Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna

Veronika Bartáková\textsuperscript{1,2}, Martin Reichard\textsuperscript{1}, Radim Blažek\textsuperscript{1}, Matej Polačik\textsuperscript{1} and Josef Bryja\textsuperscript{1,2*}

\textsuperscript{1}Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno 603 65, Czech Republic, \textsuperscript{2}Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno 611 37, Czech Republic

ABSTRACT

Aim We compared the genetic variability and phylogeographical structure of three sympatric clades of annual killifishes (the \textit{Nothobranchius furzeri} complex, \textit{N. orthontus} complex and \textit{N. rachovii} complex) inhabiting annually desiccating savanna pools. Hypotheses on the mechanisms affecting intraspecific structure and speciation were tested.

Location Temporary pools in Mozambique (Africa).

Methods The study is based on spatially detailed samples covering the entire range of all three species complexes. A set of 12–13 microsatellites (1638 individuals, 96 populations) and cytochrome \textit{b} sequences (463 fish, 152 populations) were used as genetic markers. Phylogenetic and population genetic approaches were used to describe the spatial genetic structure and to test the respective roles of river channels and river basins on diversification.

Results Profound genetic differentiation among populations was evident; some populations located only a few kilometres apart were genetically very distinct, suggesting a significant role of genetic drift and low dispersal ability. Large rivers (Zambezi, Save, Limpopo) formed major barriers to gene flow, with minor differences among the three complexes. Further, the demographic expansion of previously isolated lineages was often limited by the river channel, and rivers were also confirmed as factors affecting speciation events. River basins and elevational gradient had a smaller, but non-negligible, role in population structuring.

Main conclusions River channels are the main barriers to gene flow in \textit{Nothobranchius} fishes. The study demonstrated low dispersal ability and congruence in the phylogeographical pattern of all three complexes. Cases where \textit{Nothobranchius} appear to have crossed river channels result from the dynamics of river morphology rather than from rare dispersal events. This conclusion is supported by simultaneous crossing events across lineages. A further division, also consistent among the three complexes, was detected between drier inland and wetter coastal areas. The phylogeographical pattern of \textit{Nothobranchius} is unique in that it combines features of both aquatic and terrestrial taxa.

Keywords Genetic structure, geodispersal, Mozambique, \textit{Nothobranchius kadleci}, \textit{Nothobranchius kuhntae}, \textit{Nothobranchius pienaari}, phylogeography, population genetics, river morphology, vernal pool.

INTRODUCTION

The population structure and diversification of freshwater fishes are primarily influenced by geomorphological changes affecting connections between river basins and lakes. The separation of water bodies often leads to diversification as a result of genetic drift or local adaptation, resulting in reproductive isolation. Similar processes are observed across different fish taxa and biogeographical regions (Bernatchez & Wilson, 1998); however, different patterns of diversification
exist among other groups of fishes. For example, lakes represent an aquatic equivalent of island habitats, providing an environment conducive to sympatric speciation (reviewed in Seehausen, 2006). Riverine fish can also demonstrate unusual diversification patterns (Markert et al., 2010), but their intraspecific structure is typically clustered by river basin (Bernatchez & Wilson, 1998; Katongo et al., 2005; Koblmüller et al., 2006).

Annual killifishes are an ecologically unique group of freshwater fishes inhabiting small temporary water pools in the savanna, often outside the active alluvia of major rivers (Reichard et al., 2009). Phylogenetic data suggest that their diversification and speciation were influenced by the expansion and contraction of savanna habitats (Dorn et al., 2014), a pattern usually associated with terrestrial animals (Lorenzen et al., 2012) rather than freshwater fish. The annual fish habitat is very fragmented, but individual pools can occasionally be connected during major floods (Watters, 2009). Other inhabitants of temporary aquatic pools are cladocerans, anostracans, aquatic hemipterans, larval beetles, odonates and amphibians (Williams, 2006). All these taxa have dormant or desiccation-resistant stages with diverse modes of dispersal, both active (adult insects and amphibians) and passive, attached to larger animals (Frisch et al., 2007; Vanschoenwinkel et al., 2011), or blown on the wind (Brendonck & Riddoch, 1999). How annual fishes disperse is not known, but they survive the dry period as diapausing embryos encased in the dry mud substrate, making terrestrial life in the rainy season, characterized by rapid sexual maturation and a shorter period of post-hatching stage during the dry season when the eggs are encapsulated in the dried substrate, and a shorter period of post-hatching life in the rainy season, characterized by rapid sexual maturation and daily reproduction until habitat desiccation (Blázek et al., 2013). The three complexes are largely sympatric throughout their ranges (see Appendix S1 in Supporting Information). The F-complex comprises N. furzeri and N. kadleci, with limited occurrence in coastal areas and no recorded population north of the Zambezi River (Reichard, 2014).

Finally, at a large scale, climatic oscillations may have fragmented and reconnected different savanna-dominated biomes with suitable conditions and habitats to support annual fish (Cotterill, 2003).

Rivers are frequent barriers to gene flow in terrestrial taxa that may constrain or limit population expansions. Major rivers are known to drive allopatric diversification and speciation both in open (e.g. Bryja et al., 2010) and forested (e.g. Kennis et al., 2011) ecosystems. During relatively humid periods, some African rivers and their adjacent riparian forests may have considerably fragmented the savanna habitat and prevented dispersal across the river channel. In the drier periods the savanna habitats and their associated fauna expanded but dispersal was still often prevented by wide river channels. There is increasing evidence for such evolutionary dynamics among diverse terrestrial animals with low dispersal and swimming ability, including rodents (Bryja et al., 2010), small antelopes (Cotterill, 2003), primates (Zinner et al., 2009), and dung beetles (Sole et al., 2005). A pilot study focusing on a single species in a limited geographical area demonstrated the potential role of rivers in structuring populations of annual fishes (Bartáková et al., 2013), providing predictions for the current analysis (Table 1).

*Notobranchius* species from southern and central Mozambique form an isolated monophyletic clade composed of three species complexes (Dorn et al., 2011, 2014). Their life cycle comprises a relatively long period of diapaused embryo stage during the dry season when the eggs are encapsulated in the dried substrate, and a shorter period of post-hatching life in the rainy season, characterized by rapid sexual maturation and daily reproduction until habitat desiccation (Blázek et al., 2013). The three complexes are largely sympatric throughout their ranges (see Appendix S1 in Supporting Information). The F-complex comprises *N. furzeri* and *N. kadleci*, with limited occurrence in coastal areas and no recorded population north of the Zambezi River (Reichard,
2010). The R-complex comprises three parapatric species – *N. pienaari*, *N. rachovii* and *N. krysanovi* – and, compared with the other complexes, is the most common in coastal plains despite its occurrence along the entire elevational gradient. The O-complex, which includes *N. orthonotus* and putative *N. kahntae* (whose taxonomic status is unclear), has the widest distribution, and is the least abundant species in the *Nothobranchius* community (Reichard et al., 2009; Poláčik & Reichard, 2010; Dorn et al., 2011).

In this study we analysed comprehensive genetic data (using both mitochondrial and nuclear markers) from almost the entire distribution of all three species complexes forming the southern clade of *Nothobranchius* (sensu Dorn et al., 2014) to answer following questions. (1) Is the genetic structure of the three species complexes spatially concordant? (2) What are the main barriers affecting gene flow in the relatively homogeneous savanna habitat? (3) Do major rivers (flowing mostly north-west to south-east) fragment the population structure of *Nothobranchius* fishes or is the phylogeographical pattern consistent with another mode of dispersal? We predicted concordance in spatial genetic structure among the different complexes if river dynamics are the main driver, and discordance if dispersal is mainly stochastic aquatic (flooding), aerial (wind), or animal-mediated (Table 1).

**MATERIALS AND METHODS**

**Sampling**

Samples were collected during eight field trips between 2008 and 2012 across southern and central Mozambique (Appendix S1), encompassing almost the entire range of the study taxa (Wildekamp, 2004). Fish were collected using dip and seine nets. Species were identified directly in the field on the basis of body shape and coloration (Wildekamp, 2004). Small fin clips from the caudal fin were taken from adult fish and stored in 96% ethanol. Most fish were released but a subsample was retained and returned to the laboratory. Two specimens of *N. furzeri* from the type locality (Gonarezhou National Park, Zimbabwe) were obtained from the breeding stock in Leibniz Institute for Age Research, Jena, Germany. Based on previous studies (Shidlovskyi et al., 2010; Dorn et al., 2011, 2014) and our recent phylogenetic reconstructions (see below), the material was divided into three species complexes that were treated separately in all subsequent analyses. In total, we used samples of the F-complex from 50 pools (= localities, populations), the O-complex from 62 pools, and the R-complex from 46 pools (Table 2, Appendix S1).

**Genotyping of microsatellites and mtDNA**

To analyse the genetic structure, we used a combination of nuclear (12–13 microsatellites) and mitochondrial (sequences of the gene for cytochrome *b*; *CYTB*) markers. The number of individuals analysed for each marker is shown in Table 2.

**Table 2** The number of populations (*N* pops) and individuals (*N* inds) of the three species complexes of *Nothobranchius* analysed for each marker. The number of microsatellite markers (*N* loci) and length of the analysed alignment (bp) are also indicated.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Microsatellites</th>
<th>CYTB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>N</em> pops</td>
<td><em>N</em> inds</td>
</tr>
<tr>
<td>F-complex</td>
<td>41</td>
<td>826</td>
</tr>
<tr>
<td>O-complex</td>
<td>29</td>
<td>416</td>
</tr>
<tr>
<td>R-complex</td>
<td>26</td>
<td>396</td>
</tr>
</tbody>
</table>

Full details of the genotyping methods, including primer sequences, details of microsatellite multiplexing, and PCR protocols are given in Appendix S2.

**Data analysis**

**Phylogenetic analysis of the genus Nothobranchius in Mozambique**

Sequences of *CYTB* were edited and aligned in SeqScape 2.5 (Applied Biosystems, Foster City, CA, USA), producing a final alignment of 780 bp. We used jModelTest 2.1.6 (Darriba et al., 2012) to identify the most appropriate substitution model out of 24 models evaluated using the corrected Akaike information criterion and Bayesian information criterion. Both criteria indicated the general time reversible model with a gamma-distributed rate variation across sites (GTR+G) as the model best fitting the data. Two sequences of the genus *Aphyosemion*, GenBank accession numbers DQ522263 (*A. bivittatum*) and EU885235 (*A. herzogi*), were used as outgroups in all phylogenetic analyses.

The phylogenetic relationships were inferred using maximum likelihood (ML) and Bayesian (BI) approaches. ML analysis was performed in RAxML 8.0 (Stamatakis, 2014), applying the GTR+G model and the default bootstrap procedure with 1000 replications. Bayesian analysis of evolutionary relationships was performed in MrBayes 3.2.1 (Ronquist & Huelsenbeck, 2003). Three heated and one cold chain were employed in all analyses, and runs were initiated from random trees. Two independent runs were conducted with two million generations per run; and trees and parameters were sampled every 500 generations. Convergence was checked using Tracer 1.5 (Rambaut & Drummond, 2007). For each run, the first 30% of sampled trees were discarded as burn-in. Bayesian posterior probabilities were used to assess branch support of the Bayesian tree. The final trees were visualized in FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

**Distribution of mtDNA variation and historical demography**

Within-complex sequence variation of *CYTB* was visualized as a haplotype network using the median-joining algorithm.
in Network 4.610 (Bandelt et al., 1999). All sequences were geo-referenced and the distribution of the haplogroups was plotted onto a map using PanMap software (http://www.pangaea.de/software/PanMap).

To analyse historical demography from mitochondrial data, a reduced dataset was used. Populations where recent secondary contact was detected, populations with unclear assignment to a particular population cluster, and some peripheral genetically distinct populations (probably due to intensive drift) were not included in this analysis because it requires homogeneous populations. Groups of populations were defined on the basis of structure (see below) and haplotype network analysis. Specifically, 172 sequences (687 bp long) of the F-complex, 126 sequences (741 bp) of the O-complex and 106 sequences (779 bp) of the R-complex were used. The list of populations and their assignment to particular groups is detailed in Appendix S1. Diversity estimates, i.e. number of polymorphic sites (Np), number of haplotypes (Nh), haplotype diversity (Hd), nucleotide diversity (π), average number of nucleotide differences (k) and Watterson’s estimate of θ (θ = 4Neμ), were calculated using Dnasp 5.00.04 (Librado & Rozas, 2009).

Demographic histories were evaluated using the neutrality indices, Tajima’s D and Fu’s Fs, sensitive to population size changes (Fahey et al., 2014) in Dnasp (Librado & Rozas, 2009). They return significantly negative values in the case of recent population expansion, while population decline and/or structuring tend to return positive values.

Bayesian clustering and genetic diversity – microsatellite data

An individual-based Bayesian clustering procedure implemented in Structure 2.3.3 (Hubisz et al., 2009) was applied to detect the best genetic structuring among samples. The Bayesian model assumes K (unknown) populations with different allele frequencies at a set of independent loci. The calculation was performed for K from 1 to 15, separately for all three species complexes. The analysis was run with 20 independent simulations for each K, with 10\(^6\) iterations following a burn-in period of 10\(^5\) iterations. In all simulations, an admixture ancestry model and correlated allele frequency model (with λ = 1) were used. The likelihood of K (ln Pr(X|K)) was used to assess the number of real populations in the datasets using the method of Evanno et al. (2005). The results of 20 replicate runs for each value of K were combined using the Greedy algorithm of clump 1.1.2 (Jakobsson & Rosenberg, 2007). Individual barplots for each value of K were displayed graphically using Distruct 1.1 (Rosenberg, 2004) and the Q-values for each population were mapped for the most informative K model.

The rarefaction method in Fstat 2.9.3.2 (Goudet, 2001) was used to calculate allelic richness (AR) standardized for 5, 8, and 11 individuals (i.e. the lowest sample sizes) for the F-, O- and R-complexes, respectively. For this analysis, populations with less than five sampled individuals (P240 and P262 populations from the F-complex, P256 and P559 from the O-complex and P530 from the R-complex) and the inbred captive population (GRZ from the F-complex) were not used. The AR of the R-complex was estimated from only seven microsatellite loci, two (Nfu_0023_FLI and Nfu_0027_FLI) were excluded because of high polymorphism (> 100 alleles, which can bias the estimate of mean AR) and three (Nfu_0006_FLI, Nfu_0029_FLI and Nfu_0041_FLI) because of high frequency of missing genotypes in several populations.

Spatial genetic structure and testing the role of rivers

Genetic differentiation between the populations examined was quantified by computing pairwise estimators of \( F_{ST} \) according to Weir & Cockerham (1984) and their significance was tested with 1000 permutations in genetix 4.03 (Belkhir et al., 1996–2004). The captive population of N. furzeri (GRZ) was excluded for all \( F_{ST} \) analyses. Isolation by distance in the whole complex was analysed by regressing pairwise estimates of \( F_{ST}(1 - F_{ST}) \) against ln (distance between sample sites). Mantel tests were used to test the correlation between matrices of genetic differentiation and geographical distances between sampling sites with 1000 permutations in Genepop 4.0.10 (Rousset, 2008), after exclusion of loci with more than 100 alleles (not permitted in Genepop) and population 240 of N. kadleci (distance to population 518 less than 1 km).

To visualize the spatial structure in microsatellite data, we employed the individual-based spatial approach of Bayesian clustering implemented in Baps 4.1 (Corander et al., 2008). Baps estimates the hidden population substructure by clustering individuals into genetically distinguishable groups based on allele frequencies and linkage disequilibrium. We performed 20 independent runs examining the spatial clustering of groups of individuals (collected at the same locality). Based on the initial results, we tested specific hypotheses by repeating the analyses for a fixed number of populations (i.e. fixed-K spatial clustering of groups of individuals).

In the next step we specifically tested the role of large rivers (Limpopo, Save, Bazi, Pungwe and Zambezi) as important barriers to gene flow. These rivers flow perpendicular to the coast, principally in a north-west to south-east direction. We did not test the role of the Chefu River, because it is a relatively small intermittent stream (only 2–4 m wide in the region where the Nothobranchius populations were located), with a large portion of the river bed dry for much of the year. We selected geographically close populations from both sides of the river and performed a hierarchical analysis of molecular variance (AMOVA; Excoffier et al., 1992), based on allele frequency information at both microsatellites and CYTB, using Arlequin 3.1 (Excoffier et al., 2005) with 10\(^4\) permutations. The molecular variance was partitioned among groups, among
populations within groups and within populations. Populations were grouped according to side of the river for each species complex. Fixation indices $F_{ST}$, $F_{CT}$ and $F_{SC}$ were calculated and their significance was assessed with 10,000 bootstraps (Excoffier et al., 1992).

RESULTS

Phylogenetic relationships between species complexes

Phylogenetic analysis of 200 unique CYTB sequences produced a highly structured but only partially resolved tree (Fig. 1), suggesting three phylogenetic lineages corresponding to the three previously described species complexes (Dorn et al., 2011). The F-complex is composed of *N. kadleci* and *N. furzeri*, the O-complex is represented by *N. orthonotus*, including a putative species *N. kunhtae*, and the R-complex includes the well-supported and distinct *N. rachovii*, *N. pienaari* and *N. krysanovi* (Shidlovskyi et al., 2010). The monophyly of the F- and R-complexes is strongly supported, while that of the O-complex was not (but see significant support in Dorn et al., 2011 on the basis of multi-gene analysis), yielding the position of the putative *N. kunhtae* unresolved. The R-complex is revealed as a sister lineage to the remaining two *Nothobranchius* lineages. The splits of main lineages within all three complexes are often determined by the presence of large rivers (Fig. 1).

Genetic diversity within species complexes: mitochondrial data

There was a strong spatial population structure in all three species complexes. Within all complexes, the morphologically defined species had a largely parapatric distribution. Intraspecific clades were geographically structured, and secondary contacts were often observed (Fig. 2, Appendix S3).

In the F-complex, 69 unique haplotypes split in two main haplogroups, consistent with species delineation (Fig. 2a). In *N. furzeri* (see also Bartáková et al., 2013), three main
haplogroups with strong geographical structure were confirmed (Chefu basin, south and north of the Limpopo River) with a single population (P406) showing secondary contact between two haplogroups north of the Limpopo River (Fig. 2b). In N. kadleci, four haplogroups were detected. The river Save forms a barrier to dispersal, with a single population recorded south of the river (P226) forming a separate haplogroup. The Buzi and Pungwe rivers also partially separated different haplogroups, but appear partly permeable (Fig. 2b).

In the O-complex, 79 unique haplotypes were primarily divided into two major groups (Fig. 2c), completely separated by the river Buzi (Fig. 2d). South of the Buzi there were three distinct lineages. The first is unique to the southernmost population south of the Incomati River. The second haplogroup is distributed on both sides of the Limpopo and in the Chefu basin. The third haplogroup is strictly limited by the Buzi in the north but only weakly by the Save in the south; the haplotypes of this group were also detected in populations on the southern bank of the Save (P113, P226, P223), and even introgressed into populations located 120 km (P418) and 340 km (P551) south of the river (Fig. 2d). A visual inspection of geomorphological maps revealed that P418 and P551 are connected to P223 by a temporary stream bed, providing a flooding-related explanation for such long-distance dispersal. North of the Buzi, two haplogroups were clearly separated between coastal and inland areas, irrespective of the main Pungwe and Zambezi channels. Both northern haplogroups are present in P231 (Fig. 2d). The lack of an inland haplogroup north of the Zambezi in our dataset may be because of the unavailability of samples from Malawi, where N. orthobonotus is presumably also present (Wildekamp, 2004).

In the R-complex, the network of 57 unique haplotypes is consistent with the delineation of three species (Fig. 2e). Nothobranchius krysanovi occurs north of the Zambezi, whereas N. rachovii is found in the basins of the Buzi and the Pungwe (Fig. 2f). Further substructure is only evident in N. pienaari, the most widespread species with the largest sample available for analysis (77 sequences, 32 populations, 38 haplotypes). In the south, three separate haplogroups

Figure 2 Haplotype networks of cytochrome b sequences (a, c, e) and distribution of main haplogroups (b, d, f) in the three largely sympatric species complexes of Nothobranchius: (a, b) F-complex, (c, d) O-complex, and (e, f) R-complex. Length of branches in networks is proportional to the number of substitutions along a given branch, and circle size is proportional to haplotype frequency. Pie-chart colours indicate the relative proportions of haplogroups at particular localities. Numbers of localities correspond to those in Appendix S1; for more detail on haplotypes see Fig. S5 in Appendix S3. The distribution of currently valid described species is marked with coloured lines.
occur in the Limpopo basin, the Chefu basin, and a coastal region outside the Limpopo basin (P81 and P509). In central Mozambique, four additional haplogroups are separated, predominantly by river channels. One haplogroup is only found in two populations south of the Save (P113, P226), one occurs between the Save and Buzi (P105, P268, P511), and one is endemic to the Beira region north of the Pungwe. The fourth haplogroup occurs on both sides of the Save (Fig. 2f).

The details of mtDNA variability and tests of historical demography are shown in Table 3. In the F-complex both neutrality tests indicate strong population expansion only in the *N. furzeri* clade in the Chefu basin, but Fu’s *F*$_S$ test suggests significant expansion also in the populations along the Limpopo River (especially on its south bank). In the O-complex Fu’s *F*$_S$ test suggests population expansion in all defined population groups except ‘North inland’. In the R-complex, the only partial evidence for population expansion (i.e. negative Tajima’s *D* and significantly negative Fu’s *F*$_S$) is observed in the population of *N. pienaari* distributed along the Limpopo (Table 3).

### Genetic structure within species complexes: nuclear microsatellite data

In all clades, populations were highly genetically structured and pairwise *F*$_{ST}$ was typically high and significantly different from zero (Appendix S3). Mean (± SD) pairwise *F*$_{ST}$ estimates were 0.079 ± 0.043, 0.141 ± 0.07 and 0.192 ± 0.091 in the F-, O- and R-complex, respectively. Only 12 (0.8%) pairs of (geographically close) populations had non-significant values of pairwise *F*$_{ST}$ (1000 bootstraps in *Genetix*; *P* > 0.01).

Bayesian clustering of populations in *structure* confirmed strong genetic fragmentation in all three species complexes (Fig. 3; barplots of sequential hierarchical separation for *K* between 2 and 15 groups in Appendix S3). In the F-complex the best model was for *K* = 2, but other suitable models included *K* = 3, 8 and 10 (Appendix S3). The geographical distribution from the model for 10 clusters is shown in Fig. 3a. Spatial clustering of *N. furzeri* populations has been reported in Bartáková *et al.* (2013); see also Appendix S3. Two major groups of *N. kadleci* appear to be separated (although not completely) by the river Buzi. In contrast (and contrary to the mtDNA data), there is no evident role of the Save on population differentiation in *N. kadleci*.

The highest support in the O-complex was for *K* = 3, although models with *K* = 4, 6 and 7 were also adequate. The distribution of genetic clusters in a model for *K* = 7 is shown in Fig. 3b. In accordance with mtDNA, the populations north and south of the Buzi belong to different genetic clusters. In the south, the populations in the Chefu basin form a separate cluster and the Limpopo separates populations on its south and north banks. Populations between the Save and Buzi are genetically distinct, but one...
population south of the Save (P226) also groups with these populations. North of the Buzi, two genetic clusters are suggested – one is distributed along the coast (brown in Fig. 3b), the other tends to occur inland (purple in Fig. 3b). There is, however, evidence of an admixture of both clusters in several populations.

In the R-complex, the model for \( K = 3 \) had the best support (Appendix S3). The results of a well-supported and more informative model for \( K = 11 \) are shown in Fig. 3c. Populations of *N. krysanovi* (red) and *N. rachovii* (purple) were very distinct, in accordance with their species delinea-

ation. Within *N. pienaari*, the pattern in the south is similar to that of the other two complexes; a separate cluster is found in the Chefu basin and the Limpopo forms a barrier separating the clusters of populations on its north and south banks. The Save also acts as an important barrier to gene flow, separating population P226 from all populations north of the river. A separate genetic group is also formed by populations around the town of Beira, on the north bank of the Pungwe.

The values of allelic richness corrected for sample size (\( AR; \) visualized as pie chart diameter in Fig. 3) were much higher in the northern populations of the O- and R-complexes than the southern, especially south of the Save. The F-complex showed a different pattern; more diverse populations were found in the centres of distribution of particular clusters, while marginal populations had much lower \( AR \) and were genetically distinct, suggesting frequent founder effects and strong genetic drift.

**Spatial genetic structure – testing the role of rivers**

The pattern of isolation-by-distance was only significant in the F-complex (Mantel test, \( P < 0.001 \)), but no association between geographical and genetic distances was found in the O- and R-complexes (Appendix S3). Closer examination indicated high \( F_{ST} \) values even between geographically close populations, suggesting that geographical distance is not the only factor affecting genetic structure in the three complexes.

The role of rivers was also apparent in the spatial genetic analysis in *baps* (Appendix S3: Fig. S7). The best models identified very deep structuring (15–19 genetic groups for the particular complexes) and marginal populations often split first from the main cluster, consistent with very limited gene flow and frequent genetic drift at the periphery of the range. The separation of larger groups of populations was generally consistent with rivers forming barriers and differen-

tiation between coastal and inland areas (primarily in the Chefu region).

The results of specific tests of the role of rivers on genetic structure in AMOVA are shown in Table 4. The Limpopo River significantly shaped genetic structure in all but a single analysis; only mtDNA structure in the R-complex was not affected by the Limpopo. Similarly, the Buzi has a significant impact on the genetic structure of the O- and R-complexes and on the microsatellite structure of the F-complex. The Pungwe affected the genetic structure of the O-complex. In contrast the Save was not a significant recent barrier (except for mtDNA structure of the R-complex) despite the fact that
it was likely to have been an important barrier in the past. The border separating the main genetic lineages has now moved southwards along with a putative river channel transformation (see Fig. 1). The Zambezi, only tested in the O-complex, was not a significant recent barrier. However, it clearly separates two species in the R-complex.

**DISCUSSION**

**The phylogeographical pattern of annual fishes is a combination of freshwater and terrestrial patterns**

All the annual fish species studied had a strong spatial genetic structure, primarily driven by genetic drift in fragmented temporary pools. At the level of species complexes, large rivers formed clear barriers to dispersal and significantly affected intraspecific structure. Within each complex, major rivers separated the main groups of populations and parapatric sister species. The main channels of these rivers are at least 100 m wide within the study area and are major permanent rivers. In contrast, smaller intermittent streams (e.g. Chelu) did not represent any barrier to dispersal. River basins did not shape phylogeographical structure. The elevational gradient along the east–west axis also frequently affected the structure of intraspecific clades, especially south of the Save River and north of the Buzi, i.e. where sampling along the elevational gradient was most detailed.

Information on phylogeographical patterns among other annual fishes is scarce. At the broadest scale, biogeographical reconstruction of the entire *Nothobranchius* genus demonstrated that diversification in African annual fish was likely to have been exclusively allopatric and shaped by geological events associated with rifting. The extant *Nothobranchius* diversity has been driven by Pleistocene climatic oscillations and periodic contractions and expansions of suitable savanna habitats (Dorn et al., 2014). Studies of another clade of annual killifish, the genus *Astrolebias* from the southern Neotropics, detected no evidence of a role for rivers in structuring their spatial genetic composition (García et al., 2012), or forming a boundary for sister species distributions (García et al., 2009). Rather, their phylogeographical pattern is associated with the dynamics of the entire river basin and is reminiscent of other freshwater fishes, including non-annual killifishes (Agnèse et al., 2006; Collier et al., 2009).
The role of an aquatic barrier to dispersal of a freshwater taxon is not unique to annual killifish. Powerful rapids apparently isolated populations of a sedentary cichlid, *Teleogramma depressum*, across small distances in the lower Congo (Markert et al., 2010) and expansions of sandy habitats act as barriers to dispersal for African lacustrine cichlids associated with rocky habitats, driving their microallopatric diversification (Koblmüller et al., 2011). In other aquatic taxa, major rivers can sometimes also reduce gene flow; the phylogeographical structure of a watersnake (*Nerodia rhombifer*) is, to some extent, affected by limited dispersal across the Mississippi River (Brandley et al., 2010). Other taxa co-occurring with annual fish in temporary savanna pools typically possess specialized dispersal stages (e.g. ephippia in Cladocera, diapores in Anostraca, flying adults in Insecta), making the role of rivers in their dispersal conceivably less important (Bilton et al., 2001). In large terrestrial animals, however, major rivers often restrict dispersal. For example the Zambezi (and its tributary, the Kafue) river channel may be more difficult for annual fish than for many terrestrial animals. Crossing a plain with a negligible slope and are therefore prone to more rapid channel dynamics. This could explain the lesser role of these two rivers as barriers to dispersal and the occurrence of an isolated group of *N. pienaari* populations to the north in the Beira region (Figs 2 & 3). Ultimately, river morphology dynamics associated with alluvial geomorphology at a small scale and geological events at a broader scale (Skelton, 2001) may be the primary mode of dispersal leading to allopatric diversification.

The role of floods was still important in annual fish dispersal. On each side of the river, genetic clusters were tightly associated with river basins. This implies that flood-related dispersal remains an important mechanism of colonization and maintenance of metapopulation dynamics at a local scale, despite being spatially constrained by the main river channels. Indeed, within the Chefu basin, spatially distant populations positioned within the same alluvium of intermittent stream were sometimes genetically similar (non-significant $F_{ST}$ for pairs of population up to 75 km apart) and most cases of mitochondrial haplotype introgression into a distant population (e.g. populations 418 and 551 in the O-complex, Fig. 2d, and 406 in the F-complex, Fig. 2b) can be readily traced to a minor, temporary stream channel.

**Comparative phylogeography – congruence and incongruence among complexes**

Several phylogeographical patterns were congruent among the three complexes. For example, the Chefu basin is inhabited by distinct genetic populations in all three complexes (though in the O-complex this was detectable only in nuclear markers). Similarly consistent is the major role of some rivers as barriers to gene flow. The Limpopo forms a border between two genetically distinct groups (at least among nuclear markers, but also among mitochondrial markers for the F-complex) and this separation is more recent than that of other major rivers. The Save is a significant barrier but has been crossed by a group of populations in each complex. These patterns suggest that genetic structuring of the population in each complex may have been shaped by the same sequence of geomorphological events, producing relatively synchronous evolutionary processes during the Pleistocene. The phylogeographical congruence among the three complexes diminishes the potential significance of rare dispersal events by birds or on the bodies of large mammals, as these would represent highly stochastic events unique to each group.

There were also some incongruences among the three complexes. The role of some rivers (e.g. the Pungwe or Buzi on the low-elevation alluvial plain) in genetic structuring varied across complexes, but this variation was relatively minor compared with the congruent pattern formed by the other rivers. The other incongruence was that the O- and R-complexes had the highest allelic richness in the north, while the F-complex was the most diverse in the southern part of the range. This suggests a relatively recent expansion of the O- and R-complexes south of the Save. Uniquely, the F-complex underwent a recent population expansion from multiple refugia. Using this information together with the phylogenetic relationships of the three complexes, we can hypothesize that the divergence separating the O- and F-complexes is a result of a vicariance event, separating the ancestors of the O-complex north of the Save and that of the F-complex south of the Save.

**CONCLUSIONS**

Populations of annual fish were highly genetically structured, indicating their low dispersal ability. All three complexes demonstrated broad congruence in phylogeographical
patterns. Rivers played an important role as barriers to dispersal. At the same time, river channels were crossed by annual fish populations on several occasions, probably simultaneously for multiple lineages, a pattern suggestive of river morphology dynamics rather than rare dispersal events. On each river bank, intraspecific groups were associated with the river drainage, suggesting that flooding is important for dispersal of *Nothobranchius* at a local scale. There was consistent differentiation between the wetter coastal region and the relatively drier inland region with a higher elevation. The highest genetic diversity was found in the flat floodplains of the lower reaches of the Buzi and Pungwe rivers where river morphology is likely to be most dynamic. The pattern of isolation by distance was weak, arguing against animal-mediated egg dispersal. Allelic richness was highest in the north in two complexes but in the south in the third, suggesting two different scenarios for colonization of the current range. Further study should test alternative evolutionary scenarios in all three complexes using explicit phylogeographical simulations and Bayesian approaches.

**ACKNOWLEDGEMENTS**

Financial support came from the Czech Science Foundation (P506/11/0112). We thank M. Vrtilek and P. Vallo for help in the field, R. Poláčková and A. Bryjová for help in the design of genotyping, and Rowena Spence, Lei Fumin and three anonymous referees for comments on the manuscript. Sampling complied with the legal regulations of Mozambique (collection permit DPPM/053/7.10/08 and export permit 013/MP/2008 of the Ministry of Fisheries).

**REFERENCES**


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Overview of analysed populations from the three species complexes of *Nothobranchius*.

**Appendix S2** Detailed protocols for genotyping of microsatellites and mtDNA.

**Appendix S3** Additional information to population-genetic analysis.

**BIOSKETCH**

This study is a part of our long-term project on the ecology and evolution of African annual fishes. The paper forms part of the MSc thesis of V.B., supervised by J.B. and M.R., who all have a general interest in the evolutionary processes shaping African biodiversity, using fish and mammals as model taxa.

Author contributions: M.R. and J.B. conceived and designed the study; all authors collected samples; V.B. conducted genotyping; V.B. and J.B. analysed the data; and V.B., M.R. and J.B. wrote the paper.

Editor: Fumin Lei