 | 0.11 | 0.06 | 0.00 |
| **ARR** | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.02 | 0.01 | 0.00 | 0.00 |
| **TIB2** | 0.00 | 0.00 | 0.09 | 0.04 | 0.00 | 0.01 | 0.06 | 0.00 | 0.06 |
| **TIB3** | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 |
| **SET** | 0.00 | 0.01 | 0.06 | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 | 0.00 |
| **FON** | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **SCR** | 0.16 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.05 |
| **GAR** | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **PAM** | 0.00 | 0.00 | 0.00 | 0.00 | NA | NA | 0.00 | 0.00 | 0.00 |

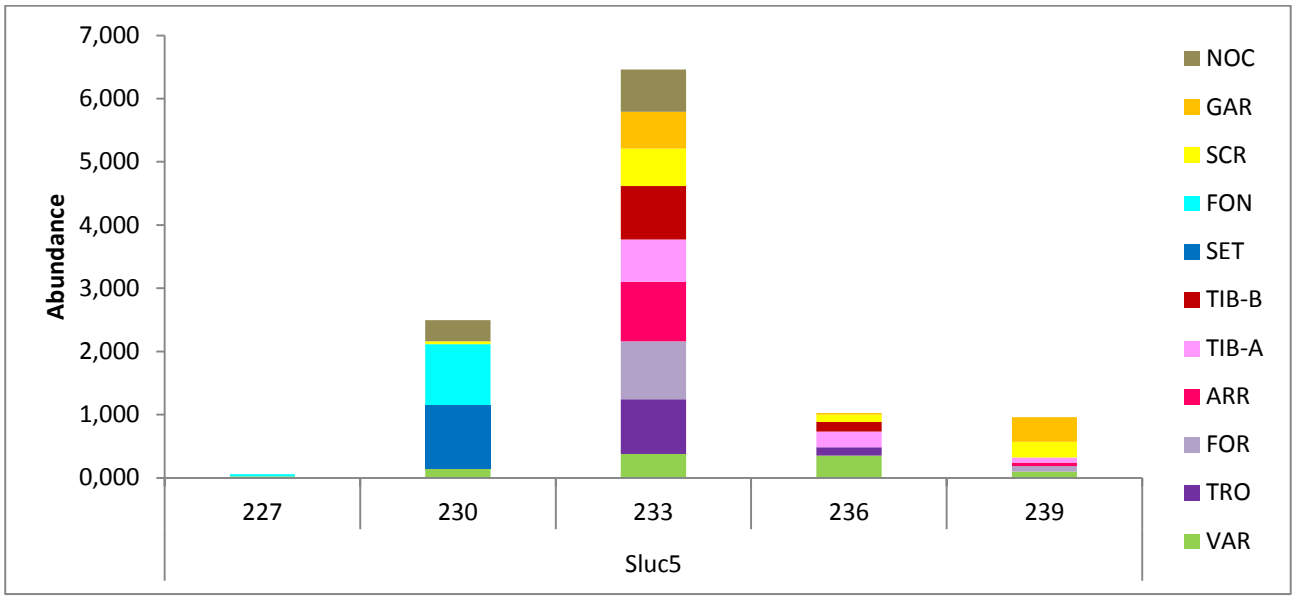
**Table S6.** $F_{ST}$s per locus and average across all samples; significance thresholds: ** = $p < 0.01$; *** = $p < 0.001$.

| | FST |
|---|---|
| **Sluc5** | 0.47438*** |
| **Sluc13** | 0.09633*** |
| **BL1-T2** | 0.20939*** |
| **Ca1** | 0.28879*** |
| **Ca3** | 0.05788*** |
| **N7K4** | 0.21394*** |
| **LC27** | 0.05802** |
| **Lsou05** | 0.10283*** |
| **LleA** | 0.11237*** |
| **Average** | 0.17134*** |

**Table S7:** Fixation index $\Phi_{ST}$ (CR) calculated using sequences from Petrosino *et al.*, 2022 and Petrosino *et al.*, 2023. Significance thresholds after Holm correction: * = P < 0.05; ** = P < 0.01; *** = P < 0.001.

| Haplogroup | Population | VAR2 | TRO | FOR | ARR | TIB2 | TIB3 | SET | FON | SCR | GAR | PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HpA/C | VAR2 | - | | | | | | | | | | |
| HpA | TRO | 0.27 | - | | | | | | | | | |
| HpA | FOR | 0.26 | 0.05 | - | | | | | | | | |
| HpA | ARR | 0.28* | 0.68*** | 0.64*** | - | | | | | | | |
| HpA | TIB2 | 0.20 | 0.07 | 0.09 | 0.53*** | - | | | | | | |
| HpA | TIB3 | 0.29* | 0.16 | 0.15 | 0.48*** | 0.00 | - | | | | | |
| HpB | SET | 0.56*** | 0.97*** | 0.96*** | 0.95*** | 0.95*** | 0.94*** | - | | | | |
| HpB | FON | 0.54*** | 0.98*** | 0.97*** | 0.96*** | 0.96*** | 0.95*** | -0.04 | - | | | |
| HpA/B | SCR | 0.29** | 0.18* | 0.17 | 0.36*** | 0.08 | 0.09 | 0.90*** | 0.90*** | - | | |
| HpA | GAR | 0.29* | 0.07 | 0.08 | 0.61*** | -0.04 | 0.04 | 0.96*** | 0.97*** | 0.11 | - | |
| HpA/B | NOC | 0.12 | 0.42*** | 0.38*** | 0.31** | 0.19 | 0.22 | 0.87*** | 0.87*** | 0.12 | 0.36** | - |

a)




b)

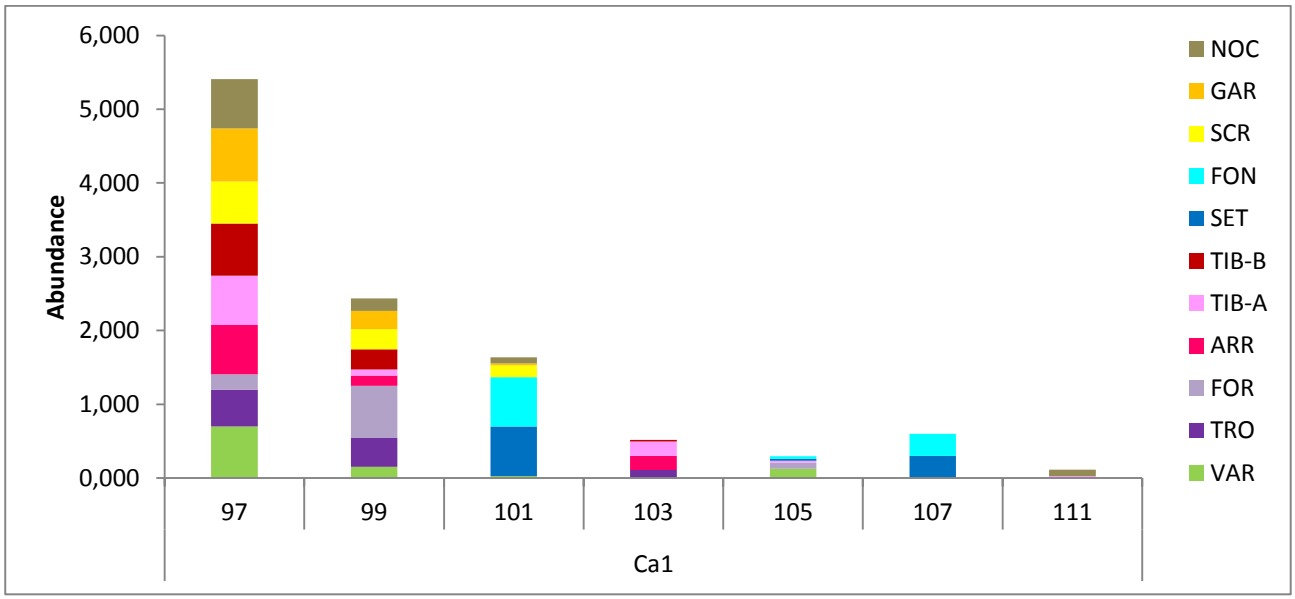



**Figure S1.** Graphical representation of relative abundances of alleles in each population at locus Sluc5 (a) and Ca1 (b).

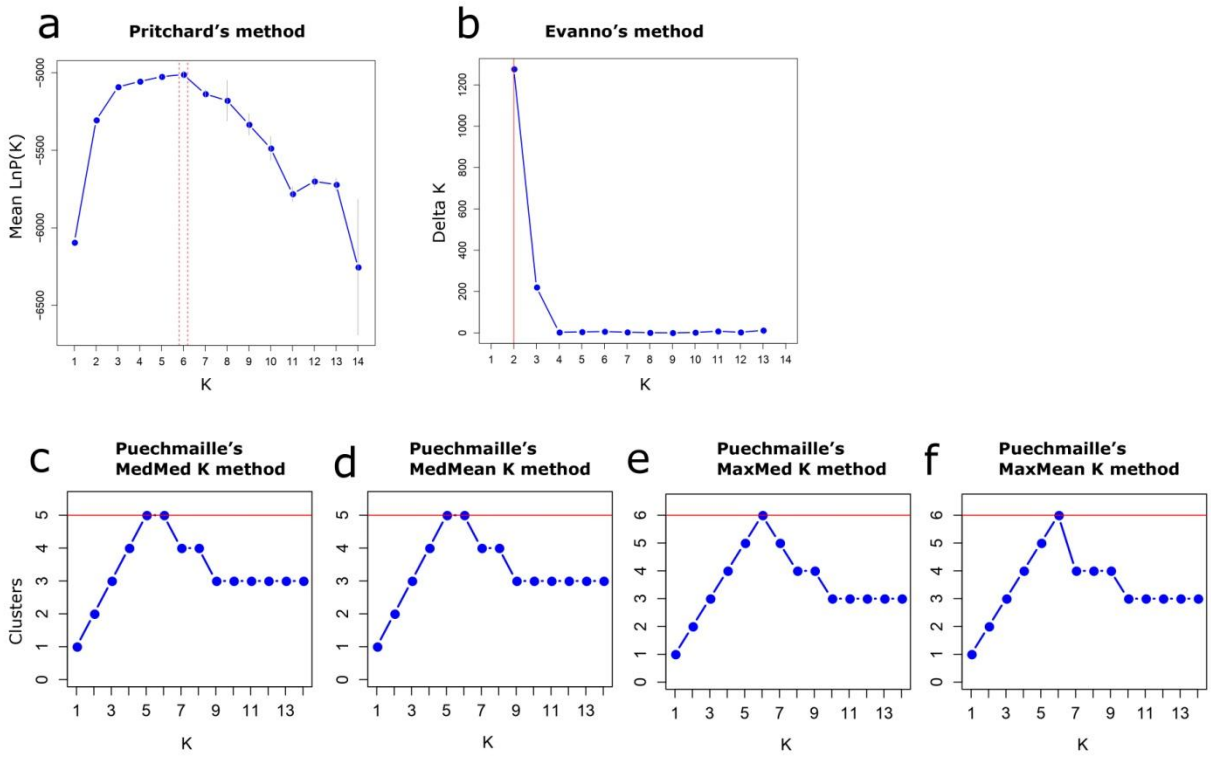



**Figure S2.** Support to different K clustering according to Pritchard's (a), Evanno's (b) and Puechmaille's (c-f) methods.

# GENERAL CONCLUSIONS

The genetic and morphological analyses reported in this thesis provided new valuable data about *Sarmarutilus rubilio* intraspecific diversity, i.e. genetic structure, phylogeography and ecomorphological variation. The combined use of multiple molecular markers, geometric morphometrics and environmental parameters allowed us to achieve the phD aims, test the different hypotheses on the drivers shaping species diversity and provide data useful for the management of this Italian endemic and threatened species.

## Phylogenetic and phylogeographic analysis

The mitochondrial markers (COI and CR) first revealed that the Albanian *S. rubilio*-like fishes are not strictly related to the genus *Sarmarutilus*, as they clustered instead with other Balkanian taxa belonging to the genus *Leucos*. This allowed us to exclude the presence of this Italian species in that geographic area. At the same time, all the Italian specimens visually identified as *S. rubilio* belonged to a monophyletic clade genetically distinct from the other species of roaches (*Rutilus* and *Leucos* genera). Intraspecific investigations, using also the nuclear marker Cyfun P, coupled with geological data of the sampled sites revealed that past events, i.e. hydrogeological evolution of basins and their isolation, caused the allopatric divergence of three different mitochondrial lineages within the South European roach, named A, B and C. The first split, originating lineage C, occurred at the northern border of this species range, the Magra-Vara basin, during the Early Pleistocene; the high divergence of this lineage suggested its status as a putative cryptic species. Subsequently, other geological events allowed the divergence between lineage A and B, in Central and South Italy respectively. The current distribution of mitochondrial haplotypes revealed that in more recent historical times, i.e. during the last glacial periods, lineage A extended its range and secondary contact events occurred at the northern and southern part of the overall species distribution. Lineage C, restricted to the Magra-Vara basin, is always present together with A, while in South Italy (Fondi plain, Southern Latium region) populations characterized by exclusive lineage B remained isolated. Native distribution of *S. rubilio* lineages is congruent with the

genetic-geographic structure of other Italian endemic freshwater fish species and contributed to a better definition of the borders between ichthyogeographic districts, which is still debated.

## Genetic and environmental contributions to morphological differences

Geometric morphometrics analysis confirmed in *S. rubilio* the existence of intraspecific morphological variations between populations. The main changes were observed in the overall body shape, i.e. more streamlined fishes vs more oval ones. On average, lineage B specimens showed a slightly deeper body than A ones; however, analysis of variance pointed out the little contribution of genetic differences in explaining the observed differences, which were mostly influenced by the sampling site (population) factor, i.e. by differences in overall environmental characteristics. Indeed, streamlined body shapes were observed in sites scarcely altered by human intervention and with fast water flow, and on the opposite deeper body shapes were observed in canals and one reservoir with slow/still water flow. In freshwater fishes streamlined body shapes are generally suited for swimming in high water flow, as they minimize drag, while deep body shapes are conversely fitted for complicated locomotor patterns where water flow is slower; thus, changes in shapes likely occurred in *S. rubilio* to optimize swimming in different freshwater environments. The observed differences are therefore the effect of phenotypic plasticity in response to different hydro-morphological features and hydrodynamic patterns. Finally, the existence of specific lineage-linked morphological characteristics that may be the result of allopatric genetic drift rather than natural selection cannot be excluded.

## Population structure and gene flow

To deepen the outcomes of the mitochondrial markers investigation, I used nine microsatellite loci, which are better suited for providing insights about recent isolation and contemporary gene flow between populations. Magra-Vara population is differentiated from any other, and mitochondrial lineage A and C are completely admixed, so it’s not possible to identify allele assemblages diagnostic for each of them; coupling this result with the lack of morphological

differences observed in Chapter 1 and 2, lineage C can not be identified as a cryptic species. As no specimens with a genotype identical to that of other populations were found in this area, this allowed us to exclude the translocation of fishes in this site in the last centuries, as the cause of the coexistence of two highly divergent lineages.

Populations from the Fondi plain (where only lineage B is present), are still isolated and characterized by high frequencies of typical alleles, absent in populations from Central Italy characterized by lineage A. These populations are highly differentiated from any other, likely due to strong genetic drift and the. South Italy populations, in sites where the coexistence of A and B lineages was recorded, clustered together and their distinct alleles assemblage testify the result of the admixing between different lineages and their subsequent independent evolution in this area (like in Magra-Vara). In the remaining population from Central Italy, where only lineage A was found, microsatellites revealed further substructures for populations between different yet adjacent basins, not detected by the previous mitochondrial analysis. A noteworthy lack of differentiation was observed within the Tiber basin and between the Tiber, on the Tyrrhenian slope, and the Tronto River, on the Adriatic one. Currently, the amplification and the genotyping of four new loci are still going on for the analyzed specimens and additional ones, and hopefully, they will clarify if current gene flow exists within basins and whether past geological events or human-mediated translocations allowed the South European roach to cross inter-basin barriers between slopes.

## Implications for conservation and management

Both mitochondrial and microsatellite molecular markers pointed out the genetic uniqueness of *S. rubilio* populations belonging to two different geographic areas: the Magra-Vara basin at the northern border of the species range, and the Fondi plain between Central and South Italy. Moreover, in Fondi plain there are two of the few lacustrine known populations for this species, which exhibit extremely deeper body shape not observed elsewhere. Due to their limited geographic extension, populations inhabiting these areas may be more susceptible to

anthropogenic threats such as habitat degradation, water pollution, abstraction and introduction of allochthonous species, and also be negatively affected by habitat alteration due to climate change. Therefore, they should be considered as distinct Management Units and prioritized in conservation actions required by the European Habitats Directive for *S. Rubilio*, e.g. habitat conservation and restoration; due to the observed genetic differences even between adjacent basins, in case *ex-situ* interventions (i.e. release of hatchery-reared juveniles) we suggest to use breeders collected from the same areas of intervention to preserve populations' genetic identity.

Finally, this thesis provided useful data, and protocols for new investigations about the genetic diversity and ecomorphological response not only for *S. rubilio* but also for other freshwater fish species. Indeed phylogeography of South European roach allowed us to better define the borders between Italian ichthyogeographic districts (e.g. TL and AC) and revealed further subdivisions within them (i.e. Magra-Vara basin within TL), suggesting the existence of similar genetic-geographic patterns in other freshwater taxa. I hope that my work will inspire others to further explore diversity among neglected freshwater fish species, as its acknowledgement is critical to counteracting their current decline.