

RESEARCH ARTICLE

Open Access



# Semi-permeable species boundaries in Iberian barbels (*Barbus* and *Luciobarbus*, Cyprinidae)

Hugo F Gante<sup>1,2,3\*</sup>, Ignacio Doadrio<sup>4</sup>, Maria Judite Alves<sup>2</sup> and Thomas E Dowling<sup>1,5</sup>

## Abstract

**Background:** The evolution of species boundaries and the relative impact of selection and gene flow on genomic divergence are best studied in populations and species pairs exhibiting various levels of divergence along the speciation continuum. We studied species boundaries in Iberian barbels, *Barbus* and *Luciobarbus*, a system of populations and species spanning a wide degree of genetic relatedness, as well as geographic distribution and range overlap. We jointly analyze multiple types of molecular markers and morphological traits to gain a comprehensive perspective on the nature of species boundaries in these cyprinid fishes.

**Results:** Intraspecific molecular and morphological differentiation is visible among many populations. Genomes of all sympatric species studied are porous to gene flow, even if they are not sister species. Compared to their allopatric counterparts, sympatric representatives of different species share alleles and show an increase in all measures of nucleotide polymorphism ( $S$ ,  $H_d$ ,  $K$ ,  $\pi$  and  $\theta$ ). High molecular diversity is particularly striking in *L. steindachneri* from the Tejo and Guadiana rivers, which co-varies with other sympatric species. Interestingly, different nuclear markers introgress across species boundaries at various levels, with distinct impacts on population trees. As such, some loci exhibit limited introgression and population trees resemble the presumed species tree, while alleles at other loci introgress more freely and population trees reflect geographic affinities and interspecific gene flow. Additionally, extent of introgression decreases with increasing genetic divergence in hybridizing species pairs.

**Conclusions:** We show that reproductive isolation in Iberian *Barbus* and *Luciobarbus* is not complete and species boundaries are semi-permeable to (some) gene flow, as different species (including non-sister) are exchanging genes in areas of sympatry. Our results support a speciation-with-gene-flow scenario with heterogeneous barriers to gene flow across the genome, strengthening with genetic divergence. This is consistent with observations coming from other systems and supports the notion that speciation is not instantaneous but a gradual process, during which different species are still able to exchange some genes, while selection prevents gene flow at other loci. We also provide evidence for a hybrid origin of a barbel ecotype, *L. steindachneri*, suggesting that ecology plays a key role in species coexistence and hybridization in Iberian barbels. This ecotype with intermediate, yet variable, molecular, morphological, trophic and ecological characteristics is the local product of introgressive hybridization of *L. comizo* with up to three different species (with *L. bocagei* in the Tejo, with *L. microcephalus* and *L. sclateri* in the Guadiana). In spite of the homogenizing effects of ongoing gene flow, species can still be discriminated using a combination of morphological and molecular markers. Iberian barbels are thus an ideal system for the study of species boundaries, since they span a wide range of genetic divergences, with diverse ecologies and degrees of sympatry.

**Keywords:** Speciation-with-gene-flow, Sympatry, Secondary contact, Introgression, Hybrid ecotype, *Barbus*, *Luciobarbus*

\* Correspondence: hugo.gante@asu.edu

<sup>1</sup>School of Life Sciences, Arizona State University, 85287-4601 Tempe, AZ, USA

<sup>2</sup>Museu Nacional de História Natural e da Ciência, Centre for Ecology, Evolution and Environmental Changes (Ce3C), Universidade de Lisboa, Rua da Escola Politécnica 58, 1250-102 Lisbon, Portugal

Full list of author information is available at the end of the article

## Background

The study of species boundaries provides invaluable information on the evolution of reproductive barriers and the impacts of gene flow on species divergence. This mechanistic approach is aimed at understanding how gene pools become subdivided and further gene flow is restricted or prevented. Gene exchange among taxa is likely to continue [1] and examples of divergence-with-gene-flow speciation [2] are now common in the literature. Speciation-with-gene-flow can range from divergence initiated in sympatry, to the evolution of additional isolating barriers after the establishment of secondary contact [3].

As taxa diverge, shared traits can reflect recent common ancestry and incomplete sorting of characters, reestablished gene exchange, or a combination of both. The two processes are difficult to separate, but the patterns they generate can potentially be discriminated using multiple independent markers and geographical information [4]. While traits sort and correlations among independent characters are generated as taxa diverge, gene flow breaks down those correlations. If introgressive hybridization is important, pre-existing correlations may be broken down by gene flow in sympatry but still persist in allopatry. Therefore, the signature of past isolation and divergence will be evident through the association among traits and geography. For instance, if species constitute monophyletic groups and only individuals in sympatry share alleles with other taxa, introgression rather than ancestral polymorphism is the likely cause of the pattern of allele sharing. Furthermore, allele sharing in areas of sympatry due to introgression is also expected to increase overall levels of molecular diversity, such as number of alleles and haplotype diversity, a pattern that is not likely to be generated by incomplete lineage sorting. Importantly, upon reestablishing some degree of gene flow the resulting patterns of variation will vary on a locus-by-locus basis, depending on impacts of selection and drift and on the architecture of the trait (e.g. [5–11]).

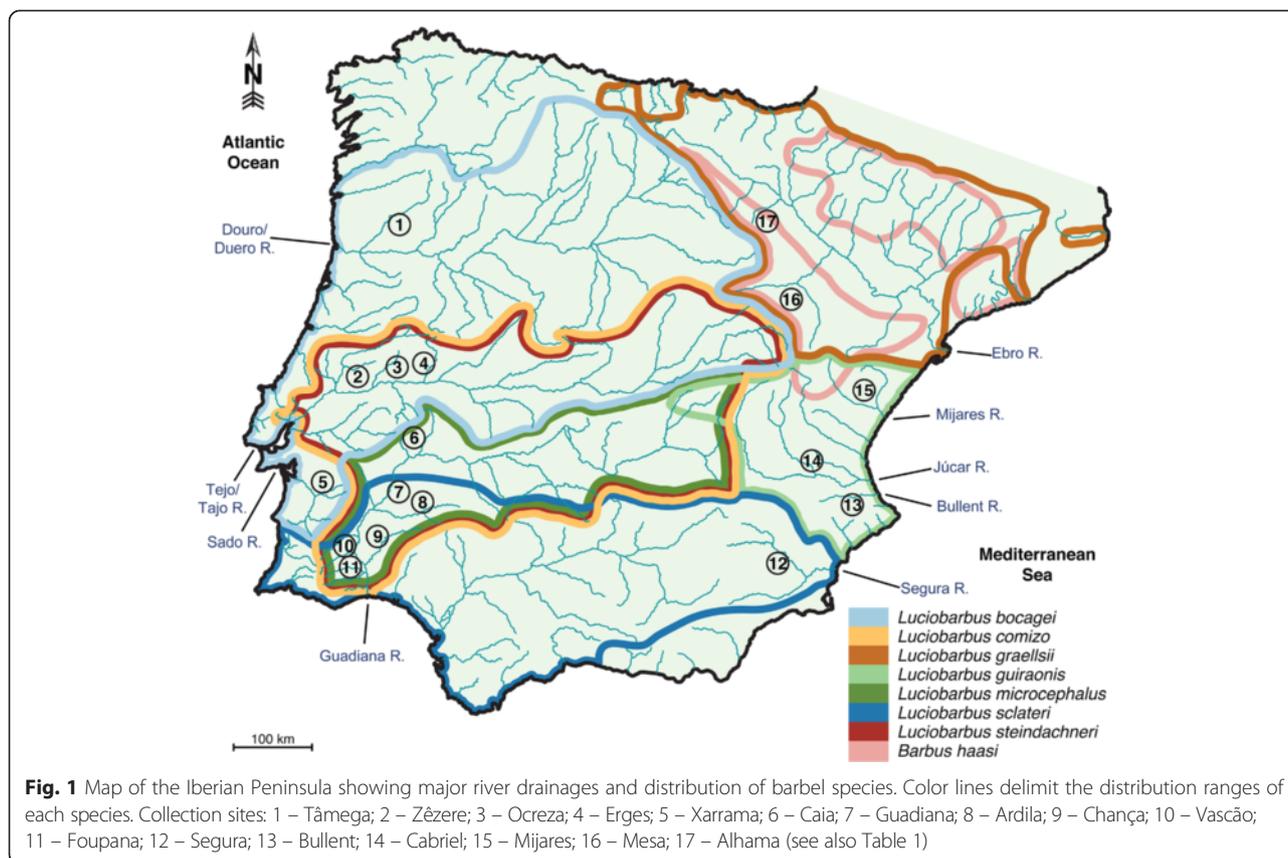
The relative impacts of homogenizing gene flow and disruptive selection on species boundaries (i.e. genomic divergence) is most easily studied in recently speciated groups that show some degree of geographic overlap, such as finches, butterflies and lizards (e.g. [12–17]). Among taxa that present these ideal characteristics are the speciose sister genera *Barbus* and *Luciobarbus*. This group of pseudotetraploid fishes started diversifying throughout fresh waters of the circum-Mediterranean region more than 20 ma (million years ago) [18–20]. In the Iberian Peninsula, *Barbus* is represented by two species, *B. meridionalis* and the endemic *B. haasi*, while *Luciobarbus* contains seven endemic species (*L. bocagei*, *L. comizo*, *L. graellsii*, *L. guiraonis*, *L. microcephalus*, *L. sclateri*, and *L. steindachneri*; [21–26]), which started diversifying around 8 ma [18–20]. Different species are

usually confined to specific ichthyogeographic provinces, inhabiting from one up to several river basins, and their ranges overlap to different extents (Fig. 1). This mostly allopatric distribution pattern indicates that speciation in Iberian barbels follows the evolution of river basins [23, 24, 26–28]. Secondary contacts in riverine species can be facilitated by several geomorphological processes affecting drainage patterns, such as river capture, marine regression and divide overtopping [29]. Allopatric diversification and range expansion with secondary contacts has been explicitly demonstrated for one polytypic species [30].

In diverse groups with varying degrees of range overlap such as barbels, species identification can be difficult. For instance, taxonomic status of *L. steindachneri* has been problematic due to its morphological variability and similarity to *L. comizo*, and has been considered a junior synonym of the latter by some authors (after [31]). Nevertheless, from an ecological standpoint it has been shown that *L. steindachneri* prefers habitats further upstream from those occupied by *L. comizo* and *L. microcephalus*, but lower than those inhabited by *L. sclateri* in the Guadiana River Basin [32–34]. *Luciobarbus steindachneri* also exhibits different trophic adaptations (e.g. intermediate mouth protrusion) and some degree of food partitioning relative to *L. bocagei* and *L. comizo* in the Tejo/Tajo River Basin [35–37] and *L. microcephalus* in the Guadiana River Basin [38]. Together, these data suggest that ecology plays a major role in species coexistence and has led some authors to instead regard *L. steindachneri* as an ecotype of *L. comizo* (e.g. [34]).

Despite coexistence, hybridization of Iberian barbels in areas of species overlap has been hypothesized based on specimens with intermediate phenotypes [21, 39–41] and mtDNA polyphyly [42], but inference of hybridization in these studies suffers from species identification problems [26] and circularity. In fact, hybridization in Iberian barbels has not been unequivocally demonstrated except in a few cases in which nuclear markers were examined [30, 43]. Unfortunately, most studies used a single type of character, precluding the examination of patterns of covariation among them. Furthermore, pseudotetraploidy of barbels limited these studies to the use of low-resolution nuclear markers, which in principle restricts their use to study more divergent species pairs. However, this limitation has been overcome with the development of paralog-specific primers for the amplification of nuclear loci in barbels [44].

Given that Iberian barbels have various degrees of divergence and geographic overlap, these species provide an excellent natural system to understand how reproductive barriers build up. We examine the morphological and genetic distinctiveness of all Iberian barbels, whether reproductive isolation is complete, and how patterns of gene flow vary among loci and species. To this end we use multiple



classes of characters (external morphological traits, mitochondrial and nuclear DNA sequence data) and their patterns of covariation on the entire constellation of endemic *Barbus* (one species) and *Luciobarbus* (seven species), sampled from sympatric and allopatric populations (Table 1).

## Results

### Multivariate analysis of morphological data

Multivariate analysis of meristic traits is effective in separating the Iberian endemic species of *Barbus* and *Luciobarbus*. The first two principal components explain 80.1 % of the observed variation, with all variables contributing to both axes, although traits that contribute the most to one axis contribute the least to the other (Table 2). Plotting the two components against each other provides clear visual separation of most of the recognized species (Fig. 2). *Barbus haasi* is the most morphologically distinct Iberian barbel, with all *Luciobarbus* analyzed being more similar to each other in meristic morphospace. The first component separates *L. comizo*, *L. bocagei* and *L. sclateri* from *L. microcephalus*, *L. guiraonis* and *L. graellsii*, and these groups from *B. haasi*, while the second component separates species within the first two groups of *Luciobarbus*. Therefore, species living in sympatry can be correctly discriminated using a few morphological traits, with the notable exception of *L. steindachneri*. Specimens of

*L. steindachneri* show substantial variation in meristic traits, some individuals overlapping with either *L. comizo* or *L. bocagei* in the Tejo River and with *L. comizo* or *L. sclateri* in the Guadiana River. As a whole, *L. steindachneri* occupies an intermediate position in the morphospace relative to the other Iberian *Luciobarbus*.

Intraspecific differences are also observable, including between populations of *L. comizo* from the Tejo and Guadiana rivers, between populations of *L. guiraonis* from the Bullent and Mijares rivers relative to that from the Júcar River, and between populations of *L. sclateri* from the Guadiana and Segura rivers (albeit to a lesser degree). Populations of *L. guiraonis* from the Bullent and Mijares rivers overlap with each other and with *L. microcephalus* for these specific meristic traits (numbers of scales and cephalic pores); however, these taxa are readily discriminated by other traits, such as position of the mouth, shape of the dorsal fin and robustness of the last simple dorsal ray (Fig. 2 and [22]).

### Mitochondrial phylogeny

Analysis of 151 individuals representing the eight endemic barbel species identified 14 different *cyt b* haplotypes (Table 3), as found in previous studies [26, 45]. Two haplotypes are typical of *L. comizo*, two of *L. bocagei*, two of *L. sclateri*, two of *L. guiraonis*, four of *L. microcephalus*, one

**Table 1** Sample sizes of *Barbus* and *Luciobarbus* analyzed across different traits

Species	Basin	River	Region	Morphology	mtDNA	nuDNA
<i>L. bocagei</i>	Douro	Tâmega	Iberia	10	10	10
	Tejo	Erges, Zêzere	Iberia	12	12	12
	Sado	Xarrama	Iberia	6	10	10
<i>L. comizo</i>	Tejo	Erges, Ocreza	Iberia	7	7	7
	Guadiana	Guadiana, Ardila, Vascão	Iberia	8	8	8
<i>L. graellsii</i>	Ebro	Mesa	Iberia	10	10	10
<i>L. guiraonis</i>	Júcar	Cabriel	Iberia	10	10	10
	Mijares	Mijares	Iberia	6	8	8
	Bullent	Bullent	Iberia	8	0	0
<i>L. microcephalus</i>	Guadiana	Guadiana, Ardila, Chança	Iberia	14	14	14
<i>L. sclateri</i>	Guadiana	Chança, Foupana,	Iberia	11	11	11
	Segura	Segura	Iberia	10	10	10
<i>L. steindachneri</i>	Tejo	Erges, Ocreza	Iberia	14	14	14
	Guadiana	Guadiana, Ardila, Caia, Chança, Foupana	Iberia	26	26	26
<i>B. haasi</i>	Ebro	Alhama	Iberia	8	1	1
<i>B. barbus</i>	Vistula	Propad	Slovakia	0	3	3
	–	–	France	0	2	2
<i>B. carpathicus</i>	Vistula	Propad	Slovakia	0	3	3
<i>B. prespensis</i>	Lake Prespa	Agios Germanos	Greece	0	3	3
<i>Total</i>				160	162	162

mtDNA: mitochondrial DNA; nuDNA: nuclear DNA

of *L. graellsii* and one of *B. haasi*. The Iberian *Luciobarbus* haplotypes form three main lineages, which are thought to reflect the true species tree. One lineage is composed of haplotypes found in the sister *L. comizo* and *L. bocagei*, which share a common ancestor with a second lineage, composed of the polytypic *L. sclateri*. The third lineage is composed of *L. graellsii* and the sister *L. microcephalus* and *L. guiraonis* (Fig. 3).

Cytochrome *b* haplotypes are readily associated with a morphological species, again with the exception of

individuals of *L. steindachneri* analyzed, which have mtDNA haplotypes typical of *L. comizo* or *L. sclateri*. In addition, a small number of specimens of several species possess mtDNA typical of other sympatric taxa, but never of allopatric taxa. This pattern is especially pronounced in *L. guiraonis* from the Júcar River where most specimens examined exhibit *B. haasi* mtDNA (Fig. 2).

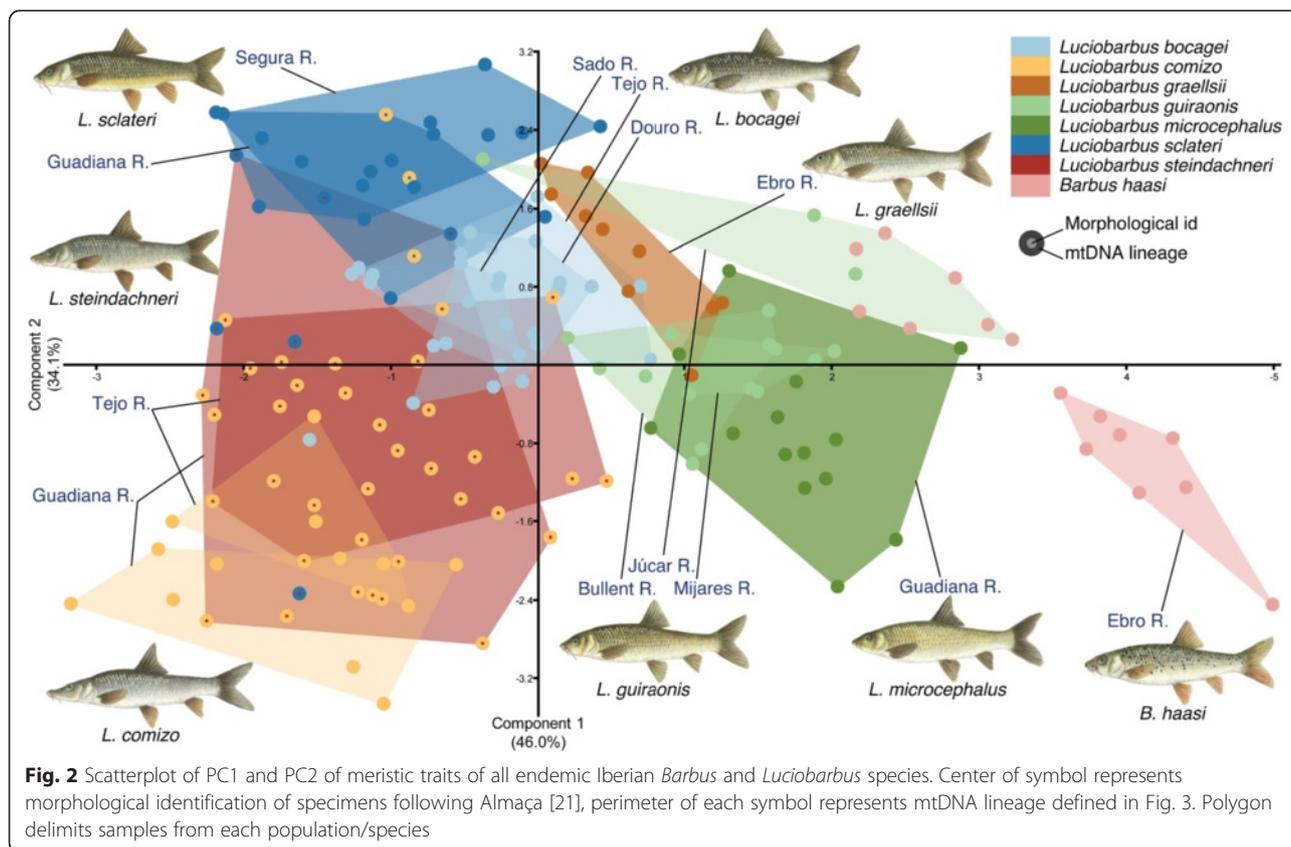
**Table 2** Principal components of meristic traits of all Iberian *Barbus* and *Luciobarbus* samples

Trait	PC1	PC2
LL	0.268	-0.544
TRA	0.332	-0.484
TRB	0.408	-0.375
IOC	-0.447	-0.370
SOC	-0.463	-0.311
POMC	-0.486	-0.313
Eigenvalue	2.76	2.05
Variance	46.0 %	34.1 %

Loadings of each original variable in components 1 and 2, eigenvalues and percentage of variance explained. LL: number of scales along the lateral line; TRA: number of scales in transverse rows above the LL; TRB: number of scales in transverse rows below the LL; SOC: supraorbital canal pores, IOC: infraorbital canal pores; POMC: preopercular-mandibular canal pores

#### Nuclear gene phylogenies

Sequence analysis of four nuclear loci yielded 3,792 aligned sites (*S7-1*, 813 bp; *S7-2*, 827 bp; *Gh-1*, 1018 bp; *Gh-2*, 1134 bp). The Iberian specimens analyzed exhibited 35, 54, 41 and 41 alleles for *S7-1*, *S7-2*, *Gh-1* and *Gh-2*, respectively. In general, gene trees are similar to each other (Figs. 4, 5, 6, 7) and are in agreement with relationships previously suggested by mtDNA, allozymes and morphology (e.g. [22–27, 30, 46, 47]). In particular, two main monophyletic lineages corresponding to the two genera, *Barbus* and *Luciobarbus*, are recovered. The former lineage is comprised of the Iberian endemic *B. haasi* and central and eastern European taxa included for phylogenetic context (*B. barbus*, *B. carpathicus* and *B. prespensis*). For *S7-1*, *S7-2* and *Gh-2*, the Iberian *Luciobarbus* lineage comprises two monophyletic groups, one composed of *L. graellsii*, *L. guiraonis* and *L. microcephalus*, and another composed of *L. bocagei*, *L. comizo*, *L. sclateri* and *L. steindachneri*. *Gh-1* is the least resolved nuclear



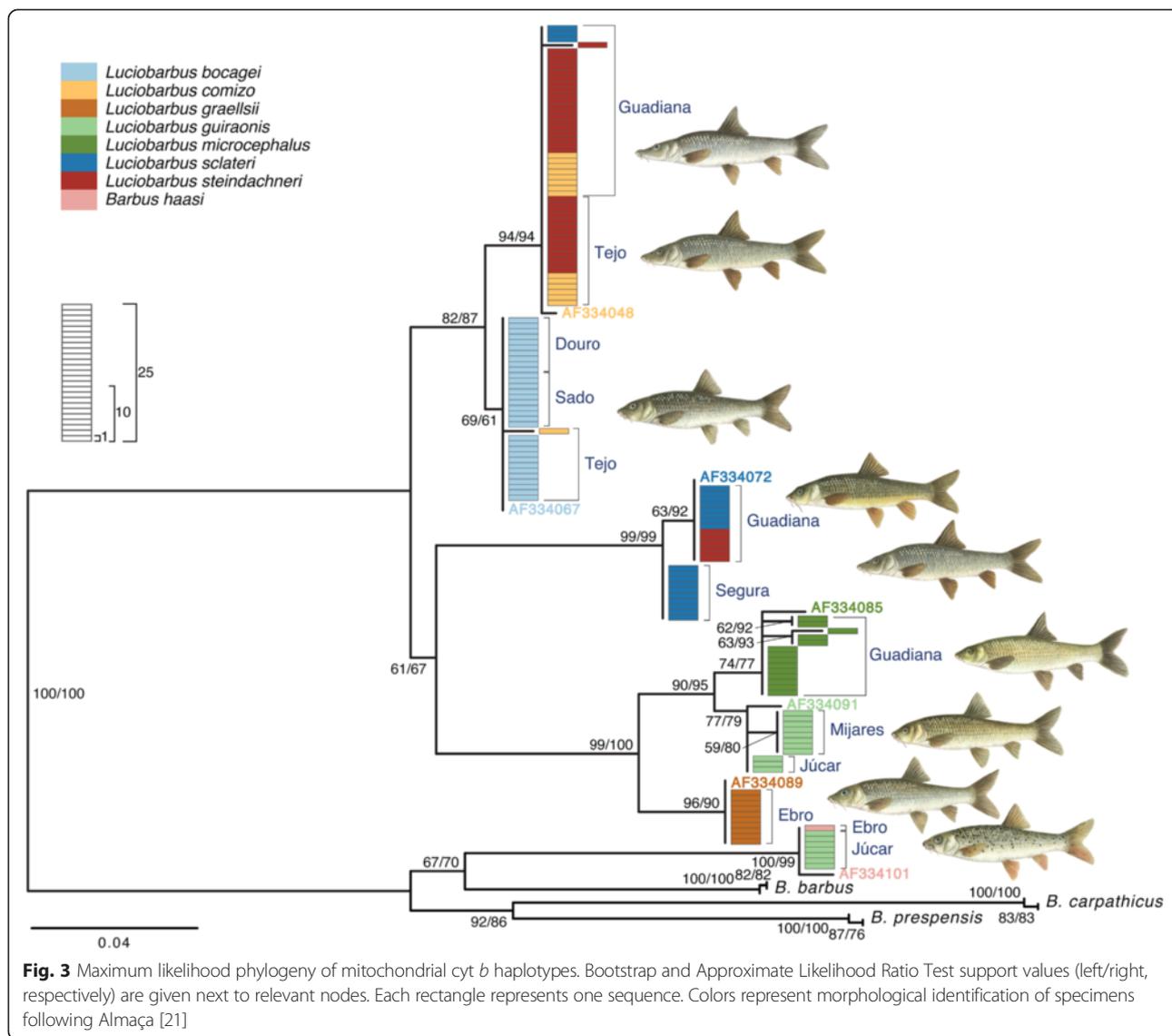
marker at the species level, revealing a monophyletic group composed of *L. graellsii* and *L. guiraonis*, while *L. microcephalus* is recovered in a weakly supported group together with other sympatric species. The Iberian *Luciobarbus* lineages are further diagnosed by particular insertion-deletion variants at *S7-1*, *S7-2* and *Gh-2*.

The relationships within the two Iberian *Luciobarbus* lineages are less resolved, with most species para- or polyphyletic for some or all nuclear loci. As in the case of mtDNA, several specimens possess alleles typical of other sympatric species (Tejo and Guadiana rivers), or from allopatric taxa found in adjacent basins (Sado River). The

**Table 3** Distribution of *cyt b* haplotypes across species and populations sampled

Species	Population	A	B	C	D	F	H	J	K	M	N	S	T	X	Z
<i>L. bocagei</i>	Douro		10												
<i>L. bocagei</i>	Tejo		12												
<i>L. bocagei</i>	Sado		10												
<i>L. comizo</i>	Tejo	1		6											
<i>L. comizo</i>	Guadiana			8											
<i>L. steindachneri</i>	Tejo			14											
<i>L. steindachneri</i>	Guadiana			19	1		6								
<i>L. microcephalus</i>	Guadiana							2	2	9	1				
<i>L. sclateri</i>	Guadiana			3			8								
<i>L. sclateri</i>	Segura					10									
<i>L. guiraonis</i>	Júcar											3			7
<i>L. guiraonis</i>	Mijares												8		
<i>L. graellsii</i>	Ebro													10	
<i>B. haasi</i>	Ebro														1

Mitochondrial haplotype codes follow [45]



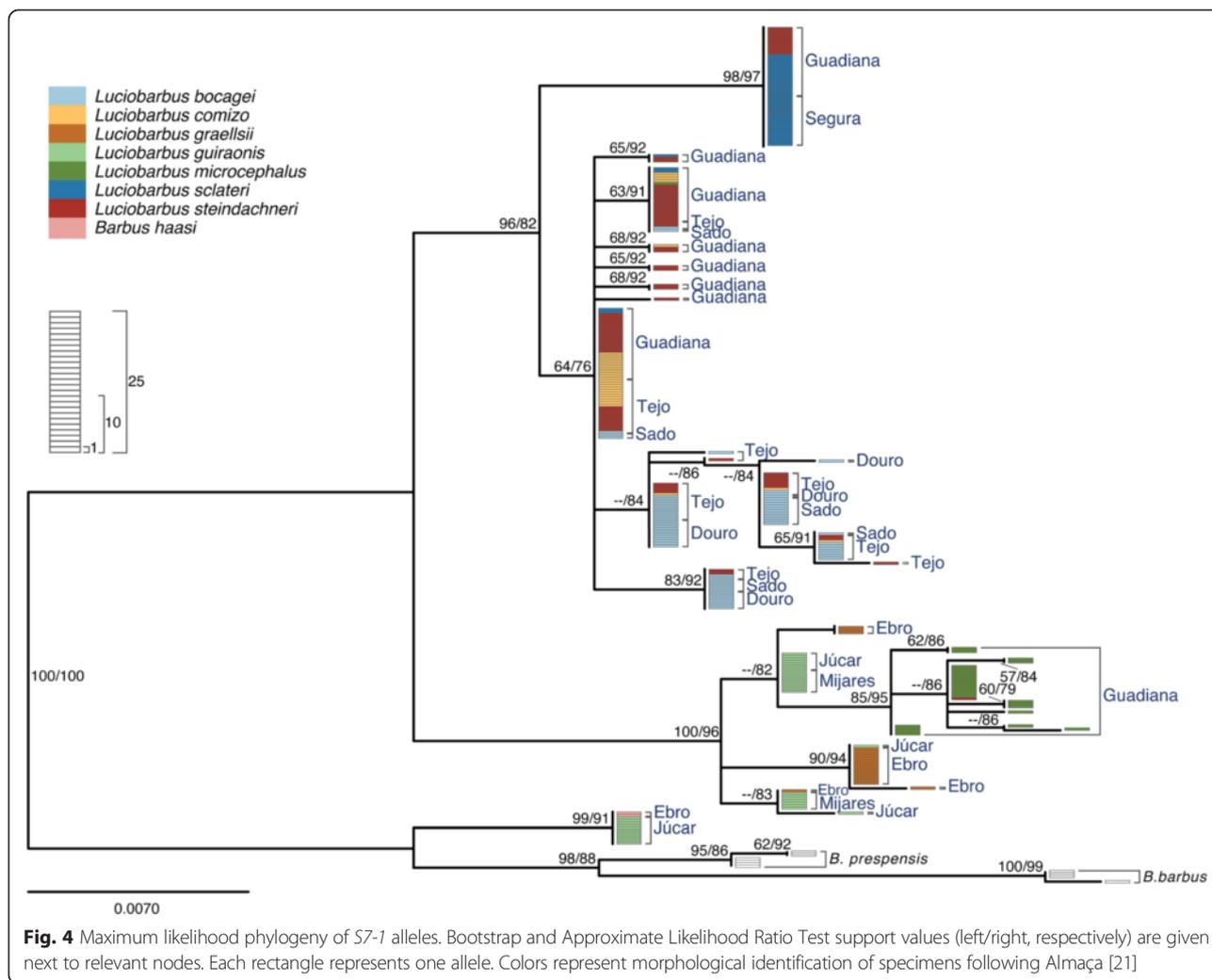
most obvious case is represented by the two divergent lineages present in *L. guiraois* from Júcar, one typical of *B. haasi* and the remaining also found in other populations of *L. guiraois*. Notably, *L. steindachneri* is not distinct from the other sympatric species as most of its alleles are also found in *L. comizo* and *L. bocagei* from the Tejo River, and in *L. comizo*, *L. microcephalus* and *L. sclateri* from the Guadiana River.

**Population-level nuclear variation and differentiation**

Levels of nuclear polymorphism vary substantially among populations, from five segregating sites (S) observed in *L. guiraois* from Mijares, to 91 segregating sites observed in *L. guiraois* from the Júcar River (Table 4). Nucleotide diversity ( $\pi$ ) also differs by an order of magnitude among these populations, but most species/populations show intermediate levels of polymorphism (Table 4). Interestingly,

allopatric populations of *Luciobarbus* exhibit lower levels of nucleotide polymorphism than sympatric populations (Mann–Whitney *U*-tests:  $z = 2.2736–2.7608$ , False Discovery Rate corrected  $P = 0.0072–0.0116$ , for all measures of polymorphism,  $S$ ,  $H_d$ ,  $K$ ,  $\pi$  and  $\theta$ ). In addition, both populations of *L. steindachneri* exhibit the highest numbers of segregating sites and levels of nucleotide diversity of sympatric species, in line with results from phylogenetic analyses. Additionally, discounting the effects of introgression, populations inhabiting smaller basins (e.g., Mijares, Sado and Segura) show lower levels of nucleotide polymorphism than those inhabiting larger river basins as Guadiana or Tejo, consistent with differences in effective population size.

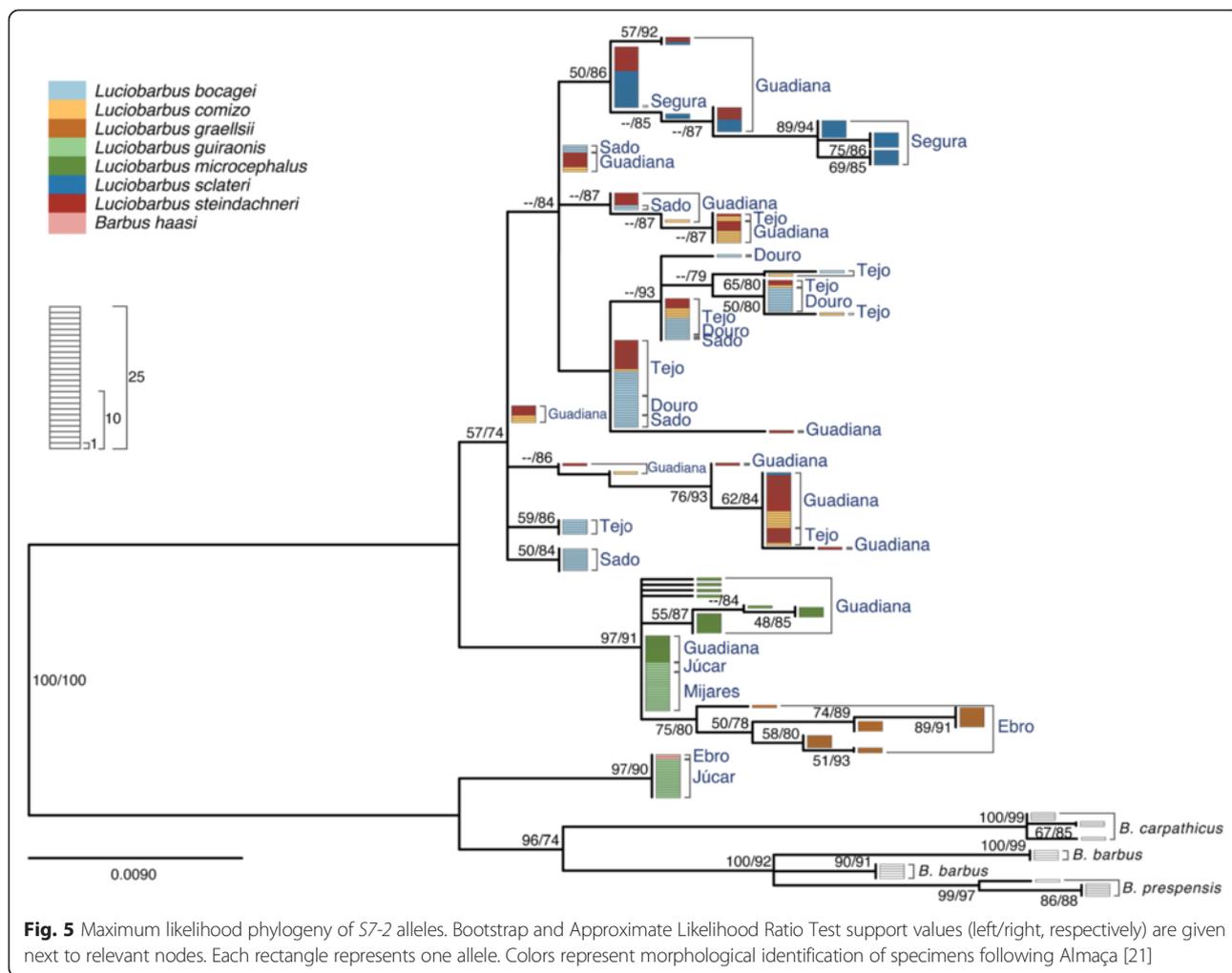
To test how polymorphisms across loci are structured among populations we generated estimates of  $D_{xy}$ ,  $D_a$  and  $F_{ST}$  (Additional file 1) and clustered them using the neighbor-joining method (Additional file 2). Population



trees built using estimates derived from mtDNA or all nuclear loci combined are generally consistent with phylogenetic results presented above. They identify two main groups, the genera *Barbus* and *Luciobarbus*. Within the latter, one subgroup is composed of *L. graellsii*, *L. guiraonis* and *L. microcephalus*, while the other contains *L. bocagei*, *L. comizo*, *L. sclateri* and *L. steindachneri*. Different populations of the same species are typically more similar to each other than to other species, with the exception of *L. steindachneri* and *L. guiraonis* (which reflect patterns of allele sharing detailed above). When nuclear loci are analyzed individually to investigate how nucleotide variation along different regions of the genome is shared across species and populations, different scenarios become clear, depending on the locus: *S7-1* and *Gh-2* follow the expected species phylogeny, while *S7-2* and *Gh-1* reflect geographical affinities. Thus, the latter genomic regions reflect geographic proximity and inter-specific gene flow, while the former reflect phylogenetic affinities.

#### Bayesian clustering of nuclear data

Due to the presence of large number of clusters (*K*) and high levels of differentiation among many populations and species, STRUCTURE sometimes converged to different solutions in independent replicates of each *K*, making determination of the best *K* challenging. Therefore, we consecutively split the complete dataset into smaller datasets, as recommended by the authors of the program as a strategy to deal with dataset multimodality. To determine the number of clusters we followed changes in LnP(D) values of consecutive *K* (i.e. when values plateau) and Evanno's *et al.*  $\Delta K$  [48]. We first split the dataset in two, based on geography and phylogenetic relationships. The first dataset is composed of all species from rivers draining to the Atlantic and allopatric populations of those species (in this case only *L. sclateri* from the Segura River). This analysis converged to *K* = 5 genetic clusters (Fig. 8, upper left), clearly separating populations inhabiting rivers draining to the western margin of the Iberian Peninsula (Douro, Tejo



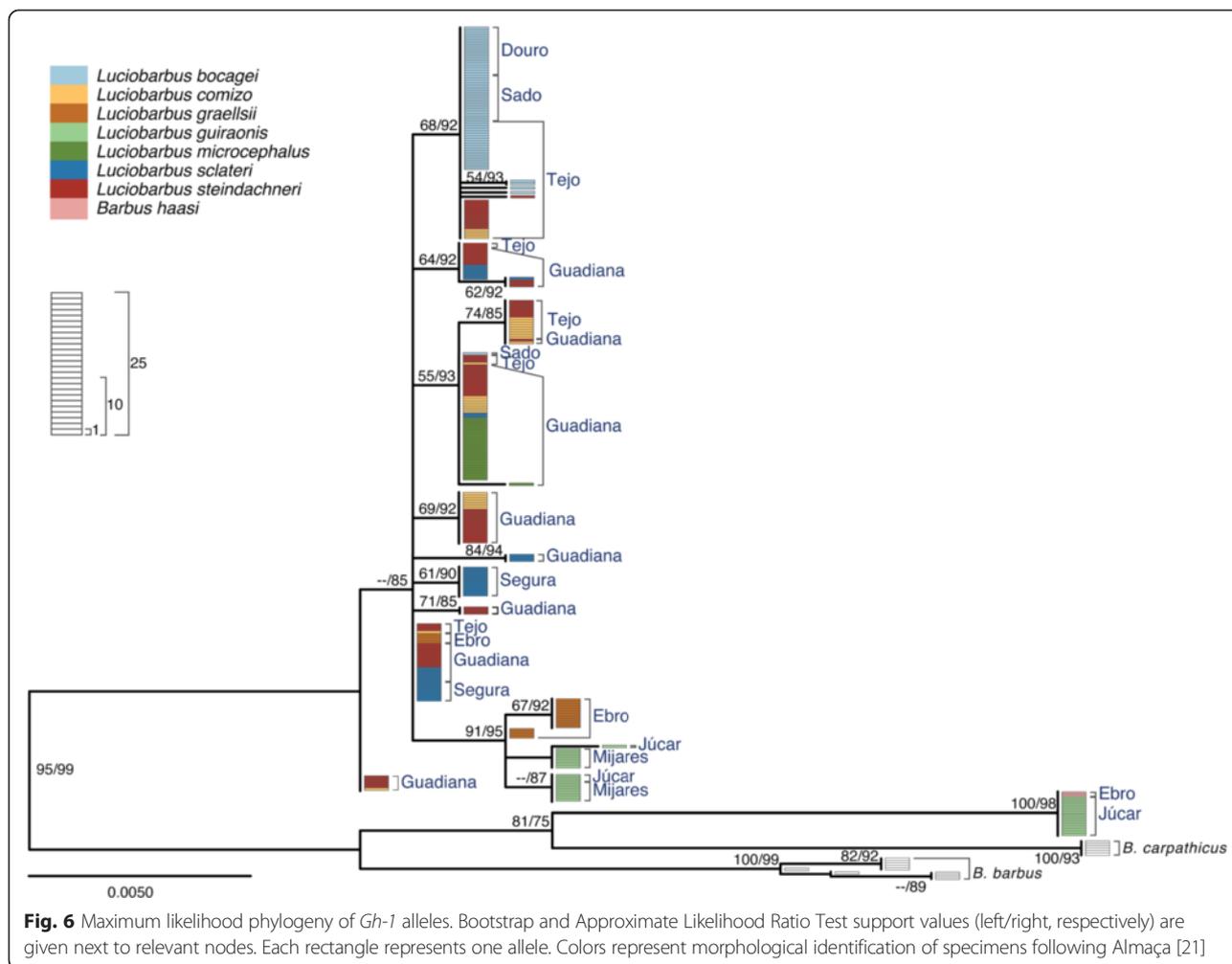
and Sado rivers) from rivers draining to the southern margin (Guadiana and Segura rivers). Further splitting the first dataset using this geographical discontinuity and patterns of gene exchange allows the identification of further intra-specific genetic structure ( $K = 3$  and  $K = 4$ ; Fig. 8, bottom). The second dataset is composed of species sympatric in rivers draining to the Mediterranean (Ebro, Júcar, Mijares and Bullent rivers) and closely related species (*i.e.*, *L. microcephalus* from the Guadiana River). This analysis converged to  $K = 4$  genetic clusters (Fig. 8, upper right), which is concordant with species identification based on morphology. The only discordant samples are *L. guiraonis* from the Júcar River, which are placed in the same group as *B. haasi*.

Overall we identified 10 distinct population clusters, which allowed the genetic discrimination of almost all species included in the study, as well as allopatric populations of *L. comizo* and *L. sclateri*, and identification of within-population differentiation in *L. bocagei* from Tejo. Conversely, and in spite of clear power to diagnose even distinct intraspecific clusters, *L. steindachneri* does not constitute a

genetically distinct unit. In turn, nuclear variation in *L. steindachneri* co-varies with sympatric species, in particular with *L. comizo* and *L. bocagei* in the Tejo River, and with *L. comizo* and *L. sclateri* (and *L. microcephalus* to a lesser extent) in the Guadiana River. This is another aspect noticeable from the STRUCTURE analyses, sympatric populations share more alleles across genetic clusters than allopatric ones, consistent with higher polymorphism levels in sympatry found above.

**Levels of gene flow across species boundaries of different ages**

We tested the hypothesis of gradual accumulation of genomic barriers to gene flow by examining the relation between the proportion of introgressed nuclear alleles (alleles shared only in areas of sympatry) between pairs of co-occurring species and their degree of divergence (see Methods section for details). We found that levels of gene flow across the species boundaries vary inversely with species divergence, *i.e.*, introgression is more reduced between more divergent species (*e.g.* between *L.*



*microcephalus* and *L. sclateri*) than between more closely related ones (e.g. between *L. bocagei* and *L. comizo*;  $r^2 = 0.773$ , Spearman's Rank  $\rho = -1$ ,  $P = 0.0416$ ; Fig. 9).

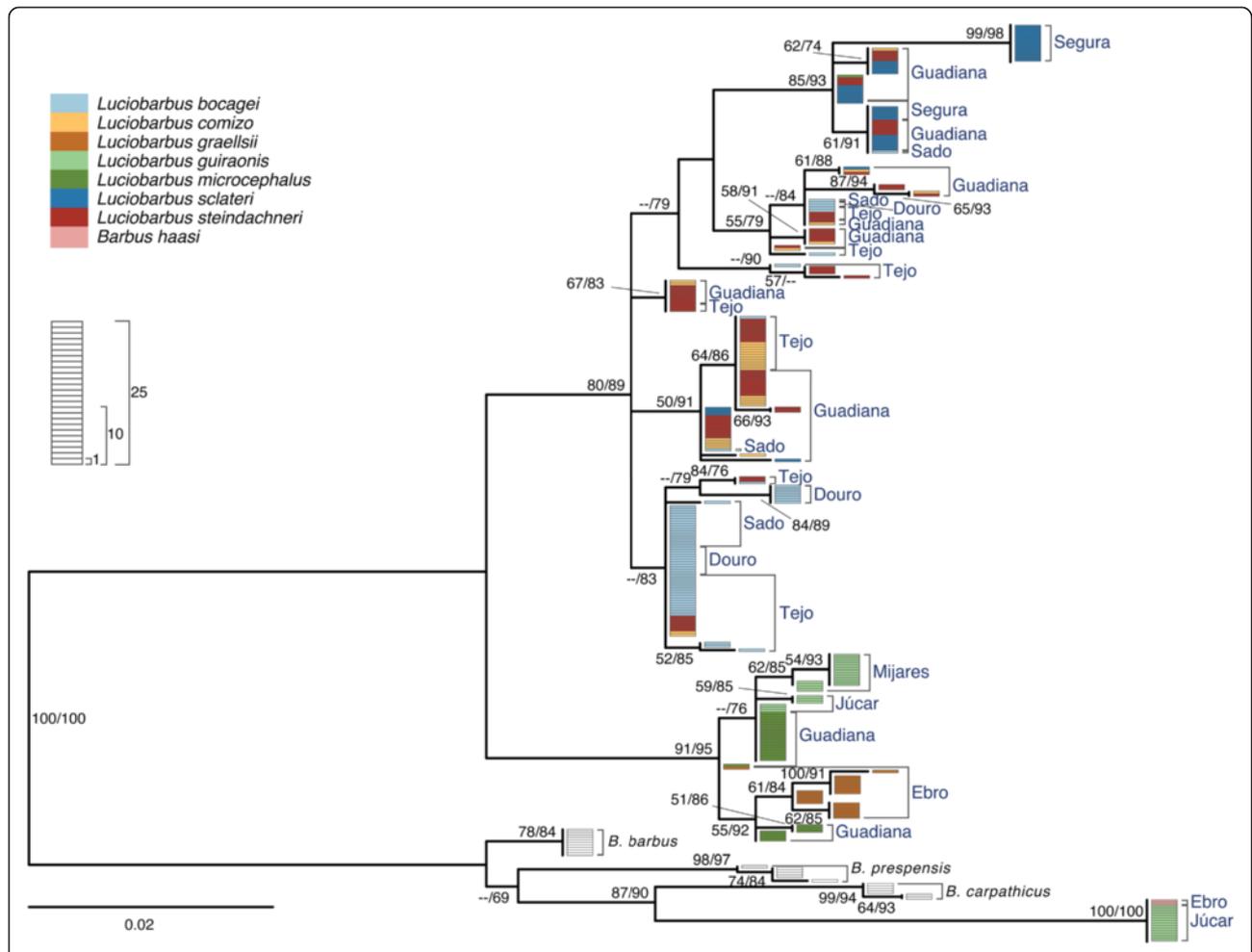
### Discussion

It has been suggested that studying pairs of populations and species spanning the speciation continuum can contribute more comprehensively to our understanding of how selection affects genomic divergence [9]. Our study successfully combines the analyses of populations and species spanning a wide degree of divergence, geographic distribution and range overlap, with different molecular markers and morphological traits, which provides exceptional perspective on the nature of species boundaries. Such a holistic approach has proven very informative in other systems (e.g. [49–54]). We show that species boundaries in Iberian *Barbus* and *Lucioibarbus* are semi-permeable, as different species (including non-sister taxa) are exchanging genes in areas of sympatry after an initial period of allopatric differentiation. Distribution of genetic variation across space, loci and species supports

a speciation-with-gene-flow scenario with levels of interspecific gene flow inversely associated with divergence. We also provide evidence for a hybrid origin of a barbel ecotype, *L. steindachneri*, suggesting that ecology plays a key role in species coexistence and hybridization in Iberian barbels (Fig. 10).

### Speciation-with-gene-flow in Iberian barbels

Patterns of molecular variation suggest that genomic differentiation in barbels generally starts in allopatry. It has been suggested that *Lucioibarbus* evolution follows river basin evolution as sister species usually inhabit one or a few adjacent river basins [23, 24, 26–28]. Gante et al. [30] showed that polytypic *L. sclateri* diversified in allopatry and expanded its distribution westwards in southern Iberia, leading to secondary contacts with *L. comizo* and *L. microcephalus* in the Lower Guadiana River less than 500 ka (thousand years ago). Geographical scenarios for the evolution of all other Iberian barbels are not equally detailed, but our data indicate that intraspecific differentiation in allopatry is the norm: in addition to discrimination

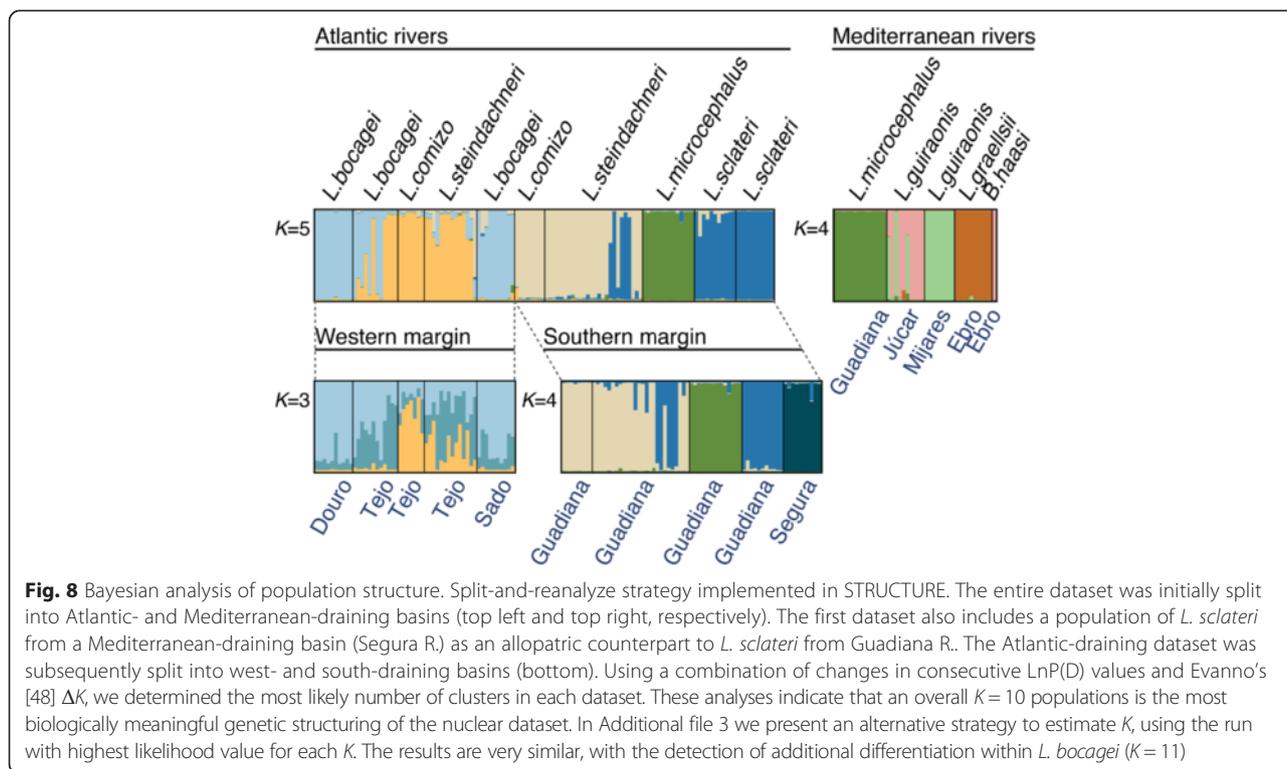


**Fig. 7** Maximum likelihood phylogeny of *Gh-2* alleles. Bootstrap and Approximate Likelihood Ratio Test support values (left/right, respectively) are given next to relevant nodes. Each rectangle represents one allele. Colors represent morphological identification of specimens following Almacá [21]

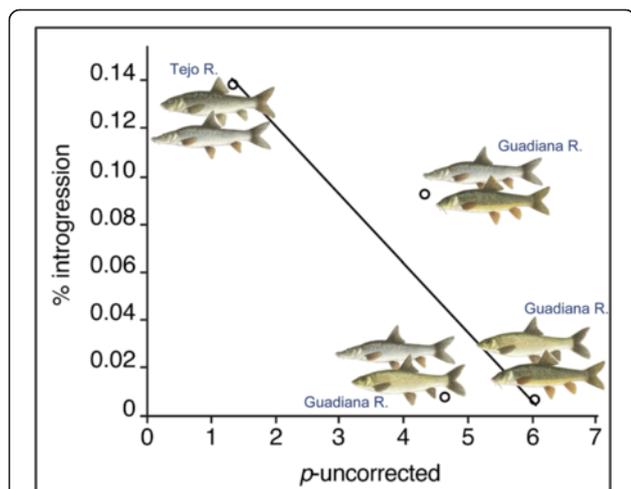
**Table 4** Nuclear sequence polymorphism in Iberian barbel populations

Species	Population	No.	S	$H_d$	K	$\pi$	$\theta$
<i>L. bocagei</i>	Douro	20	15	0.92632	5.22632	0.00177	0.00127
<i>L. bocagei</i>	Tejo	24	26	0.98913	5.84420	0.00198	0.00219
<i>L. comizo</i>	Tejo	14	23	0.97802	6.95604	0.00236	0.00213
<i>L. steindachneri</i>	Tejo	28	29	0.99471	9.31481	0.00316	0.00232
<i>L. bocagei</i>	Sado	20	21	0.94737	4.51053	0.00153	0.00191
<i>L. comizo</i>	Guadiana	16	23	1.00000	7.65000	0.00259	0.00238
<i>L. steindachneri</i>	Guadiana	52	51	0.99698	11.13801	0.00377	0.00373
<i>L. microcephalus</i>	Guadiana	28	35	0.94444	4.67989	0.00159	0.00318
<i>L. sclateri</i>	Guadiana	22	29	0.99134	7.08225	0.00240	0.00256
<i>L. sclateri</i>	Segura	20	11	0.94211	4.19474	0.00142	0.00087
<i>L. guiraois</i>	Júcar	20	91	0.70526	36.59474	0.01240	0.00888
<i>L. guiraois</i>	Mijares	16	5	0.85000	2.51667	0.00085	0.00043
<i>L. graellsii</i>	Ebro	20	17	0.98421	5.18421	0.00176	0.00162
<i>B. haasi</i>	Ebro	2	0	0.00000	0.00000	0.00000	0.00000

No.: number of alleles; S: number of segregating sites;  $H_d$ : haplotype diversity; K: average number of differences;  $\pi$ : nucleotide diversity;  $\theta$ : Watterson estimator

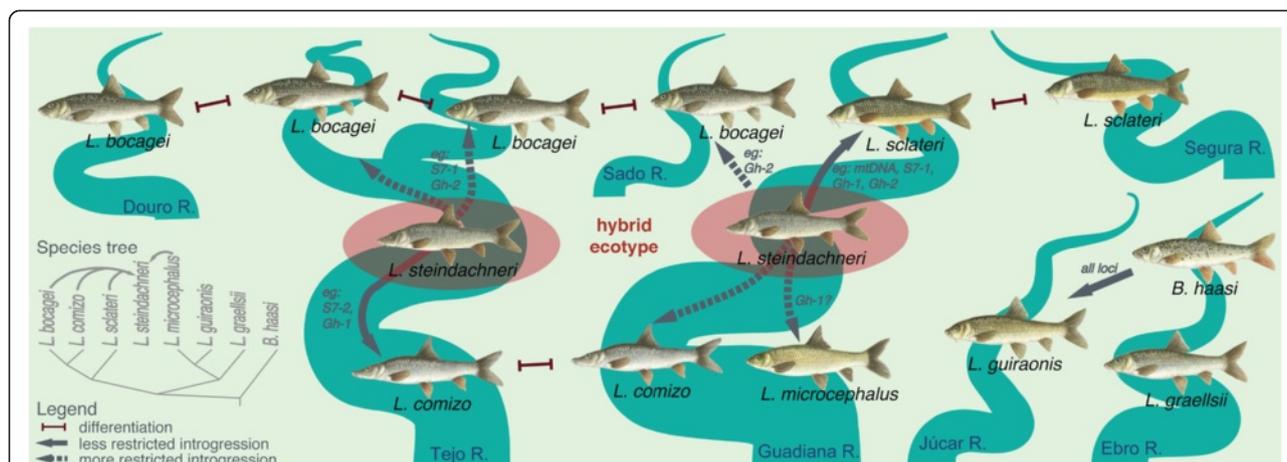


at the species levels, geographic (allopatric) differentiation is also found where samples from multiple basins are available (*L. bocagei*, *L. comizo* and *L. sclateri*). It is therefore reasonable to assume that along the speciation continuum the same mechanisms act during differentiation in allopatry



and are involved in all levels of differentiation, intra- and interspecific. We also observed within-basin differentiation in *L. bocagei*, which could be the result of its preference for low-order rivers as observed in rheophilic *Barbus* [20].

Contrasting patterns of molecular variation between areas of allopatry and sympatry is pivotal in excluding ancestral polymorphism as an explanation for the observed allele sharing and gene paralogy between co-occurring species of Iberian *Barbus* and *Luciobarbus*. Individuals are generally readily assignable to a monophyletic group representative of each species, with only individuals in sympatry sharing alleles with other taxa. This leads to higher levels of nucleotide polymorphism in sympatric populations of different species relative to their allopatric counterparts, consistent with a role for gene flow in generating these spatial patterns of molecular variation. Geographical information has also proved important in discriminating between these processes in other young systems, such as sunflowers, fruit flies, *Heliconius* butterflies and cichlid fishes (e.g. [4, 55–57]). It is clear that reproductive isolation is not yet fully complete in Iberian barbels, including between non-sister species (e.g. *L. comizo* and *L. sclateri*), indicating that genomes remain porous and allow for gene exchange, even between genera that diverged more than 20 ma [18–20]. Incomplete reproductive isolation seems to be the norm in other diverse systems in which species can be found in sympatry, including other barbels [20, 58–67]. Multiple characteristics make freshwater fish generally prone to hybridize [68]. In barbels, natural characteristics



**Fig. 10** Semi-permeable species boundaries in Iberian barbels. Species boundaries in Iberian barbels are semi-permeable to gene flow in areas of sympatry after an initial period of allopatric differentiation. Genetic data support a speciation-with-gene-flow scenario with levels of interspecific gene flow inversely associated with divergence, as represented by solid or dashed arrows. Traits involved in ecological preference (e.g. habitat, diet) are likely contributing to reproductive isolation and species coexistence. When these breakdown an ecotype of hybrid origin (*L. steindachneri*) is formed, frequent in areas with intermediate habitat characteristics

associated to external fertilization include spawning migrations and group spawning during which multiple males arrive to breeding grounds and await incoming females. In addition, the likely asymmetrical behavioral isolating mechanisms and unequal abundance of the co-occurring species increase the likelihood of interspecific gene flow. Acting alone or in combination, incorrect species recognition or sneaking behavior by small males might explain why often mispairings involve females of the largest species [20, 69, 70]. Pseudotetraploidy and long generation times (*i.e.* slow evolutionary rates) might also buffer against possible effects of intrinsic genomic incompatibilities and slow down the evolution of reinforcement mechanisms, which would allow introgression between genera separated by more than 20 ma of evolution [18–20]. Anthropogenic sources of decreased habitat complexity and construction of obstacles preventing spawning migrations to correct breeding sites (such as reservoirs and dams) might also increase the chances of hybridization. Although it is likely that anthropogenic activities have influenced the extent of hybridization and directionality of introgression we currently observed, it is probable that some interspecific gene flow has occurred soon after species became sympatric due to changes in drainage geometry. Nevertheless and in spite of ongoing gene flow in areas of sympatry, Iberian barbels can be distinguished using a combination of morphological and molecular data, which suggests that reproductive barriers, albeit variable in strength, do exist for at least some parts of the genome.

How then do reproductive barriers become established? Theory predicts that loci involved in adaptation to different niches and speciation should show increased differentiation while neutral or nearly neutral loci not

tightly linked to them should move more freely between populations and species [6, 71]. Therefore, genealogies derived from neutral or nearly neutral loci will be more susceptible to the impacts of interspecific gene flow than gene trees based on loci under divergent selection, which are prevented from introgressing and should better reflect species histories predating introgression after secondary contact. Additionally, frequency of new mutations is more likely to increase if they arise in regions of low introgression, potentially leading to a feedback loop of increased genomic differentiation [1]. Thus boundaries between older species pairs, which have accumulated more genetic differences, will likely exhibit greater genetic incompatibility and be less permeable to gene flow. We observe both scenarios in Iberian barbels: introgression across species boundaries varies among loci and gene flow seems to decay with increased divergence between species pairs, such that boundaries become less diffuse and more easily identifiable in more divergent, older species pairs. Therefore, reproductive isolation is not a property of the entire genome, nor does it arise instantaneously (*e.g.* [6]). Instead, these patterns are consistent with genetic differentiation (*i.e.* barriers to gene flow) starting in more or less localized genomic regions and then accumulating along the genome as species diverge [6, 9, 11, 58, 71–81]. Such ‘islands’ or ‘continents’ of the genome under stronger disruptive selection would exhibit reduced levels of introgression among species, relative to nearly neutral regions, retaining phylogenetic signal without noise from gene exchange among lineages. The relative impact of these two forces gradually changes with divergence until the influence of introgression is eventually ceased in all genomic regions. The

precise mechanisms by which genomes become less porous and the signatures they leave behind are a major focus of speciation research [11, 79, 82–84]. Thus, all lines of evidence suggest that population differentiation and speciation in Iberian barbels are initiated in allopatry and accumulate as populations diverge. However, reproductive barriers are not (yet) sufficient to fully prevent gene flow where species meet in regions of secondary contact, as seen in other species of barbels [20, 69, 85]. In addition to levels of introgression differing between species pairs, the environment is also expected to play a role in hybridization outcomes in particular when ecology influences their degree of sympatry. Possible environmental effects on hybridization dynamics could be disentangled from genomic divergence effects by using replicated hybrid zones of the same species pairs.

#### ***Luciobarbus steindachneri* is a hybrid ecotype**

The above considerations about species delimitation and mode of evolution fit all except for one nominal species, *L. steindachneri*. Its taxonomy has proved challenging since its description, as morphological and mtDNA similarity with *L. comizo* has already been identified [26, 31]. Here we provide evidence that explains these difficulties. Nuclear sequence data examined indicate that the genome of *L. steindachneri* shares its alleles across genetic clusters, mostly between *L. comizo* and all other sympatric species, both in the Tejo and Guadiana river basins. In addition, it has the highest numbers of segregating sites and levels of nucleotide diversity of sympatric species, which altogether support a scenario of admixture. It has the highest morphological variance of all Iberian barbels and is intermediate between several species, not only in meristic traits as seen in the present study and [21] but also in trophically relevant ones, such as mouth shape and position [35–37]. The latter are most likely under selection and have important fitness consequences [86, 87]. Differences in feeding apparatus likely explain observed food partitioning relative to *L. bocagei* and *L. comizo* in the Tejo River [35–37] and *L. microcephalus* in the Guadiana River [38], where it occupies an intermediate trophic position. Although *L. steindachneri* is always found in sympatry with other species, it prefers intermediate habitats and medium order rivers: it inhabits more frequently river stretches further upstream from those preferred by *L. comizo* and *L. microcephalus*, but downstream from typical *L. sclateri* habitat in the Guadiana River [32–34] and downstream from *L. bocagei* in the Tejo River [Gante pers. obs.]. Observed habitat preferences likely reflect selection on trophic morphology and trophic niche. These lines of evidence suggest that ecology plays a major role in the coexistence and hybridization among sympatric barbel species. *Luciobarbus steindachneri* is thus the local product of introgressive hybridization between *L. comizo* and *L. bocagei* in the Tejo, and *L.*

*microcephalus* and *L. sclateri* in the Guadiana, for which we have no evidence to be a stabilized hybrid species. Therefore, following Morán-López *et al.*'s [34] terminology, we show that *L. steindachneri* is an ecotype of hybrid origin that is using a niche intermediate to those of its parents [88]. In spite of this general intermediacy, whether its ecology varies between the Tejo and Guadiana rivers, and with parental crosses involved has not been assessed but is likely to be the case. Overall these findings raise questions on how one of the most common barbels found in Iberian rivers should be dealt with from taxonomic and conservation perspectives, as it is of hybrid origin and not stabilized. Conversely, it becomes a great model in evolutionary ecology. Sampling of replicated hybrid zones would help elucidate how stable this hybrid is and other aspects, such as geographical differences in parental contribution and levels of hybridization.

#### **Conclusions**

A fundamental question in speciation research is the evolution of barriers to gene flow and the long-term persistence of taxa in sympatry. We provide evidence for semi-permeable species boundaries in Iberian *Barbus* and *Luciobarbus* (summarized in Fig. 10). Different species are exchanging genes in areas of sympatry after an initial period of allopatric differentiation. Genomic barriers to gene flow are heterogeneous in strength and accumulate with increased divergence. Particularly puzzling is the case of intergeneric hybrids in the Júcar River, where only a hybrid swarm was found. Additional sampling outside this area would be necessary to tease apart possible causes for the apparent collapse or absence of isolating mechanisms. We also show that *L. steindachneri* is an ecotype of hybrid origin, suggesting that ecology plays a key role in species coexistence and hybridization. Since Iberian barbels span a wide range of genetic divergence, diverse ecologies, and degrees of sympatry, they provide an ideal opportunity for studying the evolution of reproductive isolation mechanisms. Given the complexities of this system (*i.e.*, ancient polyploidy) recent improvements in technology have made it possible to study systems that were traditionally less tractable due to technical challenges. In particular, newer deep sequencing technologies will allow further insight into the patterns of genomic divergence in the face of gene flow and mechanisms shaping reproductive barriers in barbels by sampling many more loci than the ones surveyed here. For instance, it would allow determining which genomic regions are responsible for reproductive isolation and when they arose relative to the establishment of secondary contacts.

#### **Methods**

##### **Specimen collection and sample processing**

Specimens of Iberian *Barbus* and *Luciobarbus* were collected by electrofishing throughout the species' ranges

(Fig. 1). All specimens were formalin fixed and ethanol preserved, and tissues ethanol preserved and deposited in the zoological collections 'Museu Bocage' (MB) of Museu Nacional de História Natural e da Ciência, Portugal, and in Museo Nacional de Ciencias Naturales (MNCN), Spain. Some museum samples were used only for morphological analyses, while others from central and eastern Europe were used only to provide context in phylogenetic analyses of molecular characters (Table 1).

We identified individuals to species based on Almaça's [21] qualitative traits, such as dorsal fin shape, robustness of the last dorsal spine, head shape, eye and mouth position, barbel length and coloration of body and fins. Assignment of individuals to *L. comizo* and *L. steindachneri* can be difficult due to morphological similarity that has motivated their synonymy [31]; conservatively only specimens with clear phenotypes were assigned to *L. comizo*. Covariation of meristic morphological traits and molecular markers (see below) would reject the null hypothesis of 'no species differentiation' and confirm that qualitative traits are truly representative of species identity [89].

#### Choice, scoring and analysis of morphological traits

For testing species discrimination using morphological traits, we chose additional independent characters based on their potential information content, scoring reliability and reproducibility, their consistent preservation in different conditions and ease of use. For these reasons, external meristic traits were preferred over morphometric traits, as the latter are further complicated by allometry, which allow us to include also juvenile specimens in the study. Even though there is substantial overlap in scale counts across different *Luciobarbus* species (e.g. [21, 31]), they vary widely and also exhibit different modal values. Three different scale counts were taken: number of scales along the lateral line (LL); number of scales in transverse rows above the LL, counted anteriorly-posteriorly from the LL to the sagittal line in front of the first dorsal-fin ray, not including the LL (TRA); number of scales in transverse rows below the LL, counted anteriorly-posteriorly from the LL to the insertion of the first pelvic-fin ray, not including the LL (TRB). Allowance was made for abnormal scale development, such as duplicated or fused scales, evaluated by comparison with scales from rows above and below the scales in question. Numbers of cephalic canal pores found to be useful in the diagnosis of *Barbus* species [90] were also examined, including those on the supraorbital canal (SOC), the infraorbital canal (IOC) and the preopercular-mandibular canal (POMC). Counts were taken preferentially on the left side under a dissecting scope using Cyanine Blue 5R temporary stain to enhance contrast of structures [91].

Covariation among morphological traits was assessed using principal components analysis (PCA), as implemented

in PAST [92]. PCA explores co-linear variation of the original variables, reducing multidimensionality of the data into new orthogonal variables (principal components). The variance-covariance matrix of standardized meristic variables was used.

#### Scoring and sequencing of mitochondrial and nuclear loci

The mitochondrial gene cytochrome *b* (*cyt b*) has proved to be useful in phylogenetic and phylogeographic studies of *Barbus* and *Luciobarbus* (e.g. [24, 26, 47, 93]). We screened for single-stranded conformational polymorphisms (SSCPs) of a 275 bp fragment following methods in Gante et al. [45]. Sequences of multiple (>10 %) representative SSCP bands from each gel were confirmed by Sanger sequencing on an ABI 3730 DNA Analyzer.

Several PCR primers for nuclear loci have become available for cyprinid fishes. Due to pseudotetraploidy of barbels, paralog-specific primers were used following Gante et al. (2011). Briefly, we used a hybrid annealing strategy, combining both universal exon-primed intron-crossing (EPIC) and paralog-specific intron-primed exon-crossing (IPEC) primers for amplification and sequencing of four nuclear loci: *S7-1*, *S7-2*, *Gh-1* and *Gh-2*. Sequences were obtained by Sanger sequencing on an ABI 3730 DNA Analyzer. Different methods were employed to resolve haplotypic phase of heterozygous individuals. Where individuals were heterozygous for insertions or deletions (indels), haplotypes were either manually phased using the method described by Flot et al. [94] or using the program CHAMPURU [95]. Haplotypes with known phases were subsequently used to phase the single nucleotide polymorphism (SNP) heterozygotes with PHASE [96]. PHASE input files were generated using SEQPHASE [97]. Consistency of the inferred haplotypes was assessed in five independent PHASE runs, each of 100 iterations, burn-in of 100 and thinning interval of 1.

#### Phylogenetic analyses of nucleotide sequences

The models of nucleotide sequence evolution for the different datasets were identified with JMODELTEST 2 [98, 99] using the corrected Akaike and Bayesian Information Criteria. Maximum likelihood allele phylogenies were built in PHYML 3.0 [100] using an HKY85 model of nucleotide evolution with four substitution rate categories, estimated transition/transversion ratios, with a proportion of invariable sites and gamma distribution parameter, and using empirical nucleotide frequencies. Previous published complete *cyt b* sequences of all species [26] were included in the analysis to check relationships of haplotypes found in the present study. Topology searches used NNI (nearest-neighbor interchange) and SPR (subtree pruning and regrafting) on a BioNJ starting tree. Node support was estimated using 1000 bootstrap replicates and approximate Bayes likelihood ratio tests.

### Levels of nucleotide polymorphism and population differentiation

Number of segregating sites ( $S$ ), haplotype diversity ( $H_d$ ), average number of differences ( $K$ ), nucleotide diversity ( $\pi$ ) and Watterson estimator ( $\theta$ ), and pairwise population divergence measures ( $D_a$ ,  $D_{xy}$  and  $F_{ST}$ ) were calculated using the program DNASP v5.10.01 [101]. Neighbor-joining (NJ) networks of pairwise population differentiation statistics were constructed in MEGA 5.05 [102] for each locus separately and all nuclear loci combined. Node support for each network was assessed by bootstrapping nucleotide sequence alignments (100 times) in SEQBOOT from the PHYLIP package [103] and re-calculating estimates.

### Bayesian clustering of nuclear DNA data

Covariation among nuclear loci was assessed using the Bayesian clustering program STRUCTURE v2.3.3 [104, 105]. STRUCTURE identifies clusters by assigning individuals to  $K$  populations in a way that minimizes deviations from Hardy–Weinberg equilibrium within clusters and maximizes linkage disequilibrium between them (*i.e.* species). Each unique allele was identified using the program MACCLADE v4.03 [106] and numerically coded (*i.e.* we coded each unique combination of SNPs at each locus and not each individual SNP). Due to the complexity (multimodality) of the complete nuclear dataset, STRUCTURE sometimes converged to different solutions in independent runs for the same  $K$ . Therefore, smaller, more tractable datasets consisting of combinations of sympatric and sister species and allopatric references were analyzed independently as recommended by Pritchard & Wen [107]. This approach allows for assessment of the importance of introgression and ancestral polymorphism on allele sharing. To assess reliability of solutions, 20 iterations were run for each  $K$ . Each iteration consisted of 500,000 MCMC (Markov Chain Monte Carlo) generations as burn-in, followed by 1,000,000 MCMC replicates to estimate the posterior sample distribution, using the admixture and correlated allele frequency models.  $\text{LnP}(D)$ , the probability of the data given  $K$ , was tracked over the course of the burn-in and the run to ensure that these values had stabilized by the end of the burn-in period. Two different methods were used to determine the number of groups ( $K$ ) identified by STRUCTURE runs of each dataset. The first method, suggested by Pritchard & Wen [107], identifies the most likely value of  $K$  by comparing changes in  $\text{LnP}(D)$  values of consecutive  $K$  (*i.e.* when values plateau). The second method, developed by Evanno *et al.* [48] finds the ad hoc quantity based on the second order rate of change of the likelihood function with respect to  $K$  ( $\Delta K$ ). Plots of these two metrics were obtained using STRUCTURE HARVESTER [108]. Consensus clustering across iterations for each for each  $K$  was generated using the greedy algorithm in CLUMPP [109] and

visualized using the program DISTRUCT [110]. In addition, we present and discuss an alternative strategy to determine the best  $K$  using the complete dataset in the supplementary information (Additional file 3).

### Levels of introgression between pairs of hybridizing species

To test the hypothesis that reproductive barriers accumulate in the genome as species diverge, we calculated the proportion of introgressed alleles between pairs of hybridizing taxa. We restricted the analysis to pairs of species that we see from our data to co-occur and exchange genes in sympatry to avoid obscuring any patterns with zero-inflation from non-hybridizing species pairs (*e.g.* because they are never in sympatry). We also excluded the intergeneric hybridizing pair *B. haasi* and *L. guiraonis*, as ‘pure’ *B. haasi* has not yet been collected in the Júcar River. It is not presently clear if it inhabits some area of the basin yet to be sampled or if it went locally extinct. Therefore we analyzed the hybridizing pairs *L. bocagei*–*L. comizo*, *L. comizo*–*L. sclateri*, *L. microcephalus*–*L. sclateri* and *L. comizo*–*L. microcephalus*. Diagnostic alleles were identified from nuclear allele phylogenies: alleles shared only in sympatric populations of species A, but absent in their allopatric counterparts while common in sympatric or allopatric populations of species B were considered typical of the latter. Because *L. microcephalus* is fixed for *Gh-1* and inhabits only the Guadiana River we do not have allopatric samples to help us distinguish whether it shares *Gh-1* alleles through introgression and complete replacement, or by retained ancestral polymorphism, we conservatively excluded this locus from comparisons involving this species. We used pairwise uncorrected mean divergence between taxa ( $p$ -uncorrected) derived from mtDNA as a surrogate for levels of species divergence (from [26]). Therefore, this estimate of divergence between taxa is independent from nuclear introgression levels determined in the present study and should better reflect species divergence in the absence of interspecific gene flow.

### Ethics statement

Direcção Geral dos Recursos Florestais (DGRF) provided the necessary collection permits. Fish were euthanized with an overdose of MS-222 (3-aminobenzoic acidethyl ester methanesulfonate) prior to handling.

### Availability of supporting data

The datasets supporting the results of this article are deposited in the Dryad Digital Repository [<http://dx.doi.org/10.5061/dryad.m1572>] and include five sequence alignment files and one morphological (meristic) traits file [111].

## Additional files

**Additional file 1: Divergence measures across population pairs of Iberian barbels.**

**Additional file 2: Population polymorphisms networks across loci.**

**Additional file 3: STRUCTURE analysis using different numbers of samples.**

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

All authors read and contributed to the final manuscript, HFG, MJA and TED designed the experiment, HFG and ID collected samples, HFG analyzed the data and drafted the first version of the manuscript. All authors read, reviewed and approved the final version of the manuscript.

### Acknowledgements

We thank Petr Kotlík, J.A. Carmona and Francisco J. Oliva-Paterna for providing samples. We are indebted to V. Cristão, C. Cristão Jr, T. Marques, F. Ribeiro, L. da Costa, C.D. Santos and M. Torralva for help in the field. This work was supported financially by Fundação para a Ciência e a Tecnologia (FCT) project grant POCTI/BSE/35121/2000 to MJA, and PhD fellowships SFRH/BD/13067/2003 from FCT (co-participated by FEDER, POCI 2010 and FSE), Proc.85547 from Fundação Calouste Gulbenkian and from the Arizona State University School of Life Sciences to HFG. This research received support from the SYNTHESYS Project ES-TAF-1605 to HFG, <http://www.synthesys.info/>, which is financed by European Community Research Infrastructure Action under the FP7 "Capacities" Programme, at the Museo Nacional de Ciencias Naturales (CSIC). Dirección Geral dos Recursos Florestais (DGRF) provided the necessary collecting permits for Portugal.

### Author details

<sup>1</sup>School of Life Sciences, Arizona State University, 85287-4601 Tempe, AZ, USA. <sup>2</sup>Museu Nacional de História Natural e da Ciência, Centre for Ecology, Evolution and Environmental Changes (Ce3C), Universidade de Lisboa, Rua da Escola Politécnica 58, 1250-102 Lisbon, Portugal. <sup>3</sup>Current address: Zoological Institute, University of Basel, 4051 Basel, Switzerland. <sup>4</sup>Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, c/José Gutiérrez Abascal 2, 28006 Madrid, Spain. <sup>5</sup>Current address: Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, 48202 Detroit, MI, USA.

Received: 17 January 2015 Accepted: 28 May 2015

Published online: 12 June 2015

### References

- Gavrilets S: Fitness Landscapes and the Origin of Species. Princeton, New Jersey: Princeton University Press; 2004.
- Rice WR, Hostert EE. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* (N Y). 1993;47:1637–53.
- Smadja CM, Butlin RK. A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol*. 2011;20:5123–40.
- Noor MAF, Bennett SM. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* (Edinb). 2009;103:439–44.
- Rieseberg LH, Whitton J, Gardner K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*. 1999;152:713–27.
- Wu C-I. The genic view of the process of speciation. *J Evol Biol*. 2001;14:851–65.
- Machado CA, Kliman RM, Markert JA, Hey J. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol*. 2002;19:472–88.
- Bull V, Beltrán M, Jiggins CD, McMillan WO, Bermingham E, Mallet J. Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol*. 2006;4:11.
- Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. 2009;18:375–402.
- Carneiro M, Ferrand N, Nachman MW. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*. 2009;181:593–606.
- Nosil P, Feder JL. Genomic divergence during speciation: causes and consequences. *Philos Trans R Soc Lond B Biol Sci*. 2012;367:332–42.
- Sato A, O'hUigin C, Figueroa F, Grant PR, Grant BR, Tichy H, et al. Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proc Natl Acad Sci*. 1999;96:5101–6.
- Beltrán M, Jiggins CD, Bull V, Linares M, Mallet J, McMillan WO, et al. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol Biol Evol*. 2002;19:2176–90.
- Shaw KL. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sci U S A*. 2002;99:16122–7.
- Heckman KL, Mariani CL, Rasoloarison R, Yoder AD. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Mol Phylogenet Evol*. 2007;43:353–67.
- Mullen SP, Dopman EB, Harrison RG. Hybrid zone origins, species boundaries, and the evolution of wing-pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: *Limenitis*). *Evolution*. 2008;62:1400–17.
- Pinho C, Harris DJ, Ferrand N. Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. *BMC Evol Biol*. 2008;8:63.
- Gante HF. Diversification of circum-Mediterranean barbels. In Chang Divers Chang Environ. Edited by Grillo O, Venora G. Rijeka, Croatia: InTech; 2011:283–298.
- Levin BA, Freyhof J, Lajbner Z, Perea S, Abdoli A, Gaffaroğlu M, et al. Phylogenetic relationships of the algae scraping cyprinid genus *Capoeta* (Teleostei: Cyprinidae). *Mol Phylogenet Evol*. 2012;62:542–9.
- Buonerba L, Zaccara S, Delmastro GB, Lorenzoni M, Salzburger W, Gante HF. 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.
- Almaça C. Estudo das populações portuguesas do gén. *Barbus* Cuvier, 1817 (Pisces, Cyprinidae). *Rev da Fac Ciências, Lisboa, 2a série*. 1967;14:151–400.
- Doadrio I. Phylogenetic relationships and classification of western palaearctic species of the genus *Barbus* (Osteichthyes, Cyprinidae). *Aquat Living Resour*. 1990;3:265–82.
- Zardoya R, Doadrio I. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. *Proc Biol Sci*. 1998;265:1365–72.
- Zardoya R, Doadrio I. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *J Mol Evol*. 1999;49:227–37.
- Machordom A, Doadrio I. Evolutionary history and speciation modes in the cyprinid genus *Barbus*. *Proc Biol Sci*. 2001;268:1297–306.
- Doadrio I, Carmona JA, Machordom A. Haplotype diversity and phylogenetic relationships among the Iberian barbels (*Barbus*, Cyprinidae) reveal two evolutionary lineages. *J Hered*. 2002;93:140–7.
- Machordom A, Doadrio I. Evidence of a cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (*Luciobarbus*, Cyprinidae). *Mol Phylogenet Evol*. 2001;18:252–63.
- Mesquita N, Cunha C, Carvalho GR, Coelho MM. Comparative phylogeography of endemic cyprinids in the south-west Iberian Peninsula: evidence for a new ichthyogeographic area. *J Fish Biol*. 2007;71(sa).
- Burridge CP, Craw D, Jack DC, King TM, Waters JM. Does fish ecology predict dispersal across a river drainage divide? *Evolution* (N Y). 2008;62:1484–99.
- Gante HF, Micael J, Oliva-Paterna FJ, Doadrio I, Dowling TE, Alves MJ. Diversification within glacial refugia: tempo and mode of evolution of the polytypic fish *Barbus sclateri*. *Mol Ecol*. 2009;18:3240–55.
- Doadrio I. Sobre la taxonomía de *Barbus comiza* Steindachner, 1865 (Ostariophys: Ciprinidae). *Doñana Acta Vertebr*. 1988;15:19–28.
- Filipe AF, Cowx IG, Collares-Pereira MJ. Spatial modelling of freshwater fish in semi-arid river systems: a tool for conservation. *River Res Appl*. 2002;18:123–36.
- Filipe AF, Marques TA, Seabra S, Tiago P, Ribeiro F, Costa LMDA, et al. Selection of priority areas for fish conservation in Guadiana River Basin, Iberian Peninsula. *Conserv Biol*. 2004;18:189–200.
- Morán-López R, Pérez-Bote JL, Da Silva RE, Corbacho Amado C. Summer habitat relationships of barbels in south-west Spain. *J Fish Biol*. 2005;67:66–82.

35. Encina LE, Granado-Lorencio C. A quantitative comparison of the jaw apparatus in three species of *Barbus* (Cyprinidae, Teleostei). *J Anim Morphol Physiol*. 1989;36:9–30.
36. Encina LE, Granado-Lorencio C. Morfoecología trófica en el género *Barbus* (Pisces, Cyprinidae). *Limnetica*. 1990;6:35–46.
37. Encina LE, Granado-Lorencio C. Food habits and food resource partitioning in three coexisting *Barbus* species. *Folia Zool*. 1997;46:325–36.
38. Pires AM, Cowx IG, Coelho MM. Diet and growth of two sympatric Iberian barbel, *Barbus steindachneri* and *Barbus microcephalus*, in the middle reaches of the Guadiana Basin (Portugal). *Folia Zool*. 2001;50:291–304.
39. Steindachner F. Ichthyologischer Bericht über eine nach Spanien und Portugal unternommene Reise. (Fortsetzung) Über die Fische des Ebro und der Flüsse bei Bilbao. *Sitzungsbd d kais Akad d Wissenschaft*. 1866;53:198–204.
40. Almaça C. Sur la systématique des Barbeaux (genre et sous-genre *Barbus*) de la Péninsule Ibérique et de l'Afrique du Nord. *Arq do Mus Bocage*, 2a série. 1972;3:319–46.
41. Almodóvar A, Nicola GG, Elvira B. Natural hybridization of *Barbus bocagei* x *Barbus comizo* (Cyprinidae) in Tagus River basin, central Spain. *Cybiun*. 2008;32:99–102.
42. Callejas C, Ochando MD. Recent radiation of Iberian barbel fish (Teleostei, Cyprinidae) inferred from cytochrome *b* genes. *J Hered*. 2000;91:283–8.
43. Machordom A, Berrebi P, Doadrio I. Spanish barbel hybridization detected using enzymatic markers: *Barbus meridionalis* Risso x *Barbus haasi* Mertens (Osteichthyes, Cyprinidae). *Aquat Living Resour*. 1990;3:295–303.
44. Gante HF, Alves MJ, Dowling TE. Paralog-specific primers for the amplification of nuclear loci in tetraploid barbels (*Barbus*: cypriniformes). *J Hered*. 2011;102:617–21.
45. Gante HF, Alves MJ, Dowling TE. Development of cytochrome *b* primers for mitotyping of barbels (*Barbus* spp.). *Mol Ecol Resour*. 2008;8:786–9.
46. Machordom A, Doadrio I, Berrebi P. Phylogeny and evolution of the genus *Barbus* in the Iberian Peninsula as revealed by allozyme electrophoresis. *J Fish Biol*. 1995;47:211–36.
47. Tsigonopoulos CS, Durand JD, Ünlü E, Berrebi P. Rapid radiation of the Mediterranean *Luciobarbus* species (Cyprinidae) after the Messinian salinity crisis of the Mediterranean Sea, inferred from mitochondrial phylogenetic analysis. *Biol J Linn Soc*. 2003;80:207–22.
48. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–20.
49. Roe AD, Sperling FAH. Population structure and species boundary delimitation of cryptic *Dioxytricia* moths: an integrative approach. *Mol Ecol*. 2007;16:3617–33.
50. Markolf M, Rakotonirina H, Fichtel C, von Grumbkow P, Brameier M, Kappeler PM. True lemurs... true species - species delimitation using multiple data sources in the brown lemur complex. *BMC Evol Biol*. 2013;13:233.
51. Lecocq T, Dellicour S, Michez D, Lhomme P, Vanderplanck M, Valterová I, et al. Seal of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*). *BMC Evol Biol*. 2013;13:263.
52. Bendik NF, Meik JM, Gluesenkamp AG, Roelke CE, Chippindale PT. Biogeography, phylogeny, and morphological evolution of central Texas cave and spring salamanders. *BMC Evol Biol*. 2013;13:201.
53. Thomé MTC, Zamudio KR, Haddad CFB, Alexandrino J. Delimiting genetic units in Neotropical toads under incomplete lineage sorting and hybridization. *BMC Evol Biol*. 2012;12:242.
54. Jacquemyn H, Brys R, Honnay O, Roldán-Ruiz I. Asymmetric gene introgression in two closely related *Orchis* species: evidence from morphometric and genetic analyses. *BMC Evol Biol*. 2012;12:178.
55. Kane NC, King MG, Barker MS, Raduski A, Karrenberg S, Yatabe Y, et al. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*. 2009;63:2061–75.
56. Mims MC, Darrin Hulsey C, Fitzpatrick BM, Strelman JT. Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Mol Ecol*. 2010;19:940–51.
57. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res*. 2013;23:1817–28.
58. Peccoud J, Ollivier A, Plantegenest M, Simon J-C. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci U S A*. 2009;106:7495–500.
59. Wahlberg N, Weingartner E, Warren AD, Nylin S. Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). *BMC Evol Biol*. 2009;9:92.
60. De Busschere C, Hendrickx F, Van Belleghem SM, Backeljau T, Lens L, Baert L. Parallel habitat specialization within the wolf spider genus *Hogna* from the Galápagos. *Mol Ecol*. 2010;19:4029–45.
61. Fontenot BE, Makowsky R, Chippindale PT. Nuclear-mitochondrial discordance and gene flow in a recent radiation of toads. *Mol Phylogenet Evol*. 2011;59:66–80.
62. Pereira RJ, Monahan WB, Wake DB. Predictors for reproductive isolation in a ring species complex following genetic and ecological divergence. *BMC Evol Biol*. 2011;11:194.
63. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, et al. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Mol Ecol*. 2013;22:814–26.
64. Willis SC, Macrander J, Farias IP, Ortí G. Simultaneous delimitation of species and quantification of interspecific hybridization in Amazonian peacock cichlids (genus *Cichla*) using multi-locus data. *BMC Evol Biol*. 2012;12:96.
65. Gante HF, Santos CD, Alves MJ. Phylogenetic relationships of the newly described species *Chondrostoma olisiponensis* (Teleostei: Cyprinidae). *J Fish Biol*. 2010;76:965–74.
66. Sousa-Santos C, Gante HF, Robalo J, Proença Cunha P, Martins A, Arruda M, et al. Evolutionary history and population genetics of a cyprinid fish (*Iberochondrostoma olisiponensis*) endangered by introgression from a more abundant relative. *Conserv Genet*. 2014;15:665–77.
67. Gante HF, Collares-Pereira MJ, Coelho MM. Introgressive hybridisation between two Iberian *Chondrostoma* species (Teleostei, Cyprinidae) revisited: new evidence from morphology, mitochondrial DNA, allozymes and NOR-phenotypes. *Folia Zool*. 2004;53:423–32.
68. Scribner KT, Page KS, Bartron ML. Hybridization in freshwater species: a review of case studies and cytonuclear methods of biological inference. *Rev Fish Biol Fish*. 2000;10:293–323.
69. Lajbner Z, Slechtová V, Slechta V, Svátora M, Berrebi P, Kotlík P. Rare and asymmetrical hybridization of the endemic *Barbus carpathicus* with its widespread congener *Barbus barbus*. *J Fish Biol*. 2009;74:418–36.
70. Meraner A, Venturi A, Ficetola GF, Rossi S, Candiotti A, Gandolfi A. Massive invasion of exotic *Barbus barbus* and introgressive hybridization with endemic *Barbus plebejus* in Northern Italy: Where, how and why? *Mol Ecol*. 2013;22:5295–312.
71. Pinho C, Hey J. Divergence with gene flow: models and data. *Annu Rev Ecol Evol Syst*. 2010;41:215–30.
72. Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, et al. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci U S A*. 2003;100:10818–23.
73. Llopert A, Lachaise D, Coyne JA. Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics*. 2005;171:197–210.
74. Basset P, Yannic G, Brünner H, Hausser J. Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. *Evolution*. 2006;60:1718–30.
75. Harr B. Genomic islands of differentiation between house mouse subspecies. *Genome Res*. 2006;16:730–7.
76. Carling MD, Brumfield RT. Speciation in *Passerina* buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. *Mol Ecol*. 2009;18:834–47.
77. Carneiro M, Blanco-Aguilar JA, Villafuerte R, Ferrand N, Nachman MW. Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution*. 2010;64:3443–60.
78. Scascitelli M, Whitney KD, Randell RA, King M, Buerkle CA, Rieseberg LH. Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Mol Ecol*. 2010;19:521–41.
79. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. *Trends Genet*. 2012;28:342–50.
80. Kautt AF, Elmer KR, Meyer A. Genomic signatures of divergent selection and speciation patterns in a “natural experiment”, the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Mol Ecol*. 2012;21:4770–86.

81. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci*. 2012;367:343–53.
82. Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*. 2010;64:1729–47.
83. Flaxman SM, Feder JL, Nosil P. Spatially explicit models of divergence and genome hitchhiking. *J Evol Biol*. 2012;25:2633–50.
84. Flaxman SM, Wacholder AC, Feder JL, Nosil P. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol Ecol* 2014;23:4074–4088.
85. Tsigenopoulos CS, Kotlík P, Berrebi P. Biogeography and pattern of gene flow among *Barbus* species (Teleostei: Cyprinidae) inhabiting the Italian Peninsula and neighbouring Adriatic drainages as revealed by allozyme and mitochondrial sequence data. *Biol J Linn Soc*. 2002;75:83–99.
86. Lavin PA, McPhail JD. Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). *Can J Fish Aquat Sci*. 1986;43:2455–63.
87. Barlow GW. *The Cichlid Fishes: Nature's Grand Experiment In Evolution*. New York: Basic Books; 2000.
88. Mallet J. Hybrid speciation. *Nature*. 2007;446:279–83.
89. Douglas ME, Minckley WL, Tyus HM. Qualitative characters, identification of Colorado River chubs (Cyprinidae: genus *Gila*) and the "art of seeing well". *Copeia*. 1989;2009:653–62.
90. Banarescu P, Bogutskaya N. *The Freshwater Fishes of Europe Volume 5/II: Cyprinidae 2: Part II: Barbus*. Aula Verlag: Wiesbaden; 2003.
91. Saruwatari T, López JA, Pietsch TW. Cyanine blue: a versatile and harmless stain for specimen observation. *Copeia*. 1997;1997:840–1.
92. Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron* 2001, 4:[http://palaeo-electronica.org/2001\\_1/past/issue1\\_01.htm](http://palaeo-electronica.org/2001_1/past/issue1_01.htm).
93. Kotlík P, Bogutskaya NG, Ekmekçi FG. Circum Black Sea phylogeography of *Barbus* freshwater fishes: divergence in the Pontic glacial refugium. *Mol Ecol*. 2004;13:87–95.
94. Flot J-F, Tillier A, Samadi S, Tillier S. Phase determination from direct sequencing of length-variable DNA regions. *Mol Ecol Notes*. 2006;6:627–30.
95. Flot J-F. champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Mol Ecol Notes*. 2007;7:974–7.
96. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68:978–89.
97. Flot J-F. seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Mol Ecol Resour*. 2010;10:162–6.
98. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9:772.
99. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003;52:696–704.
100. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
101. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.
102. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
103. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. 2005.
104. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
105. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
106. Maddison DR, Maddison WP. MacClade 4: Analysis of phylogeny and character evolution. 2003.
107. Pritchard JK, Wen W: Documentation for STRUCTURE software: Version 2. Available from <http://pritch.bsd.uchicago.edu>. 2003.
108. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2011;4:359–61.
109. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
110. Rosenberg NA. Distruct: a program for the graphical display of population structure. *Mol Ecol Notes*. 2003;4:137–8.
111. Gante HF, Doadrio I, Alves MJ, Dowling TE. Data from: Semi-permeable species boundaries in Iberian barbels (*Barbus* and *Luciobarbus*, Cyprinidae). *Dryad Digit Repos* 2015;<http://dx.doi.org/10.5061/dryad.m1572>.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

