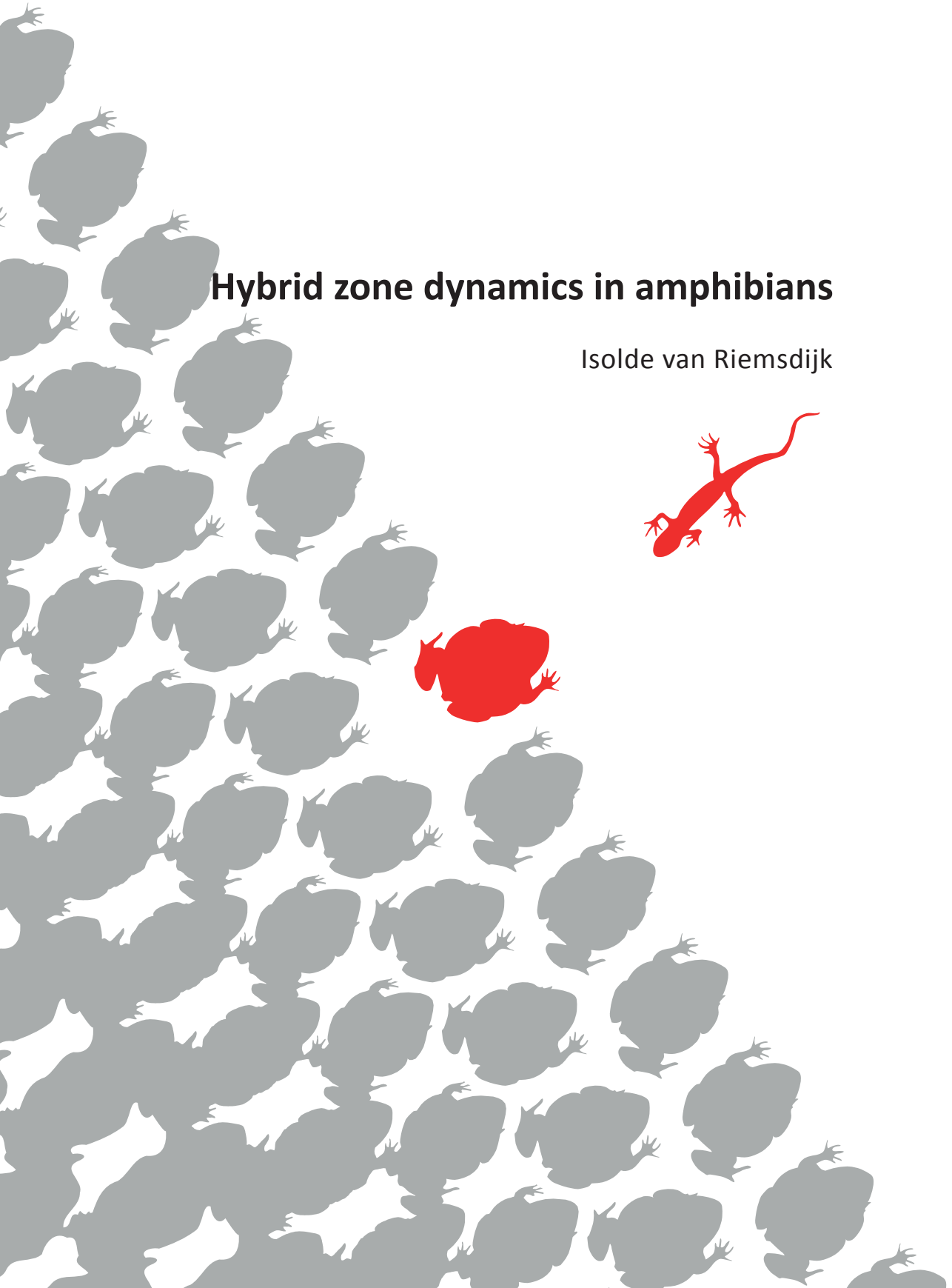
Hybrid zone dynamics in amphibians

Isolde van Riemsdijk



Hybrid zone dynamics in amphibians

Isolde van Riemsdijk

Isolde van Riemsdijk, 2019, Hybrid zone dynamics in amphibians
PhD Thesis Leiden University, the Netherlands
ISBN: 978-94-6380-475-2
Cover design & layout: Isolde van Riemsdijk
Printed by: ProefschriftenMaken

This research was carried out at Naturalis Biodiversity Center and the Institute of Biology Leiden (IBL) at Leiden University. This research was funded by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). Part of this project was carried out at the Shaffer laboratory in Los Angeles, at the University of California, sponsored by the Leiden University Fund / Swaantje Mondt Fonds (D7102).

Hybrid zone dynamics in amphibians

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 11 september 2019
klokke 13:45 uur

door

Isolde van Riemsdijk
geboren te Gouda, Nederland
in 1991

Promotor:	Prof. Dr. Menno Schilthuizen, *Leiden University*
Copromotores:	Dr. Pim J. W. Arntzen *Leiden University*
	Dr. Ben M. Wielstra *Leiden University*
Promotiecommissie:	Prof. Dr. Gilles P. van Wezel (voorzitter) *Leiden University*
	Prof. Dr. Nicole J. de Voogd (secretaris) *Leiden University*
	Prof. Dr. Roger K. Butlin *University of Sheffield*
	Dr. Katja T. C. A. Peijnenburg *University of Amsterdam*
	Prof. Dr. H. Bradley Shaffer *University of California, Los Angeles*
	Dr. Klaas Vrieling *Leiden University*
	Dr. Maurijn van der Zee *Leiden University*

Voor Keetje

Contents

Introduction

1

Phylogeography of secondary contact zones

Speciation is the process during which one lineage (or species) becomes two lineages (Darwin, 1859). The formation of new species is often classified into allopatric, parapatric, and sympatric speciation. These are parts of the speciation continuum, starting on the side of allopatric speciation with complete genetic isolation by an external barrier (Butlin, Galindo, & Grahame, 2008). An example of an external barrier is a mountain range which impedes dispersal for a low-land dwelling lineage. The continuum ends with sympatric speciation, which takes place through genetic isolation by an internal barrier, such as speciation by genome duplication. The process of allopatric speciation starts with a split in the distribution of a lineage by a physical barrier, followed by an increase in genetic differentiation through processes such as mutation, natural selection and genetic drift (Mayr, 1942; Lande, 1980; Wu & Ting, 2004). Allopatric speciation is completed by reproductive isolation, when upon renewed contact of the two diverged lineages (secondary contact) the exchange of genetic material is inhibited (Rice, 1998).

Allopatric speciation was thought to be the most common type of speciation in nature (Mayr, 1982; Coyne & Orr, 2004). Recent models are more nuanced in the paths that speciation may follow. Ideally, speciation models allow the continuous variation of the three components; spatial distribution of the diverging lineages, genetic diversity, and reproductive isolation (Butlin et al., 2008). When strictly following the biological species concept, two lineages can only be considered two species if no exchange of genes (gene flow) exists between them (Dobzhansky, 1970; Ashlock & Mayr, 1991). However, many biologists define species (and subspecies) as diverged lineages with reduced fitness in hybrid offspring (Coyne & Orr, 2004; Gavrilets, 2004), as will I in this thesis.

One excellent place to study the process of speciation is in hybrid zones (Hewitt, 1988), and recent developments in the field of biology and computer science are offering new insights in hybridisation and speciation processes (Martin & Jiggins, 2017). In a hybrid zone, intermediates between two species are found, which may show a considerable reduction in fitness (Mallet, 2005). The classic model of the origin of genetic incompatibility is the Dobzhansky-Muller incompatibility model (Muller, 1942; Dobzhansky, 1947; Turelli & Orr, 2000; Bateson, 2002). During the isolated stage of speciation the two genomes evolve independently, and once the two species reproduce during secondary contact, two new variants of genes may be incompatible. As a result, the offspring of mixed ancestry is less fit than individuals of either parent species. The deleterious effect of the alleles involved in the Dobzhansky-Muller incompatibility may be expressed both in pre- or postzygotic life stages.

More general than Dobzhansky-Muller incompatibilities are the effects of barrier genes, which are defined as “positions in the genome that contribute to barriers to gene flow between populations” (Ravinet et al., 2017). These may include genes under divergent ecological selection, divergent mate choice, or post-zygotic isolation, and can be neutral within a population. In the early stages of secondary contact between two species, barrier genes may not have the same genomic and geographic distribution.

As time passes, the occurrence of barrier genes starts to coalesce, either because selection on individual barrier genes or on population processes favours existing barrier effects to co-occur, or as an indirect response to avoid cost of hybridisation (Bierne, Welch, Loire, Bonhomme, & David, 2011; Butlin & Smadja, 2017). The barrier effect can then be defined as "a reduction of effective migration rate relative to actual migration between populations that occurs at the barrier locus (i.e. the direct effect) but can also extend beyond it (i.e. the indirect effect)" (Ravinet et al., 2017).

Hybrid zones can be formed in numerous situations. An example is the genetic isolation of taxa in refugia during glaciations, and subsequent dispersal which leads to secondary contact zones (Hewitt, 2004, 2011a). In Europe, the Iberian and Italian peninsulas are often found to have provided refugia during glaciations (Gómez & Lunt, 2007; Feliner, 2011). Subsequent dispersal during the current interglacial has led to the formation of hybrid zones throughout Europe (Hewitt, 2011b). Much of the current theory and methods for hybrid zone analyses are based on two well-studied hybrid zones from Europe, the hybrid zone between two subspecies of the house mouse, *Mus mus musculus* and *M. m. domesticus* (Forejt, 1996), and between two species of the fire--bellied toad, *Bombina bombina* and *B. variegata* (Arntzen, 1978). For regions studied less than Europe, hybrid zones are starting to emerge whilst numbers of publications increase. The Anatolian coasts, for example, provided refugia for several terrestrial species during glacial periods, and hybrid zones on the crossings of routes of dispersal from different refugia are common (Bilgin, 2011).

Two types of hybrid zones can be classified. First, in mosaic hybrid zones, one species outcompetes the other, depending on which species has a higher fitness under the local conditions, resulting in a fragmented transition from one species to the other (Harrison & Rand, 1989). In tension zones, the transition of one species to the other is not patchy, but the hybrid zone is narrow (Moore, 1977; Barton & Hewitt, 1985). In the classic literature, hybrid zones are thought to stabilise soon after formation, for example at an ecological barrier to dispersal or where population density is low (Endler, 1977; Barton & Hewitt, 1985). In Australia the existence of a tension zone between two subspecies of the red-backed fairy-wren, *Malurus melanocephalus melanocephalus* and *M. m. cruentatus*, is well supported by multiple lines of evidence. First, a narrow hybrid zone with intermediates of both subspecies was studied using genetic markers, revealing a steep transition from one subspecies to the other (Baldassarre, White, Karubian, & Webster, 2014). Second, males are more likely to respond to their own subspecies' song and the change in subspecies' song is strongly correlated to the change in the subspecies' genetic markers (Greig & Webster, 2013; Baldassarre et al., 2014). Third, the location of the hybrid zone coincides with the Carpentarian barrier, where vegetation cover is low and floods frequently occur (Chivas et al., 2001). The population density of red-backed fairy-wrens is low in this area (Baldassarre et al., 2014). Together, these properties make the *Mularus* hybrid zone a typical example of a tension zone.

An increasing amount of literature supports the idea that hybrid zone movement is more common than previously assumed (Arntzen & Wallis, 1991; Buggs, 2007; Roy, O'Connor, & Green, 2012; Leaché, Fujita, Minin, & Bouckaert, 2014; Taylor et al., 2014; Wielstra, Burke, Butlin, & Arntzen, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017; Ryan et al., 2018). Hybrid zone movement can be seen as the replacement of one species

by another, caused by a difference in fitness between the two, through a competitive advantage, asymmetric hybrid fitness effect, or environmental change (Buggs, 2007). The most reliable method of detecting hybrid zone movement is to monitor a hybrid zone at multiple moments in time, and record the movement that has occurred (Buggs, 2007). For some species, for example if they have long generation times or disperse slowly, observing hybrid zone movement over time may not be feasible. Secondly, evidence of hybrid zone movement can be found in the presence of enclaves, or relict populations (e.g. Wielstra et al., 2017a). Enclaves are patches of one species remaining present after another species has replaced the former in the surrounding area. The classic example of enclaves supporting hybrid zone movement can be found in Central and Eastern Europe, where the mountain-dwelling *B. variegata* is surrounded by the lowland species *B. bombina* (Arntzen, 1978). In other cases, introgression may also provide evidence of hybrid zone movement, but they are hard to distinguish from introgression of positively selected alleles (Buggs, 2007). To be able to distinguish the effect of hybrid zone movement from the effect of positive selection I first discuss how hybrid zones can be described using genetic data, and follow up with the hypotheses available to distinguish introgression from hybrid zone movement from positive selection.

Describing hybrid zones using genetic data

To be able to follow the transition from one species to the other, species specific (or diagnostic) genetic markers are used (Haldane, 1948; Endler, 1977). These markers ideally have one variant in one species, and another in the other species. The first generation offspring of parents of both species then receives one allele of each parent species, and therefore is heterozygote for any species specific marker. Backcrosses will be heterozygote for some of these markers and homozygote for others. Measures and tests for introgression all involve assumptions which influence the interpretation, therefore measures and tests should be interpreted with caution (Martin & Jiggins, 2017). I thus combine several approaches. To investigate patterns of introgression, I use geographical clines (1). To investigate patterns of recombination between the two species, I use admixture linkage disequilibrium and subsequent calculation of effective selection against hybrids (2). To investigate barrier loci and other types of outlier behaviour in the larger dataset, I use Bayesian genomic clines (3).

To start, geographical clines are fitted to the species specific allele frequencies obtained from populations in a transect (Fig. 1.1a), and describe the transition from one species to the other (Schmickl, Marburger, Bray, & Yant, 2017). Such clines will be informative of processes of selection and gene flow (Barton & Hewitt, 1985, 1989; Szymura & Barton, 1986; Mallet et al., 1990; Barton & Gale, 1993; Porter, Wenger, Geiger, Scholl, & Shapiro, 1997). Based on cline position and cline shape, two types of cline variation are distinguished. Coincident clines share a cline centre, but do not necessarily share the same cline shape, whereas displaced clines have a cline centre away from the majority of the clines for other loci (Fig. 1.1b). Concordant clines are similar in shape (width and tail shape; Szymura and Barton, 1991), but do not necessarily share the same cline centre, whereas discordant clines have a shape different from the majority of the clines (Fig. 1.1c). Concordant and narrow clines (steep at the centre and with distinct tails) suggest that the hybrid zone is maintained

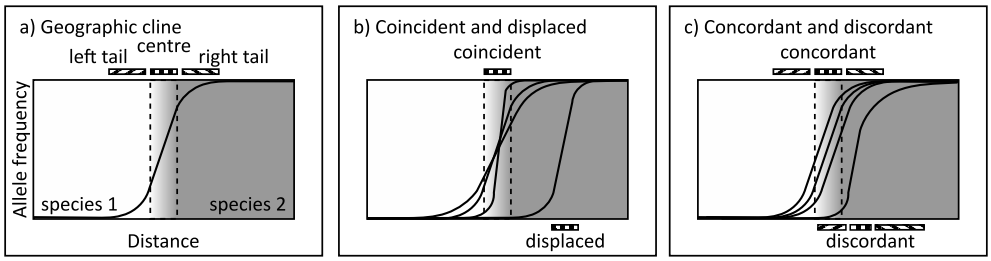


Figure 1.1: Geographical cline (a) with distance on the x axis, allele frequency on the y axis, species 1 on the left, and species 2 on the right. Left tail, centre and right tail are indicated by bars above the panel. The hybrid zone centre is indicated with dashed lines. Coincident clines (b) share the same cline centre. A displaced cline has a cline centre away from the majority of cline centres. Concordant clines (c) have the same shape. A discordant cline has a cline shape different from the majority of clines.

by a balance between strong selection and dispersal, as opposed to displaced and wide clines indicating that selection against hybrids is low (shallow slope without distinct tails; Barton and Gale, 1993). To summarise the behaviour of many geographic clines, one can use the expected frequency, which is based on a cline fitted to the average allele frequency across the available loci. However, it has been shown that genetic drift displaces clines, alters their width, and distorts their shape (Polechová & Barton, 2011).

Second, patterns of linkage disequilibrium and subsequent calculation of effective selection against hybrids show evidence of recent changes in the hybrid zone dynamics and the 'amount' of reproductive isolation of the two species. When using species diagnostic markers, in first generation offspring the linkage disequilibrium will be complete, as no crossing over of any of the species specific alleles has happened during the meiotic phase. The type of linkage disequilibrium caused by the combination of two divergent genomes is referred to as admixture linkage disequilibrium. During subsequent backcrosses, linkage disequilibrium can be broken down by cross overs (Barton & Gale, 1993; Baird, 2015). By using the admixture linkage disequilibrium in the centre of the hybrid zone, one can also calculate the average effective selection, which is the selection pressure on a locus at the zone centre due to direct selection or association with other loci under selection (Barton & Gale, 1993). In other words, this measure is indicative of the effect that barrier loci have on the ability to introgress for other loci in the hybrid zone.

Third, Bayesian genomic cline analyses can be used to study genome-wide variation of introgression among admixed individuals (Gompert & Buerkle, 2011, 2012; Gompert, Parchman, & Buerkle, 2012). The Bayesian genomic cline model uses the probability that an individual with a certain hybrid index inherited a gene variant at a locus from species one (denoted φ) and compares this to the hybrid index (Fig. 1.2a). The hybrid index can be simplified as the proportion of alleles of a population belonging to each species, and can be calculated using the following equation:

$$\text{Hybrid index} = p/N_{allele}$$

With p the frequency of the allele variants specific of one of the species across all available loci, and N_{allele} the number of alleles sampled (which translates to two times

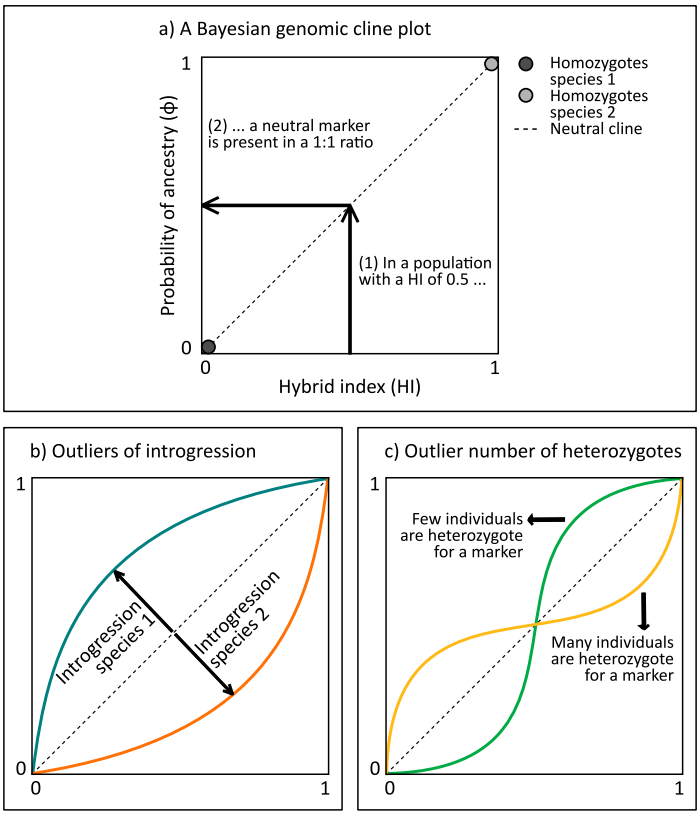


Figure 1.2: A Bayesian genomic cline plot (a) with on the x axis the hybrid index (HI), and on the y axis the probability of ancestry (φ). The arrows in panel a show that in a population with a hybrid index of 0.5, a neutral marker is expected to be present in a 1:1 ratio (frequency p=0.5). For outliers of introgression (α), the cline is moved away from the neutral expectation (b). When α > 0, introgression from an allele of species 2 into species 1, and when α < 0 introgression of markers of species 1 into species 2. A marker can be an outlier for the number of heterozygotes present for said allele when β deviates from 0 (c). When β > 0 for a marker, there are fewer individuals heterozygote for a marker than should be based on their general genetic background, whilst if β <0, there are more individuals heterozygote for a marker than there should be based on their general genetic background.

the number of (diploid) individuals sampled). In a genomic cline analysis, a population of pure individuals of species 1, with only markers diagnostic of species 1, would have both a hybrid index and a probability of ancestry of 0. A pure population of species 2, with only markers diagnostic of species 2, would have both a hybrid index and a probability of ancestry of 1. For a population of individuals with a hybrid index of 0.5, any marker would be expected to be present in a 1:1 ratio (frequency of 0.5), and any neutral marker would thus yield a probability of ancestry of 0.5. The expectance for a neutral marker is thus a diagonal cline across the plot (Fig. 1.2a, dashed line).

The parameters α and β can be used to describe the shape of the genomic cline, and can be used to detect certain outlier behaviours. The probability of ancestry of species 1 relative to expected, or the hybrid index, is described by cline parameter α (Fig. 1.2b). A positive α indicates an increase in the ancestry probability and a negative α indicates a decrease. The biological interpretation is that if α is positive (α > 0), there is introgression of the species 1 variant of the marker into species 2 compared

to the general genetic background, and if α is negative (α < 0) there is introgression of the species 2 variant of the marker into species 1 compared to the general genetic background. Cline parameter β measures the average pairwise linkage disequilibrium based on ancestry between a locus and all other loci (Fig. 1.2c). A positive β indicates an increase in the rate of transition from low to high probability of ancestry of species one as a function of the hybrid index, whereas a negative β indicates a decrease (Gompert & Buerkle, 2011; Parchman et al., 2013). The biological interpretation is that if β is positive (β > 0), there are more heterozygotes than expected, and the marker may be under positive selection when an individual is heterozygote (heterozygote advantage). If β is negative (β < 0), there are less heterozygotes than expected, and therefore this marker may be under negative selection when an individual is heterozygote (heterozygote disadvantage). Markers which are significant outliers below zero for β, are thus possibly involved in, or situated close in the genome to, the barrier effect.

Hypotheses of different hybrid zone scenarios

The null hypothesis of a tension hybrid zone is that it is stable. When a tension zone is stable, geographic clines are symmetric, concordant, and steep, and thus the expected frequency is symmetric and steep, and a peak of admixture linkage disequilibrium is situated in the centre of the hybrid zone (Fig. 1.3a). The hybrid zone between *B. bombina* and *B. variegata* displays such symmetric and steep clines in some sections of the hybrid zone, and in those sections, admixture linkage disequilibrium is positioned in the centre of the hybrid zone (Szymura & Barton, 1991).

Displaced and discordant geographic clines occur in many stable hybrid zones. Displaced and discordant clines may be caused by sex-biased gene flow or adaptive introgression (Mallet, 2005; Excoffier, Foll, & Petit, 2009; Toews & Brelsford, 2012; While et al., 2015; Sloan, Havird, & Sharbrough, 2016; Bonnet, Leblois, Rousset, & Crochet, 2017). Hybrid zone movement may also cause displaced and discordant clines when a trail of genetic material from the overtaken species remains present in the overtaking species (Rohwer, Bermingham, & Wood, 2001; Gay, Crochet, Bell, & Lenormand, 2008; Excoffier et al., 2009; Arntzen, de Vries, Canestrelli, & Martínez--Solano, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017). However, with hybrid zone movement, this introgression occurs in many physically and functionally unlinked markers, which are selectively neutral (Currat, Ruedi, Petit, & Excoffier, 2008; Wielstra, Burke, Butlin, Avcı, et al., 2017). In addition, in the leading edge of the moving hybrid zone centre a peak of admixture linkage disequilibrium is predicted, because at the leading edge, individuals with the least history of recombination are involved in reproduction (Gay et al., 2008; Wang et al., 2011). Combing these, the hypothesis of hybrid zone movement encompasses a tail of introgression in the wake of the moving zone combined with a shift of the peak in admixture linkage disequilibrium towards the leading edge of the movement (Fig. 1.3b). A southward moving hybrid zone between the gull species *Larus glaucescens* and *L. occidentalis*, was corroborated by higher levels of introgression towards the north compared to the south, and a peak of admixture linkage disequilibrium shifted towards south (Gay et al., 2008).

With the same measures, we can draw up the hypothesis of asymmetric reproductive isolation. Asymmetric reproductive isolation occurs when an

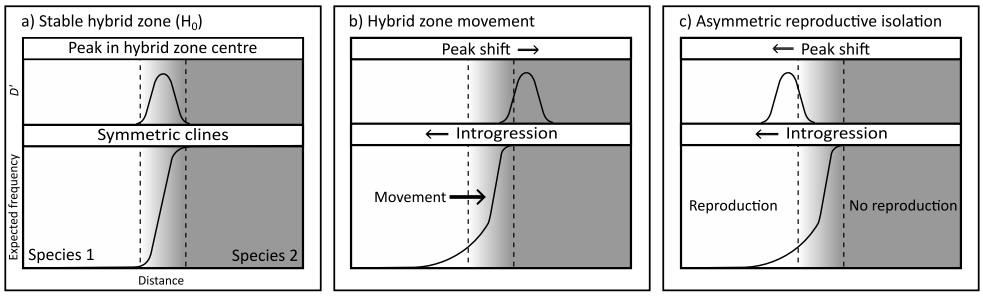


Figure 1.3: The null hypothesis of a stable hybrid zone (a), the hypothesis of hybrid zone movement (b), and the hypothesis of asymmetric reproductive isolation (c). The y-axis on the top graphs shows admixture linkage disequilibrium (D'), the y axis on the bottom graphs shows the expected frequency of species diagnostic markers. The x-axis in all plots is distance. The dotted vertical lines indicate where the current hybrid zone centre is situated.

asymmetry in the success of mating, or in the fitness of hybrid offspring, in the hybrid zone exists (Devitt, Baird, & Moritz, 2011; Johnson, White, Phillips, & Zamudio, 2015; While et al., 2015). When reproduction is more successful on one side of the hybrid zone, introgression will be higher, and the peak of admixture linkage disequilibrium will co-occur with the introgression (Fig. 1.3c). One such case is the asymmetry in the plethodontid (lungless) salamander hybrid zone (Devitt et al., 2011). Hybrid offspring nearly always possess mtDNA originating from one of the parent species, and not the other. This is convincing evidence that only first generation offspring of *Ensatina ensatina klauberi* females and *E. e. eschscholtzii* males or male hybrids can produce viable offspring, and explains the pattern of high introgression on the *E. e. eschscholtzii* side of the hybrid zone, with a coincident peak of admixture linkage disequilibrium (Devitt et al., 2011).

Amphibian moving hybrid zones and big genomes

Amphibian species play an interesting role in the study of hybrid zones. However, amphibians are hard to study compared to other vertebrates, due to their large and complex genomes. I here give an overview of a few well known studies of hybrid zone movement in amphibians, and further explain why their genomes are harder to study than those of other vertebrate species.

One of the first well-described hybrid zones is that of *B. bombina* and *B. variegata* (Arntzen, 1978), a tension zone with steep clines (Szymura & Barton, 1986, 1991). Crosses of *B. bombina* and *B. variegata* produce egg batches with higher mortality and offspring with higher frequencies of morphological abnormalities than crosses within either species (Kruuk, Jason, & Barton, 1999), confirming the existence of a tension zone. The presence of enclaves of *B. variegata* at higher altitudes, surrounded by lowlands where *B. bombina* thrives, suggests that *B. bombina* displaced *B. variegata* in the postglacial area (Arntzen, 1978).

Crested and marbled newts, *Triturus cristatus* and *T. marmoratus*, engage in a reticulate (both clinal and mosaic) hybrid zone in western France, and the zone was shown to have moved using distribution maps from different time points (Arntzen & Wallis, 1991). Continued monitoring of this hybrid zone showed that it stabilised, possibly by pond loss and decreased population density (Visser, Leeuw, Zuiderwijk,

& Arntzen, 2017). A hybrid zone between another pair of newts from the *Triturus* genus, *T. anatolicus* and *T. ivanbureschi*, was confirmed to have moved after finding asymmetric introgression of mtDNA and 50 nuclear markers (Wielstra, Burke, Butlin, Avcı, et al., 2017). Discordance between geographic clines based on mtDNA and geographic clines based on microsatellite loci showed one lineage of the spotted salamander (*Ambystoma maculatum*) replaced another lineage, possibly due to population expansion and asymmetric mate choice (Johnson et al., 2015).

Hybrid zone research in other vertebrate orders have benefited from the availability of high quality reference genomes. The house mouse hybrid zone, *Mus musculus musculus* and *M. m. domesticus*, is studied with probably the best available genomic resources of vertebrate species, allowing studies with high-resolution linkage maps (Wang et al., 2011). But there are also plenty high-resolution studies of non-model species available, e.g. using whole genome sequencing in warblers, *Setophaga coronata auduboni* and *S. c. coronate* (Brelsford, Toews, & Irwin, 2017). So what makes it so difficult to apply similar approaches in amphibians?

The largest reported variation in genome size within the class of vertebrates can be found in amphibians, with genome sizes varying from ~ 1 Gbp (the ornate burrowing frog, *Platyplectrum ornatum*) to 117 Gbp (the Gulf coast waterdog *Necturus lewisi*; Olmo, 1973; Liedtke et al., 2018). Based on a comparative quantitative analysis of factors influencing genome size, it was found that genome size has evolved gradually over time in amphibians, following a Brownian motion, with a sudden change occurring on the branch towards the Caudata (Urodela and extinct salamander-like species, Liedtke et al., 2018). The relative large size of genomes in salamanders was found to be related to a lower rate in deletions compared to insertions, high levels of long terminal repeat retrotransposons, and a relative large genic component compared to humans, mostly because amphibian genes have longer introns (Smith et al., 2009; Sun, López Arriaza, & Mueller, 2012; Sun, Shepard, et al., 2012; Sun & Mueller, 2014). In addition, sex-chromosome turnover in Ranidae (true frogs) is the highest yet observed, and sex--chromosomes in amphibians diverge fast relative to other chromosomes, because of low recombination rates in males (Jeffries et al., 2018).

These genomic differences between closely related species causes the comparability of genomes (genome synteny) to decrease relatively fast with respect to phylogenetic distance in amphibians. Therefore the use of reference genomes from even closely related species becomes problematic. To be able to sequence a large genome without a reference genome (*de novo*), one needs to be able to achieve enough read depth to cover the genome many times, and the data should allow to identify gene duplicates and stretch across transposable elements, by employing methods to obtain long read lengths (Jansen et al., 2017; Nowoshilow et al., 2018). Despite recent advances in sequencing technologies, the sequencing of whole large amphibian genomes is still costly (Smith et al., 2019).

A more practical approach to obtain genome wide information is to sequence only a part of the genome, such as is the case in reduced genome representation methods (Davey et al., 2011; Matz, 2017). Many questions about the more general evolutionary patterns in speciation may be answered using Restriction Associated DNA (RAD) sequencing. This method has two advantages. First, it can be used without prior knowledge of the genome, and second, it allows to take a random sample of the

genome (Baird et al., 2008; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Some debate around RAD sequencing exists, for example, it potentially underestimates the number of genes under adaptive selection (e.g. Hoban et al., 2016; Lowry et al., 2016). However, for more general genomic patterns in closely related species, RAD sequencing may be less prone to erroneous conclusions than other reduced genome representation methods (Andrews et al., 2016; McCartney-Melstad, Mount, & Shaffer, 2016; McKinney, Larson, Seeb, & Seeb, 2017).

Thesis outline

This thesis focusses on hybrid zone dynamics. Two amphibian species pairs with potentially moving hybrid zones were studied. Chapters two and three deal with the first pair of species. The two northern species from the genus of banded newts, *Ommatotriton nesterovi* and *O. ophryticus*, represent an understudied system in Anatolia with great potential for the existence of a (moving) hybrid zone. Chapters four and five deal with the second pair of species. The common and spined toads, *Bufo bufo* and *B. spinosus* meet in a well-studied hybrid zone in France, but previous studies employed a limited amount of genetic data. Both species pairs studied are morphologically similar (cryptic species), but can be distinguished based on genetic data.

Chapter 2

The two northern species of banded newt are morphologically similar, but appeared as separate clades in various datasets (Arntzen & Olgun, 2000; Litvinchuk, Zuiderwijk, Borkin, & Rosanov, 2005; Bülbül & Kutrup, 2013; Arntzen, Beukema, Galis, & Ivanović, 2015). These two species were not universally recognised, and it was not known what their exact species distribution was (Fig. 1.4a). The location of a potential hybrid zone was therefore also unknown. The phylogenetic relationship between the three species is also unclear, as analyses of mitochondrial DNA (mtDNA), number of rib-bearing vertebrae (NRVB), and allozyme data resulted in different hypotheses (Fig. 1.4b; Arntzen and Olgun, 2000; Litvinchuk et al., 2005; Bülbül and Kutrup, 2013; Arntzen et al., 2015).

An hybrid zone in crested newts (*Triturus anatolicus* and *T. ivanbureschi*) was found to show extensive movement in the north of Turkey, by displacement between mitochondrial and nuclear markers, and tails of introgression in the wake of the hybrid zone movement (Wielstra, Burke, Butlin, Avcı, et al., 2017). Asymmetric introgression from *Lissotriton vulgaris* nuclear genes in *L. kosswigi* could be evidence of westward movement of the hybrid zone (Nadachowska & Babik, 2009). Both these hybrid zone systems are situated in the north of Turkey, where *Ommatotriton* also occurs. It could therefore be expected that the hybrid zone between *O. nesterovi* and *O. ophryticus* also moves. The research questions dealt with in this chapter are:

Are *O. nesterovi* and *O. ophryticus* two well-diverged species?

Does a hybrid zone exist between these species, and if so, where is it situated?

The divergence of the mtDNA phylogeny showed two deeply diverged species, and allowed to clearly line out the distributions of both northern species. Despite high sequence similarity, the two nuclear markers also showed divergence between species. Although the nuclear genetic networks were not clearly reciprocally monophyletic, it did show that none of the nuclear haplotypes were shared between species, and no introgression could be detected between both species. In other words, the nuclear haplotypes perfectly matched the species distribution according to mtDNA. We thus found no hybrid populations, but we could narrow down the area where a hybrid zone could be expected to a 60 km gap between populations of both species.

Chapter 2 is published in Molecular Phylogenetics and Evolution as "The Near East as a cradle of biodiversity: a phylogeography of banded newts (genus *Ommatotriton*) reveals extensive inter- and intraspecific genetic differentiation" (2017, vol. 114, pages 73–81) by **I. van Riemsdijk**, J.W. Arntzen, S. Bogaerts, M. Franzen, S.N. Litvinchuk, K. Olgun, and B. Wielstra.

Chapter 3

A viable population of introduced individuals of *Ommatotriton* was found in an isolated group of ponds in Spain, which is far outside the natural distribution of the genus. The animals were probably released with the intention to breed and recapture them for the pet trade. From photographs with individuals from these ponds, it appeared they were intermediates of the two northern species, based on the lateral white band and disruption of the lateral white band by large blue specks (Uzum et al. in publ). The previously published mtDNA and nuDNA markers were used to determine the precise origin of a dozen individuals. The research question dealt with in this chapter is:

Are *O. nesterovi* and *O. ophryticus* able to hybridise?

We found that indeed, *O. nesterovi* and *O. ophryticus* are able to hybridise, and hybrid offspring is able to successfully reproduce, too. It is however, as of yet unknown if the two species also hybridise in nature.

Chapter 3 is published in Conservation Genetics as "Molecular data reveal the hybrid nature of an introduced population of banded newts (*Ommatotriton*) in Spain" (2018, vol. 19, pages 249-254) by **I. van Riemsdijk**, L. van Nieuwenhuize, I. Martínez-Solano, J.W. Arntzen, and B. Wielstra.

Chapter 4

The common and spined toads are morphologically similar, but genetically divergent, and were not recognised as separate species for a long time (Recuero et al., 2012; Arntzen, Recuero, Canestrelli, & Martínez-Solano, 2013). The hybrid zone, which runs from the French Atlantic Ocean coast to the French Mediterranean coast, has been investigated in multiple articles (Fig. 1.5; Arntzen et al., 2016, 2018; Trujillo et al., 2017). A pendulum (back-and-forth) movement of the hybrid zone was hypothesised

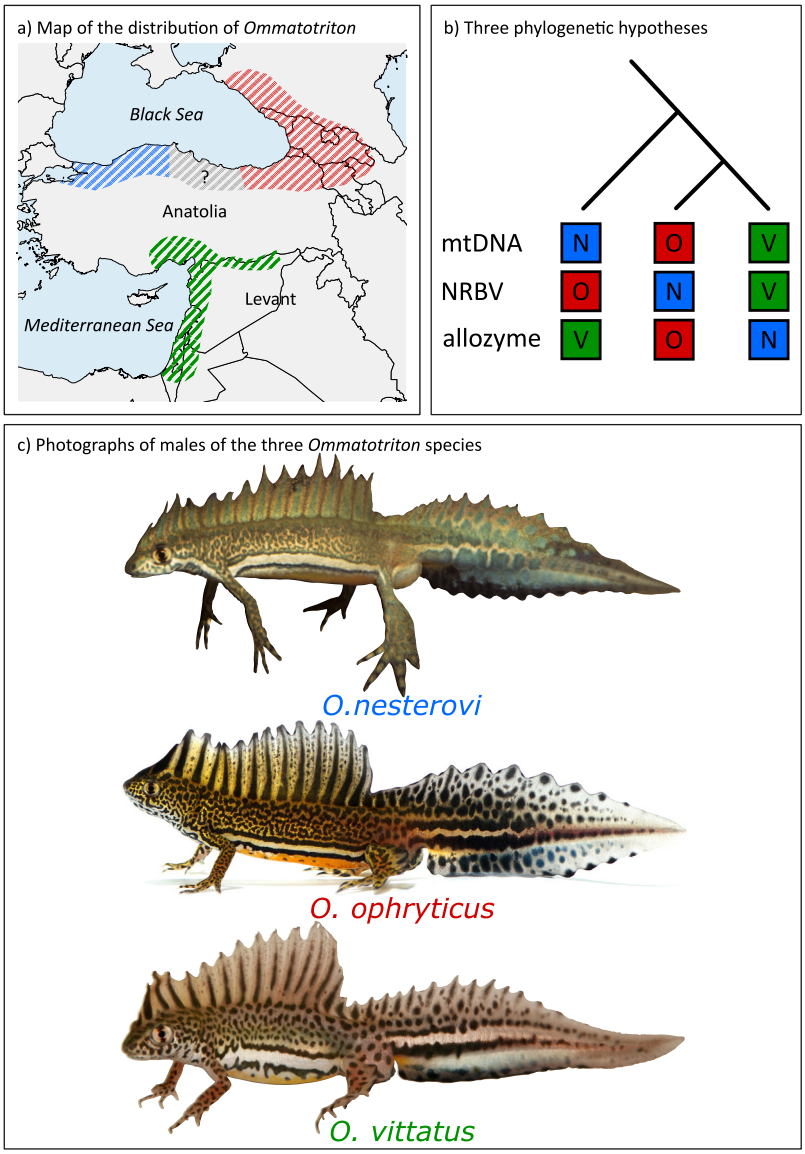


Figure 1.4: A map (a) of the distribution of the banded newt genus *Ommatotriton*, with *O. nesterovi* in blue, *O. ophryticus* in red, and *O. vittatus* in green. *Ommatotriton* is known to occur in the grey area in the north of Turkey, but it is not known which species. (b) Three different phylogenetic relationships between the three *Ommatotriton* species were supported by mitochondrial DNA (mtDNA), number of rib-bearing vertebrae (NRBV), and allozymes data. Colours indicate the three species as in the map, and the letters represent first letters of the species name. (c) Photographs of males in breeding condition of the three species of banded newt. Photographs are kindly provided by Michael Fahrbach, the late Max Sparreboom, and Sergé Bogaerts.

in the southeast section of the hybrid zone, and was based on introgression of one nuclear gene coding marker of four tested and discordance between nuclear, mtDNA and morphologic clines (Arntzen et al., 2017). In the light of this result, a previous publication on a transect in the northwest of France showed introgression towards

the north in two of four tested nuclear gene coding markers, which raised the question whether the introgression may have resulted from hybrid zone movement, too (Arntzen et al., 2016). The research questions dealt with in this chapter are:

Has the northwest section of the hybrid zone moved?

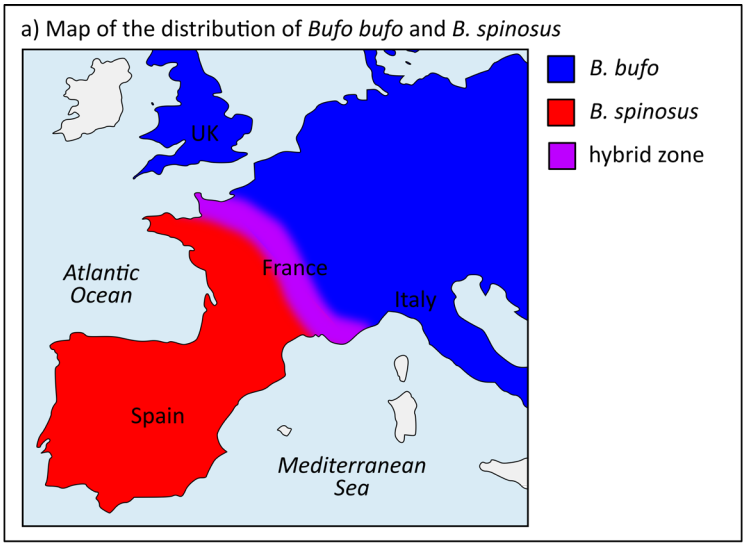
To gain resolution, the number of markers available was increased to 31 by designing genetic markers based on transcriptome data. This article treated the hypothesis of hybrid zone movement in terms of introgression and admixture linkage disequilibrium, and presented an R script for the calculation of admixture linkage disequilibrium and effective selection against hybrids following Barton and Gale (1993), which is available online. Based on literature research, I established that the expected genetic patterns of hybrid zone movement are unidirectional introgression in the wake of the hybrid zone movement and a peak of admixture linkage disequilibrium at the leading edge of the moving hybrid zone. Introgression from the southern *B. spinosus* into *B. bufo* was not significantly asymmetric, and the peak of admixture linkage disequilibrium appeared to coincide with the larger portion of introgression on the north side of the hybrid zone. I was therefore unable to establish whether the hybrid zone indeed has moved.

Chapter 4 is published in Molecular Ecology as "Testing an hypothesis of hybrid zone movement for toads in France" (2019, vol. 28, pages 1070-1083) by **I. van Riemsdijk**, R.K. Butlin, B. Wielstra, and J.W. Arntzen.

Chapter 5

Despite the lack of proof for hybrid zone movement in the northwest section of the *Bufo* hybrid zone from Chapter 4, two arguments could be made in favour of suspecting hybrid zone movement. First, a previous article hypothesised movement in the southeast section of the hybrid zone (Arntzen et al., 2017). Second, using the same markers for toad samples from England, traces of *B. spinosus* in otherwise *B. bufo* individuals were found (data unpublished), which would suggest southward movement of the hybrid zone on a geological time scale since the hybrid zone was established ~ 8,000 years ago (Arntzen et al., 2016). Replication of hybrid zone analysis in space helps identify factors driving hybrid zone movement, it allows comparison on both speed and direction of movement, and comparisons of environmental conditions and genetic impact (Buggs, 2007). The hybrid zone of *Bufo* stretches across France (Fig. 1.5a), which makes it ideal to draw independent transects. Thus, in the next study we included two transects, one in the northwest of France, and one in the southeast of France.

The hypotheses of hybrid zone movement are based on the assumption that there is no directional selection on the markers used. Because RAD sequencing involves random generation of sequences throughout the genome, the chance of markers being under directional selection is lower than when using markers based on gene coding sequences. A bigger dataset also allowed to look for the occurrence of barrier effects. The research questions dealt with in this chapter are:



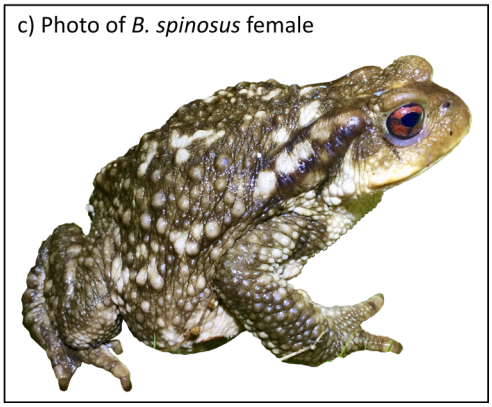


Figure 1.5: Map (a) of the distribution of *Bufo bufo* (blue) and *B. spinosus* (red), and the hybrid zone between the two species (purple). Photos of (b) a *B. bufo* female, and (c) a *B. spinosus* female. Photographs are kindly provided by Iñigo Martínez-Solano and Sergé Bogaerts.

Have either or both the northwest or southeast section of the hybrid zone moved?

If so, is movement the same in both?

Can barrier effects be detected using RAD data?

Introgression was asymmetric in northwest France, with higher levels of admixture linkage disequilibrium towards the north of the hybrid zone. Introgression was symmetric in southeast France, and higher levels of admixture linkage disequilibrium in the centre of southeast France. The use of genomic cline analyses allowed us to find barrier genes.

Chapter 5 is a draft manuscript, which we intend to publish as "Spatial variation in introgression along the common toad hybrid zone in France" by **I. van Riemsdijk**, J.W.

Arntzen, R. K. Butlin, G. Bucchiarelli, E. McCartney-Melstad, M. Rafajlovic, P. Scott, E. Toffelmier, B. Shaffer, and B. Wielstra.

References

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17, 81–92. https://doi.org/10.1038/nrg.2015.28

Arntzen, J. W. (1978). Some hypotheses on postglacial migrations of the fire-bellied toad *Bombina bombina* (Linnaeus) and the yellow-bellied toad *Bombina variegata* (Linnaeus). Journal of Biogeography, 5, 339–345.

Arntzen, J. W., Beukema, W., Galis, F., & Ivanović, A. (2015). Vertebral number is highly evolvable in salamanders and newts (family Salamandridae) and variably associated with climatic parameters. Contributions to Zoology, 84, 85–113.

Arntzen, J. W., de Vries, W., Canestrelli, D., & Martínez-Solano, I. (2017). Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Molecular Ecology, 26, 5663–5675. https://doi.org/10.1111/mec.14273

Arntzen, J. W., McAtear, J., Butôt, R., & Martínez-Solano, I. (2018). A common toad hybrid zone that runs from the Atlantic to the Mediterranean. Amphibia-Reptilia, 39, 41–50. https://doi.org/10.1163/15685381-00003145

Arntzen, J. W., & Olgun, K. (2000). Taxonomy of the banded newt, *Triturus vittatus*: morphological and allozyme data. Amphibia-Reptilia, 21, 155–168.

Arntzen, J. W., Recuero, E., Canestrelli, D., & Martínez-Solano, I. (2013). How complex is the *Bufo bufo* species group? Molecular Phylogenetics and Evolution, 69, 1203–1208. https://doi.org/10.1016/j.ympev.2013.07.012

Arntzen, J. W., Trujillo, T., Butot, R., Vrieling, K., Schaap, O. D., Gutiérrez-Rodriquez, J., & Martinez-Solano, I. (2016). Concordant morphological and molecular clines in a contact zone of the common and spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. Frontiers in Zoology, 13, 1–12. https://doi.org/10.1186/s12983-016-0184-7

Arntzen, J. W., & Wallis, G. P. (1991). Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. Evolution, 45, 805–826. https://doi.org/10.2307/2409691

Ashlock, P. D., & Mayr, E. (1991). Principles of systematic zoology. New York: McGraw-Hill.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3, 1–7. https://doi.org/10.1371/journal.pone.0003376

Baird, S. J. E. (2015). Exploring linkage disequilibrium. Molecular Ecology Resources, 15, 1017–1019. https://doi.org/10.1111/1755-0998.12424

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution, 68, 2644–2657. https://doi.org/10.1111/evo.12457

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), Hybrid zones and the evolutionary process (pp. 13–45). New York: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual Review of Ecology and Systematics, 16, 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. Nature, 341,

497–503. https://doi.org/10.1038/341497a0

Bateson, P. (2002). William Bateson: a biologist ahead of his time. Journal of Genetics, 81, 49–58. https://doi.org/10.1007/BF02715900

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bilgin, R. (2011). Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. International Journal of Molecular Sciences, 12, 4080–4103. https://doi.org/10.3390/ijms12064080

Bonnet, T., Leblois, R., Rousset, F., & Crochet, P.-A. (2017). A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution, 71, 2140–2158. https://doi.org/10.1111/evo.13296

Brelsford, A., Toews, D. P. L., & Irwin, D. E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. Proceedings of the Royal Society B: Biological Sciences, 284. https://doi.org/10.1098/rspb.2017.1106

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. Heredity, 99, 301–312. https://doi.org/10.1038/sj.hdy.6800997

Bülbül, U., & Kutrup, B. (2013). Morphological and genetic variations of *Ommatotriton* in Turkey. Animal Biology, 63, 297–312. https://doi.org/10.1163/15707563-00002413

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society B: Biological Sciences, 363, 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Butlin, R. K., & Smadja, C. M. (2017). Coupling, reinforcement, and speciation. The American Naturalist, 191, 000–000. https://doi.org/10.1086/695136

Chivas, A. R., García, A., van der Kaars, S., Couapel, M. J. J., Holt, S., Reeves, J. M., … Cecil, C. B. (2001). Sea-level and environmental changes since the last interglacial in the Gulf of Carpentaria, Australia: an overview. Quaternary International, 82, 19–46. https://doi.org/10.1016/S1040-6182(01)00029-5

Coyne, J. A., & Orr, H. (2004). Speciation. Sunderland: Sinauer.

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. Evolution, 62, 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Darwin, C. R. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London, England: John Murray.

Davey, J., Hohenlohe, P., Etter, P., Boone, J., Catchen, J., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12, 499–510. https://doi.org/10.1038/nrg3012

Devitt, T. J., Baird, S. J. E., & Moritz, C. (2011). Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. BMC Evolutionary Biology, 11. https://doi.org/10.1186/1471-2148-11-245

Dobzhansky, T. G. (1947). Genetics and the origin of species. New York: Columbia University Press.

Dobzhansky, T. G. (1970). Genetics of the evolutionary process. New York: Columbia University Press.

Endler, J. A. (1977). Geographic variation, speciation, and clines. (R. M. May, Ed.) (2nd ed.). Princeton, New Jersey: Princeton University Press.

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Feliner, G. N. (2011). Southern European glacial refugia: a tale of tales. Taxon, 60, 365–372.

Forejt, J. (1996). Hybrid sterility in the mouse. Trends in Genetics, 12, 412–417. https://doi.org/ https://doi.org/10.1016/0168-9525(96)10040-8

Gavrilets, S. (2004). Fitness landscapes and the origin of species. Princeton, New Jersey: Princeton University Press.

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution, 62, 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss & N. Ferrand (Eds.), Phylogeography of southern European refugia (pp. 155–188). Dordrecht: Springer Netherlands.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. Molecular Ecology, 20, 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x

Gompert, Z., & Buerkle, C. A. (2012). bgc: software for Bayesian estimation of genomic clines. Molecular Ecology Resources, 12, 1168–1176. https://doi.org/10.1111/1755-0998.12009.x

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 439–450. https://doi.org/10.1098/rstb.2011.0196

Greig, E. I., & Webster, M. S. (2013). Spatial decoupling of song and plumage generates novel phenotypes between 2 avian subspecies. Behavioral Ecology, 24, 1004–1013. https://doi.org/10.1093/beheco/art005

Haldane, J. B. S. (1948). The theory of a cline. Journal of Genetics, 48, 277–284.

Harrison, R. G., & Rand, D. M. (1989). Mosaic hybrid zones and the nature of species boundaries. In D. Otte & J. A. Endler (Eds.), Speciation and its consequences (pp. 111–122).

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. Trends in Ecology and Evolution, 3, 158–167. https://doi.org/10.1016/0169-5347(88)90033-X

Hewitt, G. M. (2004). The structure of biodiversity – insights from molecular phylogeography. Frontiers in Zoology, 1, 1–16. https://doi.org/10.1186/1742-9994-1-4

Hewitt, G. M. (2011a). Mediterranean peninsulas: the evolution of hotspots. In F. E. Zachos & J. C. Habel (Eds.), Biodiversity hotspots: distribution and protection of conservation priority areas (pp. 123–147). Berlin: Springer.

Hewitt, G. M. (2011b). Quaternary phylogeography: the roots of hybrid zones. Genetica, 139, 617–638. https://doi.org/10.1007/s10709-011-9547-3

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. The American Naturalist, 188, 000–000. https://doi.org/10.1086/688018

Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F.-A., Swinkels, W., … Henkel, C. V. (2017). Rapid *de novo* assembly of the European eel genome from nanopore sequencing reads. Scientific Reports, 7, 7213. https://doi.org/10.1038/s41598-017-07650-6

Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., … Perrin, N. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature Communications, 9, 4088. https://doi.org/10.1038/s41467-018-06517-2

Johnson, B. B., White, T. A., Phillips, C. A., & Zamudio, K. R. (2015). Asymmetric introgression in a spotted salamander hybrid zone. Journal of Heredity, 106, 608–617. https://doi.

org/10.1093/jhered/esv042

Kruuk, L. E. B., Jason, S. G., & Barton, N. H. (1999). Hybrid dysfunction in fire-bellied toads (*Bombina*). Evolution, 53, 1611–1616. https://doi.org/10.2307/2640907

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. The American Naturalist, 116, 463–479.

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. Systematic Biology, 63, 534–542. https://doi.org/10.1093/sysbio/syu018

Liedtke, H. C., Gower, D. J., Wilkinson, M., & Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-018-0674-4

Litvinchuk, S. N., Zuiderwijk, A., Borkin, L. J., & Rosanov, J. M. (2005). Taxonomic status of *Triturus vittatus* (Amphibia: Salamandridae) in western Turkey: trunk vertebrae count, genome size and allozyme data. Amphibia-Reptilia, 26, 305–323.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, 17, 142–152. https://doi.org/10.1111/1755-0998.12596

Mallet, J. (2005). Hybridization as an invasion of the genome. Trends in Ecology and Evolution, 20, 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Barton, N., Lamas, G. M., Santisteban, J. C., Muedas, M. M., & Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics, 124, 921–936.

Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development, 47, 69–74. https://doi.org/10.1016/j.gde.2017.08.007

Matz, M. V. (2017). Fantastic Beasts and how to sequence them: ecological genomics for obscure model organisms. Trends in Genetics, 34, 121–132. https://doi.org/10.1016/j.tig.2017.11.002

Mayr, E. (1942). Systematics and the Origin of Species from the viewpoint of a zoologist. (L. C. Dunn, H. T. Clarke, M. M. Rhoades, S. R. Detwiler, F. Schrader, & S. Hecht, Eds.) (2nd ed.). New York: Columbia University Press.

Mayr, E. (1982). Speciation and Macroevolution. Evolution, 36, 1119–1132.

McCartney-Melstad, E., Mount, G. G., & Shaffer, H. B. (2016). Exon capture optimization in amphibians with large genomes. Molecular Ecology Resources, 16, 1084–1094. https://doi.org/10.1111/1755-0998.12538

McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.12649

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52, 263–277. https://doi.org/10.1086/409995

Muller, H. J. (1942). Isolating mechanisms, evolution and temperature (6th ed.). Biological Symposia.

Nadachowska, K., & Babik, W. (2009). Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). Molecular Biology and Evolution, 26, 829–841. https://doi.org/10.1093/molbev/msp004

Nowoshilow, S., Schloissnig, S., Fei, J., Dahl, A., Andy, W., Pang, C., … Roscito, J. G. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature Publishing Group, 554, 50–55. https://doi.org/10.1038/nature25458

Olmo, E. (1973). Quantitative variations in the nuclear DNA and phylogenesis of the amphibia. Caryologia, 26, 43–68. https://doi.org/10.1080/00087114.1973.10796525

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., … Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology, 22, 3304–3317. https://doi.org/10.1111/mec.12201

Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. Genetics, 189, 227–235. https://doi.org/10.1534/genetics.111.129817

Porter, A. H., Wenger, R., Geiger, H., Scholl, A., & Shapiro, A. M. (1997). The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. Evolution, 51, 1561. https://doi.org/10.2307/2411208

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: finding barriers to gene flow. Journal of Evolutionary Biology, 30, 1450–1477. https://doi.org/10.1111/jeb.13047

Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., … Martínez-Solano, I. (2012). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Molecular Phylogenetics and Evolution, 62, 71–86. https://doi.org/10.1016/j.ympev.2011.09.008

Rice, W. R. (1998). Intergenomic conflict, interlocus antagonistic coevolution and the evolution of reproductive isolation. In D. J. Howard & S. H. Berlocher (Eds.), Endless Forms Species and Speciation (pp. 261–270). Oxford University Press.

Rohwer, S., Bermingham, E., & Wood, C. (2001). Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. Evolution, 55, 405–422. https://doi.org/10.1554/0014-3820(2001)055[0405:PAMDHV]2.0.CO;2

Roy, J. S., O'Connor, D., & Green, D. M. (2012). Oscillation of an anuran hybrid zone: morphological evidence spanning 50 years. PLoS ONE, 7. https://doi.org/10.1371/journal.pone.0052819

Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. Proceedings of the National Academy of Sciences of the United States of America, 201714950. https://doi.org/10.1073/pnas.1714950115

Schmickl, R., Marburger, S., Bray, S., & Yant, L. (2017). Hybrids and horizontal transfer: introgression allows adaptive allele discovery. Journal of Experimental Botany, 10.1093/jxb/erx297. https://doi.org/10.1093/jxb/erx297

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2016). The on-again-off-again relationship between mitochondrial genomes and species boundaries. Molecular Ecology, 26, 2212–2236. https://doi.org/10.1111/mec.13959

Smith, J. J., Putta, S., Zhu, W., Pao, G. M., Verma, I. M., Hunter, T., … Randal, S. R. (2009). Genic regions of a large salamander genome contain long introns and novel genes. BMC Genomics, 10, 1–11. https://doi.org/10.1186/1471-2164-10-19

Smith, J. J., Timoshevskaya, N., Timoshevskiy, V. A., Keinath, M. C., Hardy, D., & Voss, S. R. (2019). A chromosome-scale assembly of the axolotl genome. Genome Research, 373548. https://doi.org/10.1101/gr.241901.118

Sun, C., López Arriaza, J. R., & Mueller, R. L. (2012). Slow DNA loss in the gigantic genomes of salamanders. Genome Biology and Evolution, 4, 1340–1348. https://doi.org/10.1093/gbe/evs103

Sun, C., & Mueller, R. L. (2014). Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. Genome Biology and Evolution, 6, 1818–1829. https://doi.org/10.1093/gbe/evu143

Sun, C., Shepard, D. B., Chong, R. A., López Arriaza, J., Hall, K., Castoe, T. A., … Mueller, R. L. (2012). LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biology and Evolution, 4, 168–83. https://doi.org/10.1093/gbe/evr139

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the Fire-Bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution, 40, 1141–1159.

Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution, 45, 237–261.

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. Current Biology, 24, 671–676. https://doi.org/10.1016/j.cub.2014.01.069

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology, 21, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Trujillo, T., Gutiérrez-Rodríguez, J., Arntzen, J. W., & Martínez-Solano, I. (2017). Morphological and molecular data to describe a hybrid population of the common toad (*Bufo bufo*) and the spined toad (*Bufo spinosus*) in western France. Contributions to Zoology, 86, 1–10.

Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. Genetics, 154, 1663–1679. https://doi.org/10.2307/2410324

Visser, M., Leeuw, M. De, Zuiderwijk, A., & Arntzen, J. W. (2017). Stabilization of a salamander moving hybrid zone. Ecology and Evolution, 7, 689–696. https://doi.org/10.1002/ece3.2676

Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkovú, P., Vyskočilovú, M. M., … Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. Molecular Ecology, 20, 2985–3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x

While, G. M., Michaelides, S., Heathcote, R. J. P., MacGregor, H. E. A., Zajac, N., Beninde, J., … Uller, T. (2015). Sexual selection drives asymmetric introgression in wall lizards. Ecology Letters, 18, 1366–1375. https://doi.org/10.1111/ele.12531

Wielstra, B., Burke, T., Butlin, R. K., & Arntzen, J. W. (2017). A signature of dynamic biogeography: enclaves indicate past species replacement. Proceedings of the Royal Society Biological Sciences, 284, 1–6.

Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., … Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. Evolution Letters, 1, 93–101. https://doi.org/10.1002/evl3.9

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. Nature Reviews Genetics, 5, 114–122. https://doi.org/10.1038/nrg1269

Chapter 2

The Near East as a cradle of biodiversity: a phylogeography of banded newts (genus *Ommatotriton*) reveals extensive inter- and intraspecific genetic differentiation

Published in Molecular Phylogenetics and Evolution, 2017

Isolde van Riemsdijk[1,*], Jan W. Arntzen[1], Sergé Bogaerts[2], Michael Franzen[3], Spartak N. Litvinchuk[4], Kurtuluş Olgun[5], Ben Wielstra[1,6,7]

[1]Naturalis Biodiversity Center, P.O. box 9517, 2300 RA Leiden, The Netherlands.

[2]Lupinelaan 25, 5582 CG Waalre, The Netherlands.

[3]Zoologische Staatssammlung München (ZSM-SNSB), Münchhausenstraße 21, 81247 München, Germany.

[4]Institute of Cytology, Russian Academy of Sciences, Tikhoretsky pr. 4, 194064 St. Petersburg, Russia.

[5]Department of Biology, Faculty of Arts and Sciences, Adnan Menderes University, 09010 Aydın, Turkey.

[6]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA.

[7]Department of Animal and Plant Sciences, University of Sheffield, S10 2TN Sheffield, UK.

[*]Correspondence: Isolde van Riemsdijk, Naturalis Biodiversity Center, P.O. box 9517, 2300 RA Leiden, The Netherlands, isolde.vanriemsdijk@naturalis.nl

Abstract

The banded newt (genus *Ommatotriton*) is widely distributed in the Near East (Anatolia, Caucasus and the Levant) – an understudied region from the perspective of phylogeography. The genus is polytypic, but the number of species included and the phylogenetic relationships between them are not settled. We sequenced two mitochondrial and two nuclear DNA markers throughout the range of *Ommatotriton*. For mtDNA we constructed phylogenetic trees, estimated divergence times using fossil calibration, and investigated changes in effective population size with Bayesian skyline plots and mismatch analyses. For nuDNA we constructed phylogenetic trees and haplotype networks. Species trees were constructed for all markers and nuDNA only. Species distribution models were projected on current and Last Glacial Maximum climate layers. We confirm the presence of three *Ommatotriton* species: *O. nesterovi*, *O. ophryticus* and *O. vittatus*. These species are genetically distinct and their most recent common ancestor was dated at ~ 25 Ma (Oligocene). No evidence of recent gene flow between species was found. The species show deep intraspecific genetic divergence, represented by geographically structured clades, with crown nodes of species dated ~ 8-13 Ma (Miocene to Early Quaternary); evidence of long-term in situ evolution and survival in multiple glacial refugia. While a species tree based on nuDNA suggested a sister species relationship between *O. vittatus* and *O. ophryticus*, when mtDNA was included, phylogenetic relationships were unresolved, and we refrain from accepting a particular phylogenetic hypothesis at this stage. While species distribution models suggest reduced and fragmented ranges during the Last Glacial Maximum, we found no evidence for strong population bottlenecks. We discuss our results in the light of other phylogeographic studies from the Near East. Our study underlines the important

role of the Near East in generating and sustaining biodiversity.

Keywords: Anatolia; historical biogeography; Levant; molecular dating; phylogeny; species distribution modelling

Introduction

Phylogeography reconstructs phylogenetic patterns in space and time to identify the drivers that shape biodiversity within taxa (Avise, 2000; Hickerson et al., 2010). The Mediterranean Basin is considered a near-ideal laboratory for phylogeographical studies (Hewitt, 2011). The collision of the Eurasian, African and Arabian tectonic plates, from the Late Eocene (37-34 Ma) onwards, caused major geological changes in the Mediterranean Basin, as marine barriers were re-arranged and mountains formed (Meulenkamp & Sissingh, 2003; Popov et al. 2006). This process left its signature in the phylogeography of the species inhabiting the region (Hewitt, 2011). While the Quaternary (2.6 Ma - present) glacial cycles resulted in drastic climate shifts in temperate Eurasia, the climate in the Mediterranean Basin remained relatively stable (Stewart et al., 2009; Hewitt, 2011). As a consequence, many species had a continuous presence within the Mediterranean Basin, while they only periodically colonized more northern areas during interglacials – and subsequently were wiped out during the next glacial cycle (Hewitt, 2000).

Because of the interaction of geographic instability and climatic stability, many Mediterranean species are characterized by deep intraspecific lineages (Oosterbroek & Arntzen, 1992; Bilgin, 2011). The Mediterranean Basin features a high concentration of biodiversity and is considered a biodiversity hotspot (Myers et al., 2000). Complex phylogeographic patterns are particularly well documented for the Iberian, Italian and Balkan peninsulas (Hewitt, 2011). Despite the similar geologic and climatologic background of the Near East, historical biogeographical studies of this region are underrepresented in the phylogeographical literature (Beheregaray, 2008; Riddle, 2016).

Most amphibian species, especially salamanders, are particularly influenced by geological disruptions and climate oscillations, due to their limited dispersal capabilities and strong dependence on water, often resulting in a pronounced phylogenetic structure (Vences & Wake, 2007). Seven Salamandridae genera (*Lissotriton*, *Lyciasalamandra*, *Mertensiella*, *Neurergus*, *Ommatotriton*, *Salamandra*, and *Triturus*) occur in the Near East. Phylogeographical studies of these genera revealed long term in situ evolution and survival during glacial cycles (Steinfartz et al., 2000; Tarkhnishvili et al., 2000; Veith et al., 2008; Wielstra et al., 2010; Hendrix et al., 2014; Pabijan et al., 2015). However, banded newts (genus *Ommatotriton*) remain to be studied in-depth. Because *Ommatotriton* is the most wide-spread salamander genus within the Near East (Sparreboom, 2014), a phylogeographic survey is expected to be insightful regarding the region's historical biogeography.

Ommatotriton is endemic to the Near East (Fig. 2.1). The genus was considered a single species, *O. vittatus* (Gray, 1835), with a disjunct northern and southern distribution, until Litvinchuk et al. in 2005 proposed that the northern banded newt range segment should be regarded as a separate, distinct species, *O. ophryticus* (Berthold, 1846). This

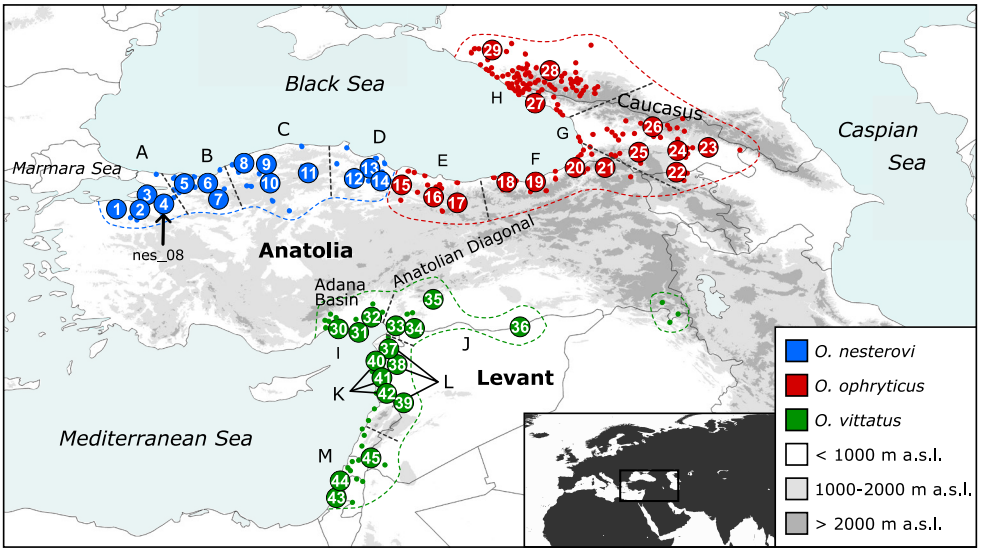


Figure 2.1: Distribution of and sampling scheme for banded newts (genus *Ommatotriton*). Coloured dashed lines roughly delineate the three species' ranges and black dashed lines delineate mtDNA-based phylogeographical groups (A-M). Numbered dots (1-45) represent sampled localities and small dots additional localities used for species distribution modelling. Grey shading reflects elevation.

treatment is now generally accepted (e.g. Sparreboom, 2014; Frost, 2016). A recent proposal by Bülbül & Kutrup (2013) to split the northern banded newt range segment into two species, *O. ophryticus* in the east and *O. nesterovi* (Litvinchuk, Zuiderwijk, Borkin and Rosanov, 2005) in the west, is not universally supported. For convenience we refer to the three taxa as species from here on. The phylogeny of *Ommatotriton* is unclear. Considering three *Ommatotriton* species, three branching orders are possible and each is supported by existing data. Phylogenetic hypothesis I ((*O. ophryticus*, *O. vittatus*), *O. nesterovi*) is supported by mtDNA data (Bülbül & Kutrup 2013), II ((*O. ophryticus*, *O. nesterovi*), *O. vittatus*) by allozyme data (Arntzen & Olgun, 2000; Litvinchuk et al., 2005), and III ((*O. vittatus*, *O. nesterovi*), *O. ophryticus*) by the number of rib-bearing vertebrae (Arntzen & Olgun, 2000; Arntzen et al., 2015; Litvinchuk et al., 2005).

Previous studies dealing with the inter- and intraspecific phylogenetic patterns of *Ommatotriton* suffered from limited sampling. We here conduct range-wide sampling. We use mitochondrial and nuclear DNA (mtDNA and nuDNA) markers to determine inter- and intraspecific phylogenetic relationships, conduct fossil-based time calibration and test for interspecific hybridization. Additionally, species distribution models are constructed to evaluate the impact of the Quaternary climate oscillations, and compared to historical demographic reconstructions. We identify interspecific and intraspecific phylogeographical patterns of *Ommatotriton* and discuss our findings in the context of a comparative phylogeography of the Near East. Our study highlights the importance of the Near East as a biodiversity hotspot.

Materials and methods

Sampling and sequencing

We collected tissue samples of 1-3 individuals (average 2.6, total 106) from 40 localities (Fig. 2.1, Table A.1). The Thermo Scientific KingFisher Purification System was used to extract DNA. Fragments of two mitochondrial protein coding markers were sequenced: cytochrome b (cyt b, 786 bp) and cytochrome c oxidase I (COI, 658 bp). We added COI sequences for 13 individuals from five additional localities from Smith et al. (2008). Fragments of two unlinked nuclear protein coding markers (Hendrix et al., 2014; Shen et al., 2012), KIAA1239 (600 bp; hereafter KIAA) and SACS (624 bp), were sequenced for a single individual per locality, but included all three individuals for the three *O. ophryticus* and three *O. nesterovi* locations closest to their parapatric contact zone as defined by the mtDNA data (see Results, total 47, Table A.1).

We adjusted the 'universal' cyt b primers (MVZ-15-NP, MVZ-16-NP) from Moritz et al. (1992) and COI primers (VF1-d, VR1-d) from Ivanova et al. (2006) to match the full mitogenome of *O. nesterovi* (Zhang et al., 2008; Table A.2). We designed *Ommatotriton*--specific primers based on a KIAA and a SACS (Shen et al., 2012) sequence for *O. nesterovi* presented in Hendrix et al. (2014). In cases where no PCR product could be obtained, presumably due to degraded DNA, internal primers were designed to amplify stretches of 100-300 bp (three primer sets for COI, KIAA and SACS, and four for cyt b; Table A.3).

PCR conditions for DNA amplification were: initial denaturation for 180 s at 94°C, 35 cycles with 30 s denaturation at 94°C, 30 s annealing at 58°C and 60 s extension at 72°C, and a 240 s final extension step at 72°C. Sanger sequencing was done commercially on an ABI 3730xl DNA analyser at BaseClear B.V., Leiden, the Netherlands. Sequences were assembled in Sequencher 4.10.1 (Gene Codes Corporation, MI USA) and aligned in Mesquite 3.04 (Maddison & Maddison, 2015). Nuclear alleles were phased with the Phase 2.1 algorithm (Stephens & Donnelly, 2003) implemented in DnaSP 5 (Librado & Rozas, 2009). GenBank accession numbers are listed in Table A.1.

Phylogenetic analyses

The mtDNA alignment was divided into six partitions, by marker (COI and cyt b) and by codon position (first, second and third). *Lissotriton vulgaris* and *Neurergus strauchii* were added as outgroups (data from Zhang et al. 2008; Table A.2). Maximum likelihood (ML) analysis was performed with RAxML 8.2.4 (Stamatakis, 2014), with 1000 rapid bootstrap replicates and a search for the best scoring maximum likelihood tree. The nucleotide substitution model GTR+G was used for each partition, as recommended in the RAxML manual. Bayesian phylogenetic reconstruction was performed with MrBayes 3.2.6 (Ronquist et al., 2012). We determined the best fitting models of sequence evolution for each of the six partitions based on the Bayesian information criterion (Guindon & Gascuel, 2003) in jModelTest 2.1.7 (Darriba et al., 2012; Table A.4). A Markov Chain Monte Carlo search was run for a million generations under default parameters (two runs, four chains, with temperature 0.2), a sampling frequency of 0.001 and a burn-in of 25% of generations. State frequencies were unlinked and rates were set to vary across each of six partitions. Tracer 1.6 (Rambaut et al., 2014) was

used to confirm stabilisation within and convergence between runs (all ESS values > 200). RAxML and MrBayes were run on the Cipres Science Gateway (Miller et al., 2010). Trees were also reconstructed for each mtDNA marker separately, following the same procedure as for the concatenated mtDNA data.

Phylogenetic trees for nuDNA were constructed using the same approach as for mtDNA, with *Neurergus strauchii* as outgroup (data taken from Hendrix et al., 2014). Because jModelTest suggested two out of three substitution models to be identical for codon positions within nuclear markers (Table A.4), each marker was analysed unpartitioned to avoid over-parameterisation of the analysis. Because these trees showed little divergence (Fig. B.7-B.10), we additionally constructed minimum spanning haplotype networks (Bandelt et al., 1999) for each nuclear marker in PopArt 1.7 (Leigh et al., 2016).

*BEAST, implemented in BEAST 1.8.3 (Heled & Drummond, 2010; Drummond et al., 2012), was used to construct a species tree for the nuDNA markers. Species were set as operational taxonomic units. We selected the best fitting model per marker in jModelTest (Table A.4). We performed a species tree analysis including all mtDNA and nuclear DNA markers (two twenty million generation runs), and a species tree analyses using only the two nuclear DNA markers (one ten million generation run) with a sampling frequency of 0.001. Tracer 1.6 was used to confirm stabilisation and convergence as for MrBayes. A burn-in of 10% was removed when combining log and tree files in LogCombiner (BEAST 1.8.3). TreeAnnotator (BEAST 1.8.3) was used to construct a maximum clade credibility tree with mean node height.

Molecular dating

COI and cyt b sequences for all Salamandridae genera (Table A.2) were taken from Zhang et al. (2008). We included 14 *Ommatotriton* haplotypes, one randomly chosen haplotype from each phylogeographical group (see Results, Fig 2a) and a distinct haplotype found at a single locality only (Nes_08 from locality 4). We excluded the cyt b sequence for *Ichthyosaura alpestris*, because it contained a large number of unique non-synonymous substitutions suggestive of a pseudogene (see also Steinfartz et al., 2007). BEAST 1.8.3 was used to generate a calibrated phylogeny, employing seven fossil-based calibration points within Salamandridae, the family to which *Ommatotriton* belongs, following the procedure outlined in Wiens et al. (2011; Table A.5). To ensure that the calibration points fitted within the age constraints, we constructed a starting tree adapted from a RAxML tree with the R package 'APE' (Paradis et al., 2004). Four runs of 20 million generations were executed on the Cipres Science Gateway with nucleotide substitution models selected with jModelTest (Table A.4) and a sampling frequency of 0.001. Log and tree files were combined and convergence was checked as above.

Species distribution modelling

We combined our localities of species occurrence with those in Borkin et al. (2003). This yielded 75 localities for *O. vittatus*, 53 for *O. nesterovi* and 233 for *O. ophryticus*. Localities were assigned to species based on ranges suggested by our genetic analyses. We used bioclimatic variables from the WorldClim database 1.4 (Hijmans et al., 2005) as explanatory variables. Following Guisan & Thuiller (2005) and Peterson (2011), we

selected variables with a Pearson's correlation < 0.7 that were considered biologically meaningful: isothermality (bio3), temperature seasonality (bio4), mean temperature of wettest quarter (bio8), mean temperature of driest quarter (bio9), precipitation of wettest quarter (bio16) and precipitation of driest quarter (bio17).

We used Maxent 3.3.3k (Phillips et al., 2006) to construct species distribution models for each *Ommatotriton* species. Because the area on which species distribution models are calibrated influences model parameters considerably (Stokland et al., 2011), we restricted sampling of pseudo-absence data to a 200 km buffer zone around *Ommatotriton* localities (VanDerWal et al., 2009). Feature type was restricted to hinge features for a smoother model fit, emphasising trends rather than idiosyncrasies of the data (Elith et al., 2010). We confirmed that the models performed better than random by testing them for significance against a null model derived from random localities, following Raes & ter Steege (2007). Random point data (99 replicates per analysis) were created with ENMTools 1.3 (Warren et al., 2010). Models were projected on climate reconstructions for the Last Glacial Maximum (~ 21 ka), based on the MIROC (Sueyoshi et al., 2013) and CCSM (Brady et al., 2013) climate simulations, available from WorldClim.

Population dynamics

Bayesian skyline plots (Ho & Shapiro, 2011) were generated with BEAST for 40 *O. nesterovi*, 40 *O. ophryticus*, and 39 *O. vittatus* mtDNA sequences with 6 million generations. Parameters were kept at default except for the nucleotide substitution models (selected with jModelTest; Table A.4) and the estimated total substitution rate (4.802×10^{-3} substitutions My^{-1} from the molecular dating analysis). Mismatch analyses were performed per species in DnaSP (Table A.6). We used COI data only because DnaSP is unable to handle missing data, and several cyt b sequences were not available (Table A.1). The observed haplotype frequencies were compared to haplotype frequencies expected at constant population size. The statistics R2 and Fu's Fs (Ramos-Onsins & Rozas, 2002) were calculated to test for significance (R2 < 0.05 and Fu's Fs < 0.02), using coalescent simulations with 1000 generations.

Results

For COI the number of segregating sites (S) is 214, nucleotide diversity (π) is 0.119, and total number of mutations (θ) is 71.26, for cyt b 100, 0.114, and 30.55, for KIAA 20, 0.007 and 4.11 and for SACS 17, 0.006 and 3.34. The mtDNA phylogenetic tree revealed three monophyletic groups, corresponding to the three *Ommatotriton* species (summarized in Fig. 2.2a, detailed MrBayes and RAxML trees in Fig. B.1 and B.2). The most recent common ancestor of *Ommatotriton* was dated at 24.5 Ma [95% HPD confidence interval (CI) 14.6-35.8 Ma, Fig. 2.2b]. *Ommatotriton nesterovi*, *O. ophryticus* and *O. vittatus* crown nodes were estimated at 13.0 Ma (CI 4.9-22.3 Ma), 8.6 Ma (CI 2.1-17.2 Ma) and 12.5 Ma (CI 5.5-20.8 Ma). The phylogenetic tree based on both mtDNA markers suggested *O. nesterovi* and *O. ophryticus* as sister species, but with low support, independent of the method used (Fig. 2.2a). Phylogenetic trees for the individual mtDNA markers strongly supported the three monophyletic clades, while relationships among clades were poorly supported (Fig. B.3-B.6). The species tree

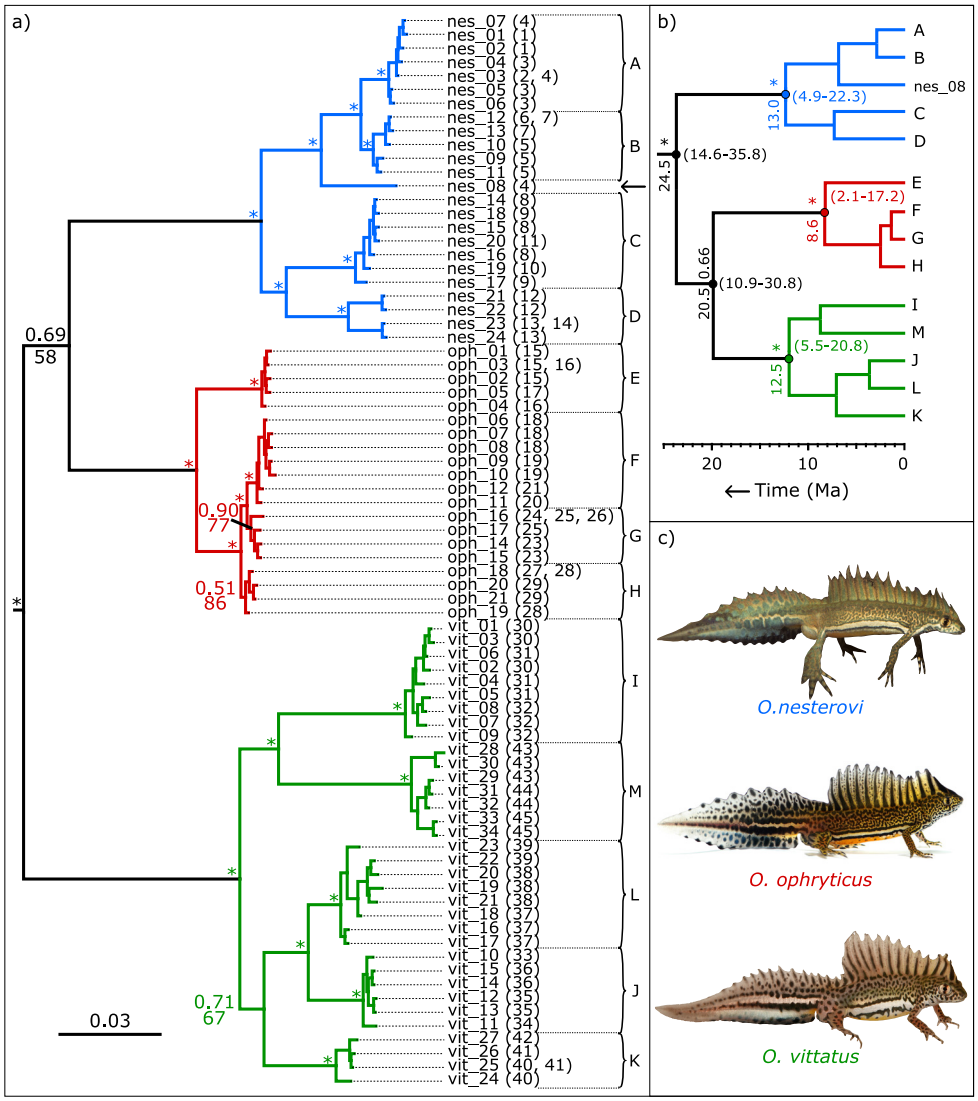


Figure 2.2: Mitochondrial DNA phylogenetic tree and dated tree for banded newts (genus *Ommatotriton*). Panel a) Bayesian consensus tree with Bayesian posterior probability value above and Maximum Likelihood bootstrap support value below branches. Asterisks indicate nodes supported with a posterior probability > 0.95 and a bootstrap support value > 80 and (values below the crown nodes of phylogeographical groups are not shown). Tips are haplotypes with codes corresponding to Table A.1 and localities in parentheses. Panel b) dated phylogeny with Bayesian posterior probability value left above branches (asterisk indicates posterior probability > 0.95) and node age on the left below with 95% HPD confidence interval in parentheses right of nodes. In a) and b) capital letters correspond to the phylogeographical groups delineated in Fig. 2.1; the outgroup is not shown. Panel c) male banded newt species in breeding condition.

based on nuDNA only suggested that *O. ophryticus* and *O. vittatus* are sister species, with high support (posterior probability of 0.98; Fig. 2.3c), while adding mtDNA to the *BEAST analysis resulted in a lower support for this clade (0.75).

For nuDNA phylogenetic trees (BI and ML, Fig. B.7-B.10), weak support was found for a monophyletic *O. nesterovi*. *Ommatotriton vittatus* was suggested, albeit with low

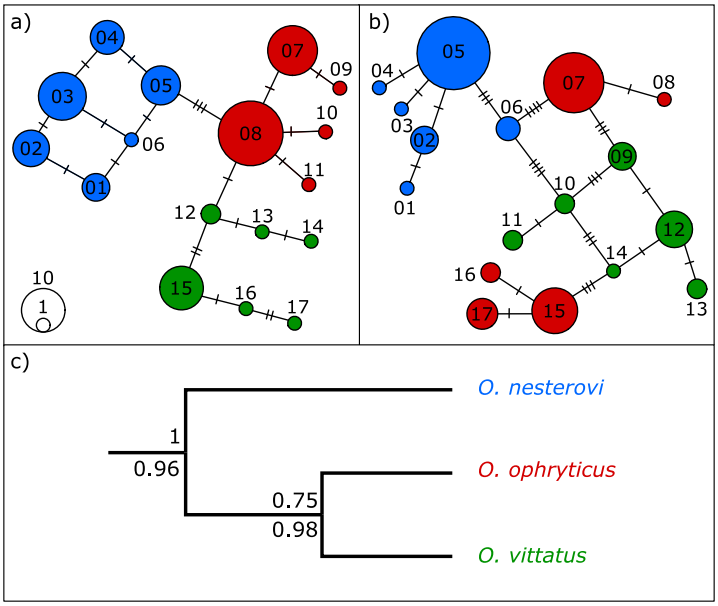


Figure 2.3: Haplotype networks and species tree based on nuclear markers for banded newts (genus *Ommatotriton*). In the haplotype networks for KIAA (a) and SACS (b) pie diameter represents the number of individuals per haplotype and bars the number of inferred substitutions. Panel c) *BEAST species tree with posterior values for all markers above, and only nuDNA markers below the branches. Colours indicate species as in Fig. 2.1.

support, to be either monophyletic or paraphyletic and nested within *O. ophryticus*, reflecting the relationships found in the haplotype network. In the KIAA network haplotypes for each species are clustered, while in the SACS haplotype network *O. nesterovi* was split up in two groups (Fig. 2.3a, 2.3b). We found no evidence for nuDNA haplotype sharing between species as defined by mtDNA (Fig. 2.2a, 2.3a, 2.3b).

Within species, 13 distinct monophyletic mtDNA clades representing geographically coherent groups were found (Fig. 2.1 and Fig. 2.2a). The sole exception was haplotype Nes_08 from locality 4, which is placed in a sister position to phylogeographic groups A and B and occurs in the geographic range of group A. Intraspecific phylogeographical groups were found to be more closely related to geographically more closely located phylogeographic groups than to groups located further apart, with the exception of the I+M-clade from Israel and the Adana Basin. For nuDNA, intraspecific differentiation was low (Fig. 2.3 and Fig. B.7-B.10). For the SACS gene, individuals from group I (localities 30 and 31) were relatively diverged from other individuals of *O. vittatus* (groups J-M). The KIAA marker suggested divergence of western (group E; localities 15-17) and eastern (groups F-H; localities 19-29) *O. ophryticus*, in line with the mtDNA-based phylogenetic tree. The mtDNA phylogeographic groups, however, were not recovered in the nuDNA analyses.

Projections on Last Glacial Maximum climate reconstructions showed the potential distribution for the three species was reduced and fragmented compared to the present (Fig. 2.4 and Fig. B.11). For *O. vittatus*, the range reduction was less pronounced than for *O. nesterovi* and *O. ophryticus*. The potential distributions of *O. nesterovi* and *O. ophryticus* overlap considerably at the present day and much less so at the Last Glacial Maximum. Unsuitable environmental conditions separated *O. vittatus*

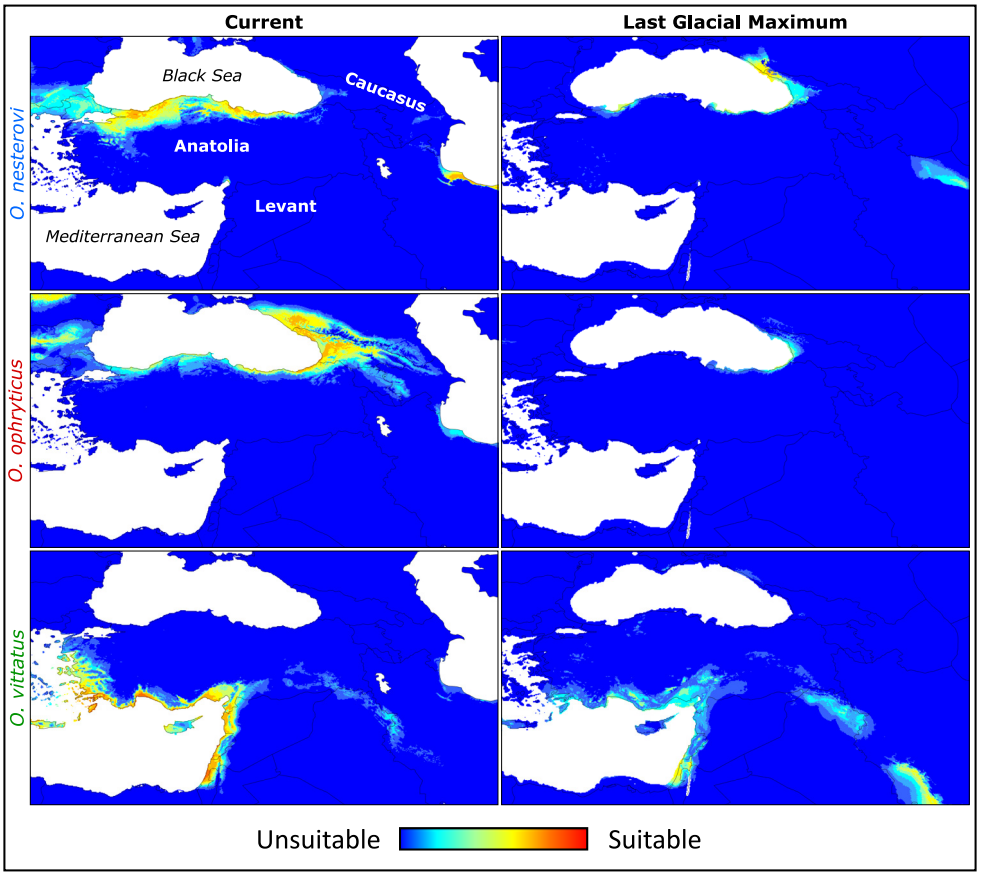


Figure 2.4: Species distribution models for banded newts (genus *Ommatotriton*). Model projections for the three species are arranged from top to bottom, with projections on current (left) and MIROC-derived Last Glacial Maximum (right) climate layers. Predicted suitability is reflected by a colour scale running from deep blue (highly unsuitable) to deep red (highly suitable).

from the other two species during either period. Bayesian skyline plots based on mtDNA indicated a more pronounced dip in effective population size for *O. nesterovi* at ~0.5 Ma than for *O. ophryticus* and *O. vittatus* at ~ 0.05 Ma and at ~ 0.5 Ma (Fig. B12). Mismatch analyses indicated that allele frequencies did not differ from the allele frequency expected for a stable population over time (Fig. B.11, Table A.6).

Discussion

Biogeographical scenario for three genetically distinct species

The mtDNA phylogeny for *Ommatotriton* reveals three genetically distinct clades, which is in line with the proposal to recognize three banded newt species: *O. nesterovi*, *O. ophryticus* and *O. vittatus*. While reciprocal monophyly is not significantly supported for individual nuDNA markers, crucially none of the nuclear haplotypes are shared between species, not even in the region where the ranges of *O. nesterovi* and *O. ophryticus* approach one another, and where we sampled more densely (Fig. 2.1). On the basis of these results we accept the three species treatment.

Hybridization is not a likely cause of the lack of reciprocal monophyly of *O. vittatus* since it is separated from the remaining *Ommatotriton* range by the Anatolian Diagonal (Fig. 2.1). This mountain range was formed at 30-25 Ma as a consequence of the Arabia-Eurasia collision (Jolivet & Faccenna, 2000) and acts as a barrier for a wide range of taxa (Kapli et al., 2013). Uplift of the Anatolian Diagonal likely isolated the stock that would become *O. vittatus* at ~25 Ma (Fig. 2.2b). Last Glacial Maximum conditions appear to have exacerbated this geographical isolation (Fig. 2.4).

The Arabia-Eurasia collision resulted in the uplift of the Anatolian Plateau (Hearn & Ni, 1994), which likely initiated speciation of what would become *O. nesterovi* and *O. ophryticus* at ~ 25 Ma (Fig. 2.2b). The two species were isolated during the Last Glacial Maximum (Fig. 2.4), implying that the current contact is of secondary, postglacial origin. Bilgin (2011) documented a north-south oriented suture zone, related to postglacial expansion from refugia in western and eastern Anatolia for plants, insects and vertebrates. The two other pairs of parapatric newt species distributed along the southern Black Sea coast show evidence of hybridisation upon secondary postglacial contact with pronounced mtDNA introgression, from the western (*Lissotriton vulgaris* and *Triturus ivanbureschi*) to the eastern species (*L. kosswigi* and *T. anatolicus;* Nadachowska & Babik, 2009; Wielstra et al., 2013, 2017), but there is no such mismatch between mtDNA and nuDNA for *O. nesterovi* and *O. ophryticus*. While we did not find evidence of (recent) introgression between the two species, to be conclusive, a denser sampling filling the 60 km gap between the closest *O. nesterovi* and *O. ophryticus* localities (localities 14 and 15, Fig. 2.1) is required.

Ancient and rapid radiation

The fossil based time calibration suggests the *Ommatotriton* species radiated ~ 25 Ma during a short time span. The mtDNA phylogeny suggests the three species form a polytomy. While our nuDNA species tree supported a ((*O. ophryticus*, *O. vittatus*), *O. nesterovi*) phylogeny, support decreased when adding mtDNA. We conclude that our data provide no convincing support of a particular branching order, and we consider the *Ommatotriton* phylogeny unresolved.

Conflicting results among markers can be expected for ancient and rapid radiations because few informative substitutions accumulate along internal branches (Philippe et al., 2011), little time is available for lineage sorting, resulting in conflicting gene genealogies (Pamilo & Nei, 1988) and uninformative substitutions accumulate along terminal branches, causing long-branch attraction (Felsenstein, 1978). Although comprehensive sampling of phylogenetic diversity can break up long branches (Bergsten, 2005), our sampling is comprehensive. However, we merely employed three unlinked genetic markers. Multispecies coalescence methods, as implemented in e.g. *BEAST, explicitly take incomplete lineage sorting into account (Heled & Drummond, 2010) and benefit from sampling more unlinked markers (Maddison & Knowles, 2006). Hence, we suggest a phylogenomic approach is required to resolve the phylogeny of banded newts.

Deep intraspecific phylogeographical structure

A deep genetic differentiation within *Ommatotriton* species is reflected by four to five intraspecific mtDNA phylogeographical groups per species. Splits between *O.*

nesterovi groups located southeast of the Sea of Marmara (A, B) and along the central southern Black Sea shore (C, D), and between *O. vittatus* groups from the Adana Basin and Israel (I, M) and Turkey east of the Amanus Mountains and Syria (J, K, L; Fig. 2.1, Fig. 2.2b), coincide with the Middle Miocene disruption, a sudden period of climate cooling between 14.0 and 13.5 Ma (Flower & Kennett, 1994). This period is linked to vicariance and genetic divergence in the Near East for skinks (*Ablepharus*; Skourtanioti et al., 2016), toads (*Bufo*; Garcia-Porta et al., 2012) and scorpions (*Mesobuthus*; Shi et al., 2013). The basal split between *O. ophryticus* groups located in central Turkey (E) and north-eastern Turkey, Georgia and on the western side of the Greater Caucasus (F, G, H) coincides with the uplift of the Armenian plateau during the Serravallian, caused by the Arabia-Eurasia collision (13-11 Ma; Gelati, 1975). This event was also inferred to have caused the split between the crested newts *Triturus karelinii* versus *T. ivanbureschi* and *T. anatolicus* (Wielstra et al., 2010).

We observe further subdivision into *O. nesterovi* groups (A and B, C and D), during the Late Miocene and Pliocene. This time period was associated with large--scale orogeny throughout the region (Oosterbroek & Arntzen, 1992; Bilgin, 2011). The different *O. nesterovi* groups are currently separated by protrusions of the Pontic Mountains, but the presence of a distinct haplotype from inside the range of group A, basal to the clade of groups A and B, shows the pattern is more complex. Divergence of the three groups distributed in the eastern part of the *O. ophryticus* range (F, G, H) suggests the Quaternary climate oscillations (2.6 Ma - present) initiated increased population fragmentation.

For *O. vittatus* we observe two allopatrically distributed groups that, even though they are sister clades, occur in the Adana Basin (I) and Israel (M). The Adana Basin is isolated from the remainder of the *Ommatotriton* range by the Amanus Mountains lifted during the lower-middle Miocene (~ 23 - 11 Ma; Aksu et al., 2005). Desiccation of the Mediterranean Sea during the Messinian salinity crisis (~ 5.96 - 5.33 Ma) would have established a direct connection allowing colonization of the Adana Basin (Duggen et al., 2003). Subsequent expansion of the stock that became groups J, K and L would explain the disjunct distribution pattern observed today. Interpreting the phylogeographical pattern of groups J, K, and L is difficult based on the current sampling. Wider sampling in this presently hostile area would be required to further elucidate the historical biogeography of *O. vittatus*. Deep phylogenetic splits and complex phylogeographical patterns caused by the uplift of the Amanus Mountains and dispersal routes facilitated by the Messinian salinity crisis in southern Anatolia and the Levant were previously suggested for amphibians (Akin et al., 2010; Skourtanioti et al., 2016) and reptiles (Guicking et al., 2009; Kapli et al., 2013).

Multiple intraspecific glacial refugia

The species distribution models for *O. vittatus* suggest relatively little range contraction during the Last Glacial Maximum (Fig. 2.4), supported by the Bayesian skyline plot suggesting a relatively small population decrease during the Quaternary (Fig. B.12). The deep phylogeographical structure suggests multiple glacial refugia (i.e. "refugia-within-refugia") were present within the current range of *O. vittatus*. Multiple refugia within the current range of *O. vittatus* have been found for reptiles (Kornilios et al., 2011; Stümpel et al., 2016), insects (Simonato et al., 2007) and plants

(Médail & Diadema, 2009).

Compared to *O. vittatus*, a more pronounced habitat reduction at the Last Glacial Maximum was suggested by the species distribution models for *O. ophryticus* and, especially, *O. nesterovi* (Fig. 2.4). This is supported by the reduced population size indicated by the Bayesian skyline plots (Fig. B12). The deep genetic structure within the two northern species also survived the Last Glacial Maximum in multiple glacial refugia, located south of the Marmara Sea, and various locations along the southern Black Sea shore and in the Colchis (Fig. 2.1). The Marmara region acted as a glacial refugium for newts (Wielstra et al., 2010; Pabijan et al., 2015), multiple glacial refugia along the southern Black Sea coast were found for amphibians (Wielstra et al., 2010; Dufresnes et al., 2016), reptiles (Guicking et al., 2009) and mammals (Dubey et al., 2006, 2007; Bilgin et al., 2008) and multiple glacial refugia in the Colchis were found for plants, insects and snails (Tarkhnishvili et al., 2012), amphibians (Tarkhnishvili et al., 2000; Veith et al., 2003), lizards (Freitas et al., 2016), and mammals (Seddon et al., 2002; Tarkhnishvili et al., 2012).

Conclusions

The genus *Ommatotriton* represents a rapid radiation, initiated by orogeny associated with the Arabia-Eurasia collision during late Oligocene. Lack of gene flow supports the recognition of three species of banded newt, but relationships among these species remain unresolved. Each species is composed of distinct, geographically coherent mtDNA clades, with splits dated from the onset of the Quaternary to as far back in time as the Middle Miocene. Multiple glacial refugia per species are positioned along the coasts of the Black and Mediterranean Seas, suggesting a refugia-within-refugia scenario for the Near East, as has been previously inferred for the Iberian (Gómez & Lunt, 2007), Italian (Canestrelli et al., 2012) and Balkan peninsulas (Poulakakis et al., 2015). *Ommatotriton* is exemplary for co-distributed taxa and the overarching pattern emerging from cumulative phylogeographical studies underlines the Near East as an important biodiversity hotspot, on par with hotspots revealed through comparative phylogeography for the better studied Mediterranean peninsulas (Riddle, 2016).

Acknowledgements

IvR is supported by the 'Nederlandse organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487. Permits to collect were made available to KO by TÜBITAK. We thank Michael Fahrbach and the late Max Sparreboom for providing *O. nesterovi* and *O. ophryticus* pictures and Aroub Almasri, Frank Pasmans and Max Sparreboom for help during fieldwork. We thank two anonymous reviewers who provided helpful comments.

Data accessibility

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

Appendix A. Supplementary Tables A.1-A.5

Appendix B. Supplementary Figures B.1-B.12

GenBank accession numbers are available in Table A.1. The raw and edited sequence data, in- and output for phylogenetic analyses and species distribution model analyses are available at the Dryad database: DOI: https://doi.org/10.5061/dryad.4q08q

References

Akin C., Bilgin C.C., Beerli P., Westaway R., Ohst T., Litvinchuk S.N., Uzzell T., Bilgin M., Hotz H., Guex G.D., & Plötner J. (2010) Phylogeographic patterns of genetic diversity in eastern Mediterranean water frogs were determined by geological processes and climate change in the Late Cenozoic. Journal of Biogeography, 37, 2111–2124.

Aksu A.E., Calon T.J., Hall J., Mansfield S., & Yaşar D. (2005) The Cilicia–Adana basin complex, Eastern Mediterranean: Neogene evolution of an active fore-arc basin in an obliquely convergent margin. Marine Geology, 221, 121–159.

Avise, J.C. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Arntzen J.W., Beukema W., Galis F., & Ivanović A. (2015) Vertebral number is highly evolvable in salamanders and newts (family Salamandridae) and variably associated with climatic parameters. Contributions to Zoology, 84, 85–113.

Arntzen J.W. & Olgun K. (2000) Taxonomy of the banded newt, *Triturus vittatus*: morphological and allozyme data. Amphibia-Reptilia, 21, 155–168.

Bandelt H.J., Forster P., & Rohl A. (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16, 37–48.

Beheregaray L.B. (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Molecular ecology, 17, 3754–74.

Bergsten J. (2005) A review of long-branch attraction. Cladistics, 21, 163–193.

Berthold A.A. (1846) Über das Vorkommen von Tritonen am Kaukasus. Nachrichten von der Georg-Augusts-Universität und der Königlichen Gesellschaf der Wissenschaften, Göttingen.

Bilgin R. (2011) Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. International Journal of Molecular Sciences, 12, 4080–4103.

Bilgin R., Karataş A., Coraman E., Disotell T., & Morales J.C. (2008) Regionally and climatically restricted patterns of distribution of genetic diversity in a migratory bat species, *Miniopterus schreibersii* (Chiroptera: Vespertilionidae). BMC Evolutionary Biology, 8, 209–223.

Borkin L.J., Litvinchuk S.N., & Zuiderwijk A. (2003) *Triturus vittatus* (Gray, 1835) - Bandmolch. Handbuch der Reptilien und Amphibien Europas (ed. by K. Grossenbacher and B. Thiesmeier), pp. 555–605. Aula-Verlag, Wiebelsheim.

Brady E.C., Otto-Bliesner B.L., Kay J.E., & Rosenbloom N. (2013) Sensitivity to glacial forcing in the CCSM4. Journal of Climate, 26, 1901–1925.

Bülbül U. & Kutrup B. (2013) Morphological and genetic variations of *Ommatotriton* in Turkey. Animal Biology, 63, 297–312.

Canestrelli D., Salvi D., Maura M., Bologna M.A., & Nascetti G. (2012) One species, three Pleistocene evolutionary histories: phylogeography of the Italian crested newt, *Triturus carnifex*. PLoS ONE, 7, e41754.

Darriba D., Taboada G.L., Doallo R., & Posada D. (2012) jModelTest 2: more models, new

heuristics and parallel computing. Nature Methods, 9, 772.
Drummond A.J., Suchard M. a, Xie D., & Rambaut A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution, 29, 1969–73.
Dubey S., Cosson J.-F., Vohralík V., Krystufek B., Diker E., & Vogel P. (2007) Molecular evidence of Pleistocene bidirectional faunal exchange between Europe and the Near East: the case of the bicoloured shrew (*Crocidura leucodon*, Soricidae). Journal of Evolutionary Biology, 20, 1799–808.
Dubey S., Zaitsev M., Cosson J.-F., Abdukadier A., & Vogel P. (2006) Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). Molecular Phylogenetics and Evolution, 38, 635–647.
Dufresnes C., Litvinchuk S.N., Leuenberger J., Ghali K., Zinenko O., Stöck M., & Perrin N. (2016) Evolutionary melting pots: a biodiversity hotspot shaped by ring diversifications around the Black Sea in the Eastern tree frog (*Hyla orientalis*). Molecular Ecology, Early view, .
Duggen S., Hoernle K., van den Bogaard P., Rüpke L., & Phipps Morgan J. (2003) Deep roots of the Messinian salinity crisis. Nature, 422, 602–606.
Elith J., Kearney M., & Phillips S. (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution, 1, 330–342.
Felsenstein J. (1978) Cases in which parsimony or compatibility methods will be possitively misleading. Systematic Zoology, 27, 401–410.
Flower B.P. & Kennett J.P. (1994) The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeography, Palaeoclimatology, Palaeoecology, 108, 537–555.
Freitas S., Vavakou A., Arakelyan M., Drovetski S. V., Crnobrnja-Isailović J., Kidov A. A., Cogălniceanu D., Corti C., Lymberakis P., Harris D.J., & Carretero M.A. (2016) Cryptic diversity and unexpected evolutionary patterns in the meadow lizard, *Darevskia praticola* (Eversmann, 1834). Systematics and Biodiversity, 14, 184–197.
Frost D. (2016) Available at: http://research.amnh.org/vz/herpetology/amphibia/Amphibia/Caudata/Salamandridae/Pleurodelinae/Ommatotriton. Last visited: 12 August 2016
Garcia-Porta J., Litvinchuk S.N., Crochet P.A., Romano A., Geniez P.H., Lo-Valvo M., Lymberakis P., & Carranza S. (2012) Molecular phylogenetics and historical biogeography of the west--palearctic common toads (*Bufo bufo* species complex). Molecular Phylogenetics and Evolution, 63, 113–130.
Gelati R. (1975) Miocene marine sequence from lake Van eastern Turkey. Rivista Italiana di Paleontologia e Stratigraphia, 81, 477–480.
Gómez A. & Lunt D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. Phylogeography of southern European refugia (ed. by S. Weiss and N. Ferrand), pp. 155–188. Springer Netherlands, Dordrecht.
Gray (1835) A manual of British vertebrate animals. Pitt Press, Cambridge.
Guicking D., Joger U., & Wink M. (2009) Cryptic diversity in a Eurasian water snake (*Natrix tessellata*, Serpentes: Colubridae): evidence from mitochondrial sequence data and nuclear ISSR-PCR fingerprinting. Organisms Diversity & Evolution, 9, 201–214.
Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology, 52, 696–704.
Guisan A. & Thuiller W. (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8, 993–1009.
Hearn T.M. & Ni J.F. (1994) Pn velocities beneath continental collision zones: the Turkish--Iranian Plateau. Geological Journal International, 117, 273–283.

Heled J. & Drummond A.J. (2010) Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution, 27, 570–580.

Hendrix R., Fleck J., Schneider W., Schneider C., Geller D., Avci A., Olgun K., & Steinfartz S. (2014) First comprehensive insights into nuclear and mitochondrial DNA based population structure of Near East mountain brook newts (Salamandridae: genus *Neurergus*) suggest the resurrection of *Neurergus derjugini*. Amphibia-Reptilia, 35, 172–187.

Hewitt G.M. (2000). The genetic legacy of the Quaternary ice ages. Nature, 405, 907-913.

Hewitt G.M. (2011) Mediterranean peninsulas: the evolution of hotspots. Biodiversity Hotspots: distribution and protection of conservation priority areas (ed. by F.E. Zachos and J.C. Habel), pp. 123–147. Springer, Berlin.

Hickerson M.J., Carstens B.C., Cavender-Bares J., Crandall K.A., Graham C.H., Johnson J.B., Rissler L., Victoriano P.F., & Yoder A.D. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution, 54, 291–301.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965–1978.

Ho S.Y.W. & Shapiro B. (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources, 11, 423–434.

Ivanova N. V., Dewaard J.R., & Hebert P.D.N. (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes, 6, 998–1002.

Jolivet L. & Faccenna C. (2000) Mediterranean extension and the Africa-Eurasia collision. Tectonics, 19, 1095–1106.

Kapli P., Botoni D., Ilgaz C., Kumlutaş Y., Avcı A., Rastegar-Pouyani N., Fathinia B., Lymberakis P., Ahmadzadeh F., & Poulakakis N. (2013) Molecular phylogeny and historical biogeography of the Anatolian lizard *Apathya* (Squamata, Lacertidae). Molecular Phylogenetics and Evolution, 66, 992–1001.

Kornilios P., Ilgaz C., Kumlutas Y., Giokas S., Fraguedakis-Tsolis S., & Chondropoulos B. (2011) The role of Anatolian refugia in herpetofaunal diversity: an mtDNA analysis of T*yphlops vermicularis* Merrem, 1820 (Squamata, Typhlopidae). Amphibia-Reptilia, 32, 351–363.

Leigh J., Bryant D., & Steel M. (2016) PopArt. Version 1.7. Available at: http://popart.otago.ac.nz.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics (Oxford, England), 25, 1451–1452.

Litvinchuk S.N., Zuiderwijk A., Borkin L.J., & Rosanov J.M. (2005) Taxonomic status of *Triturus vittatus* (Amphibia: Salamandridae) in western Turkey: trunk vertebrae count, genome size and allozyme data. Amphibia-Reptilia, 26, 305–323.

Maddison W.P. & Knowles L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. Systematic biology, 55, 21–30.

Maddison W.P. & Maddison D.R. (2015) Mesquite: a modular system for evolutionary analysis. Version 3.10. Available at: http://mesquiteproject.org.

Médail F. & Diadema K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. Journal of Biogeography, 36, 1333–1345.

Meulenkamp J.E. & Sissingh W. (2003) Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone. Palaeogeography, Palaeoclimatology, Palaeoecology, 196, 209–228.

Miller M.A., Pfeiffer W., & Schwartz T. (2010) Creating the CIPRES Science Gateway for inference

of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop.14 nov, New Orleans, LA, 1–8.
Moritz C., Schneider C.J., & Wake D.B. (1992) Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Systematic Biology, 41, 273–291.
Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., & Kent J. (2000) Biodiversity hotspots for conservation priorities. Nature, 403, 853–858.
Nadachowska K. & Babik W. (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). Molecular Biology and Evolution, 26, 829–841.
Oosterbroek P. & Arntzen J.W. (1992) Area-cladograms of circum-Mediterranean taxa in relation to Mediterranean palaeogeography. Journal of Biogeography, 19, 3–20.
Pabijan M., Zieliński P., Dudek K., Chloupek M., Sotiropoulos K., Liana M., & Babik W. (2015) The dissection of a Pleistocene refugium: phylogeography of the smooth newt, L*issotriton vulgaris*, in the Balkans. Journal of Biogeography, 42, 671–683.
Pamilo P. & Nei M. (1988) Relationships between gene trees and species trees. Molecular Biology and Evolution, 5, 568–583.
Paradis E., Claude J., & Strimmer K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, 20, 289–290.
Peterson A.T. (2011) Ecological niche conservatism: a time-structured review of evidence. Journal of Biogeography, 38, 817–827.
Philippe H., Brinkmann H., Lavrov D. V, Littlewood D.T.J., Manuel M., Wörheide G., & Baurain D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biology, 9, e1000602.
Phillips S.J., Anderson R.P., & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259.
Popov S. V., Shcherba I.G., Ilyina L.B., Nevesskaya L.A., Paramonova N.P., Khondkarian S.O., & Magyar I. (2006) Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. Palaeogeography, Palaeoclimatology, Palaeoecology, 238, 91–106.
Poulakakis N., Kapli P., Lymberakis P., Trichas A., Vardinoyiannis K., Sfenthourakis S., & Mylonas M. (2015) A review of phylogeographic analyses of animal taxa from the Aegean and surrounding regions. Journal of Zoological Systematics and Evolutionary Research, 53, 18–32.
Raes N. & ter Steege H. (2007) A null-model for significance testing of presence-only species distribution models. Ecography, 30, 727–736.
Rambaut A., Suchard M.A., & Drummond A.J. (2014) Tracer. Version 1.6. Available at: tree.bio.ed.ac.uk/software/tracer.
Ramos-Onsins S.E. & Rozas J. (2002) Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution, 19, 2092–2100.
Riddle B.R. (2016) Comparative phylogeography clarifies the complexity and problems of continental distribution that drove A.R. Wallace to favor islands. PNAS, 113, 7970–7977.
Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., & Huelsenbeck J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61, 539–42.
Seddon J.M., Santucci F., Reeve N., & Hewitt G.M. (2002) Caucasus Mountains divide postulated postglacial colonization routes in the white-breasted hedgehog, *Erinaceus concolor*. Journal of Evolutionary Biology, 15, 463–467.

Shen X.-X., Liang D., & Zhang P. (2012) The development of three long universal nuclear protein--coding locus markers and their application to osteichthyan phylogenetics with nested PCR. PLoS ONE, 7, e39256.

Shi C.M., Ji Y.J., Liu L., Wang L., & Zhang D.X. (2013) Impact of climate changes from Middle Miocene onwards on evolutionary diversification in Eurasia: insights from the mesobuthid scorpions. Molecular Ecology, 22, 1700–1716.

Simonato M., Mendel Z., Kerdelhué C., Rousselet J., Magnoux E., Salvato P., Roques A., Battisti A., & Zane L. (2007) Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East. Molecular Ecology, 16, 2273–2283.

Skourtanioti E., Kapli P., Ilgaz Ç., Kumlutaş Y., Avcı A., Ahmadzadeh F., Crnobrnja-Isailović J., Gherghel I., Lymberakis P., & Poulakakis N. (2016) A reinvestigation of phylogeny and divergence times of the *Ablepharus kitaibelii* species complex (Sauria, Scincidae) based on mtDNA and nuDNA genes. Molecular Phylogenetics and Evolution, 103, 199–214.

Smith M.A., Poyarkov N.A., & Hebert D.N. (2008) CO1 DNA barcoding amphibians: take the chance, meet the challenge. Molecular Ecology Resources, 8, 235–246.

Sparreboom, M. (2014) Salamanders of the Old World. KNNV Publishing, Zeist, The Netherlands.

Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30, 1312–1313.

Steinfartz S., Veith M., & Tautz D. (2000) Mitochondrial sequence analysis of *Salamandra* taxa suggests old splits of major lineages and postglacial recolonizations of Central Europe from distinct source populations of *Salamandra salamandra*. Molecular Ecology, 9, 397–410.

Steinfartz S., Vicario S., Arntzen J.W., & Caccone A. (2007) A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary patterns of the Salamandridae with emphasis on *Triturus newts*. Journal of Experimental Zoology (Molecular and Developmental Evolution), 306B, 1–24.

Stephens M. & Donnelly P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. American Journal of Human Genetics, 73, 1162–1169.

Stewart J.R., Lister A.M., Barnes I., & Dalén L. (2009) Refugia revisited: individualistic responses of species in space and time. Proceedings of the Royal Society Biological sciences, 277, 661–671.

Stokland J.N., Halvorsen R., & Støa B. (2011) Species distribution modelling - effect of design and sample size of pseudo-absence observations. Ecological Modelling, 222, 1800–1809.

Stümpel N., Rajabizadeh M., Avcı A., Wüster W., & Joger U. (2016) Phylogeny and diversification of mountain vipers (*Montivipera*, Nilson et al., 2001) triggered by multiple Plio–Pleistocene refugia and high-mountain topography in the Near and Middle East. Molecular Phylogenetics and Evolution, 101, 336–351.

Sueyoshi T., Ohgaito R., Yamamoto A., Chikamoto M.O., Hajima T., Okajima H., Yoshimori M., Abe M., O'ishi R., Saito F., Watanabe S., Kawamiya M., & Abe-Ouchi A. (2013) Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM. Geoscientific Model Development, 6, 819–836.

Tarkhnishvili D., Gavashelishvili A., & Mumladze L. (2012) Palaeoclimatic models help to understand current distribution of Caucasian forest species. Biological Journal of the Linnean Society, 105, 231–248.

Tarkhnishvili D.N., Thorpe R.S., & Arntzen J.W. (2000) Pre-pleistocene refugia and differentiation between populations of the Caucasian salamander (*Mertensiella caucasica*).

Molecular phylogenetics and evolution, 14, 414–22.
VanDerWal J., Shoo L.P., Graham C., & Williams S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecological Modelling, 220, 589–594.
Veith M., Lipscher E., Oz M., Kiefer A., Baran I., Polymeni R.M., & Steinfartz S. (2008) Cracking the nut: geographical adjacency of sister taxa supports vicariance in a polytomic salamander clade in the absence of node support. Molecular Phylogenetics and Evolution, 47, 916–31.
Veith M., Schmidtler J.F., Kosuch J., Baran I., & Seitz A. (2003) Palaeoclimatic changes explain Anatolian mountain frog evolution: a test for alternating vicariance and dispersal events. Molecular Ecology, 12, 185–199.
Vences M. & Wake D.B. (2007) Speciation, species boundaries and phylogeography of amphibians. Amphibian Biology, vol.6, Systematics (ed. by H. Heatwole and M. Tyler), pp. 2613–2660. Surrey Beatty & Sons, Chipping Norton, Australia.
Warren D.L., Glor R.E., & Turelli M. (2010) ENMTools: a toolbox for comparative studies of environmental niche models. Ecography, 33, 607–611.
Wielstra B., Baird A.B., & Arntzen J.W. (2013) A multimarker phylogeography of crested newts (*Triturus cristatus* superspecies) reveals cryptic species. Molecular Phylogenetics and Evolution, 67, 167–175.
Wielstra B., Burke T., Butlin R.K., Avcı A., Üzüm N., Bozkurt E., Olgun K. & Arntzen J.W. (2017) A genomic footprint of hybrid zone movement in crested newts. Evolution Letters, 1 (2), 93-101
Wielstra B., Espregueira Themudo G., Güçlü Ö., Olgun K., Poyarkov N.A., & Arntzen J.W. (2010) Cryptic crested newt diversity at the Eurasian transition: the mitochondrial DNA phylogeography of Near Eastern *Triturus* newts. Molecular Phylogenetics and Evolution, 56, 888–896.
Wiens J.J., Sparreboom M., & Arntzen J.W. (2011) Crest evolution in newts: implications for reconstruction methods, sexual selection, phenotypic plasticity and the origin of novelties. Journal of Evolutionary Biology, 24, 2073–2086.
Zhang P., Papenfuss T.J., Wake M.H., Qu L., & Wake D.B. (2008) Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Molecular Phylogenetics and Evolution, 49, 586–597.

Chapter 3

Molecular data reveal the hybrid nature of an introduced population of banded newts (*Ommatotriton*) in Spain

Published in Conservation Genetics, 2018

Isolde van Riemsdijk[1*], Laurens van Nieuwenhuize[1], Iñigo Martínez-Solano[2], Jan W. Arntzen[1], Ben Wielstra[3,4]

[1]Naturalis Biodiversity Center, P.O. box 9517, 2300 RA Leiden, The Netherlands

[2]Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

[3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA

[4]Department of Animal and Plant Sciences, University of Sheffield, S10 2TN Sheffield, UK

[*]Correspondence: Isolde van Riemsdijk, Naturalis Biodiversity Center, P.O. box 9517, 2300 RA Leiden, The Netherlands, isolde.vanriemsdijk@naturalis.nl

Abstract

The three species of banded newt (genus *Ommatotriton*) are endemic to the Near East. Recently an introduced banded newt population was discovered in Catalonia, Spain. To determine the species involved and the geographical source, we genotyped 11 individuals for one mitochondrial and two nuclear genetic markers, and compared the observed haplotypes to a range-wide phylogeography of *Ommatotriton*. All haplotypes identified in Spain are identical to haplotypes known from the native range. The mitochondrial haplotypes derive from *O. ophryticus* and were originally recorded in northeast Turkey. The nuclear haplotypes reveal that all individuals are genetically admixed between *O. ophryticus* and *O. nesterovi*. While the geographical resolution for the nuclear markers is low, the source of the *O. nesterovi* ancestry must be Turkey, as this species is a Turkish endemic. Species distribution models suggest a large potential distribution for the two *Ommatotriton* species, extending over northern Iberia and southern France. The ecology of hybrids can differ substantially from that of the parent species, making the impact of the Spanish hybrid banded newt population difficult to predict.

Keywords: Amphibia; exotic species; genotyping; hybridisation; invasive species; species distribution modelling

Introduction

Human activities are causing a worldwide mass extinction of amphibians (Wake and Vredenburg 2008). Next to habitat destruction, the most important anthropogenic factor causing species extinctions is the introduction of exotic species (Sax and Gaines 2003). While many exotic species disappear quickly after introduction, some

occasionally become invasive and threaten local wildlife (Pimentel et al. 2001). Competition with and predation by exotic taxa can cause major declines in native taxa (Kats and Ferrer 2003). Introduced species may also carry alien pathogens against which native species do not have an adequate immune response (Martel et al. 2014). Moreover, introduced species may threaten the genetic integrity of natives by hybridisation (Fitzpatrick and Shaffer 2007; Meilink et al. 2015).

In May 2011, an introduced population of banded newts (*Ommatotriton* sp.), a genus endemic to the Near East (c. 3000 km eastwards), was discovered in two adjacent ponds near Lleida in Catalonia, Spain (Fontelles et al. 2011, Fig. 3.1). The newts were found in an area with many ponds and watering holes for cattle. Efforts to eradicate the population are underway, including the direct removal of adults and larvae, and desiccation of ponds (Fontelles et al., 2011; D. Martínez, pers. comm). In October 2016, after the peak of emergence of juveniles, few metamorphic individuals were recorded in the vicinity of the ponds, suggesting a successful population reduction.

The genus *Ommatotriton* comprises three species: *O. nesterovi*, *O. ophryticus*, and *O. vittatus*. The number of rib-bearing vertebrae (NRBV) can be used to distinguish *O. ophryticus* (NRBV=13) from the other two (NRBV=12: Arntzen & Olgun, 2000; Litvinchuk et al., 2005; Kutrup & Bülbül, 2011). Based on X-Rays of a single individual with NRBV=13, the Spanish population was identified as *O. ophryticus* (Fontelles et al. 2011). However, NRBV is known to be environmentally plastic in salamanders (Arntzen et al. 2015). Furthermore, pictures of Spanish banded newts show features reminiscent of *O. nesterovi*, in particular the lateral white band that continues from the front limbs up to the eye and is interrupted by large blue specks on the tail (B. Wielstra, unpublished data).

Considering the uncertainty regarding specific identity and hence the geographic origin of the introduced banded newt population, we genotyped 11 Spanish individuals for one mitochondrial and two nuclear markers and compared the results to a rangewide *Ommatotriton* phylogeography (van Riemsdijk et al.,2017). We also used species distribution modelling to investigate the potential of the Spanish banded newts to spread over Iberia.

Materials and methods

Tail and toe tips from 11 Spanish *Ommatotriton* individuals were collected with permission from the Servicio de Fauna y Flora, Generalitat de Catalunya, Spain. Samples are stored in the Museo Nacional de Ciencias Naturales Tissue and DNA Collection in Madrid, Spain (Online Resources 1). DNA was extracted with the Qiagen DNeasy Blood & Tissue Kit. Fragments of one mitochondrial (COI, 658 bp) and two nuclear markers (KIAA, 600 bp and SACS, 624 bp), all nuclear markers for which reference data from the natural range are available, were amplified following van Riemsdijk et al. (2017). Sanger sequencing was done commercially at BaseClear B.V., Leiden, the Netherlands. Sequences were edited in Sequencher 4.10.1 (Gene Codes Corporation, MI USA) and nuclear alleles were phased with the PHASE 2.1 algorithm (Stephens and Donnelly 2003) in DnaSP 5 (Librado and Rozas 2009). Sequences were combined with the dataset of van Riemsdijk et al. (2017) in Mesquite (Maddison and Maddison 2015) and collapsed into unique haplotypes (Online Resources 1). For the

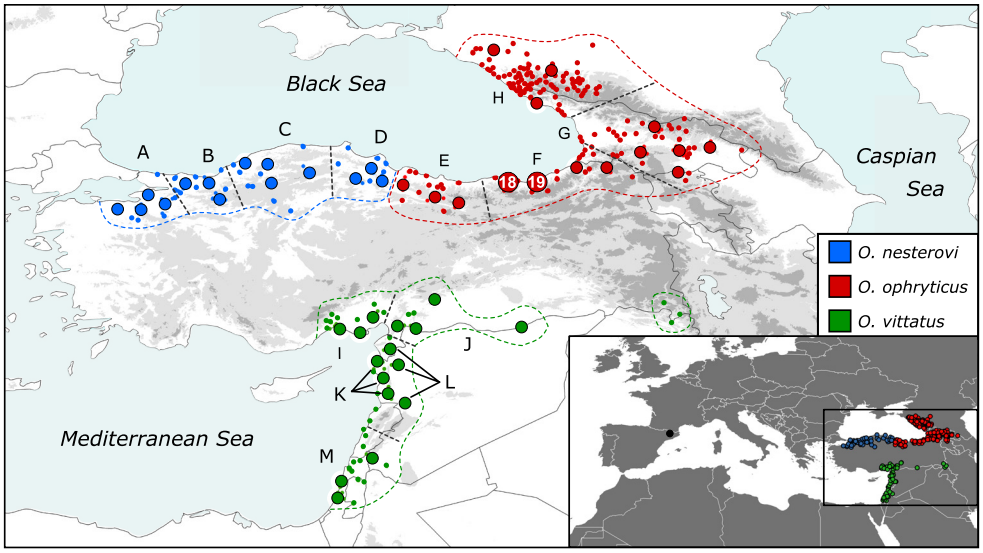
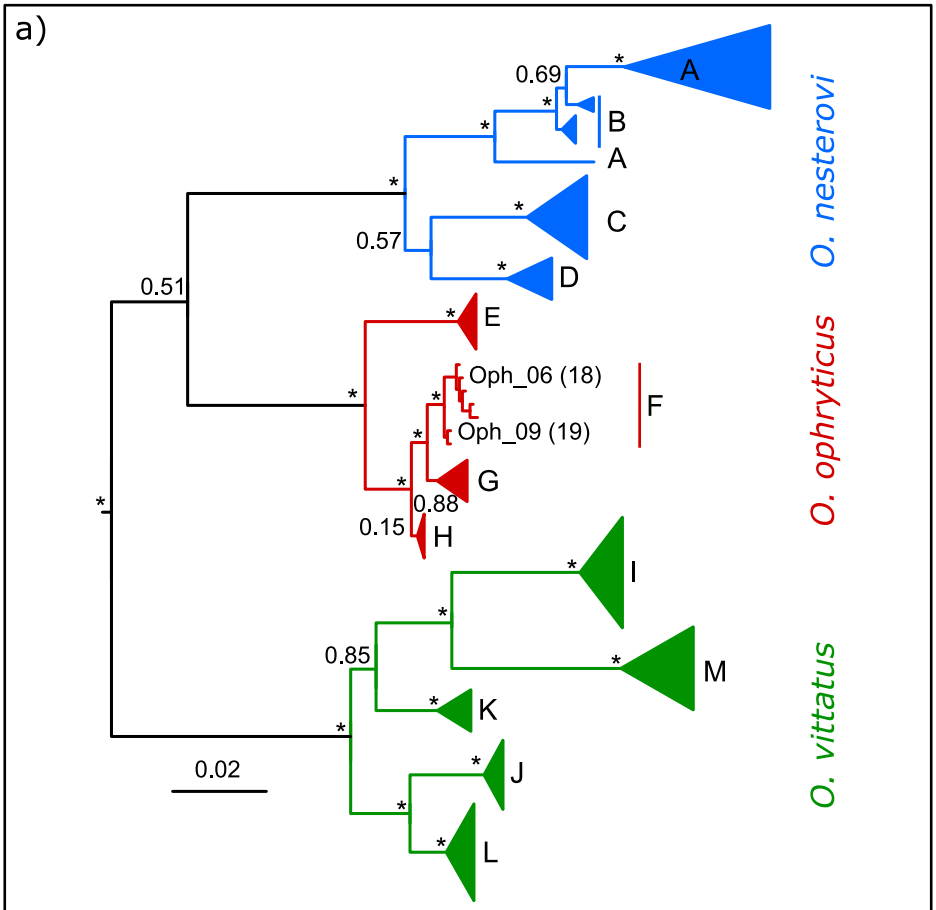


Figure 3.1: Distribution map of the three banded newt *Ommatotriton* species, native to the Near East. The inset shows the position of the introduced population in Spain (black dot). On the main map, large dots are populations for which genetic data are available and small dots are additional locations used for species distribution modelling (Riemsdijk et al. 2017). Dots are coloured according to species. Letters correspond to phylogeographic clades in Fig. 3.2. Spanish mitochondrial haplotypes were previously reported from localities 18 and 19.

nuclear markers, haplotype networks were built using PopArt 1.7 (Bandelt et al. 1999; Leigh et al. 2016). Species distribution models for *O. ophryticus* and *O. nesterovi*, based on bioclimatic values and presented in van Riemsdijk et al. (2017), and an additional model combining distribution data for both species were projected onto current climate layers of northern Iberia, to approximate the suitable area in the introduced range and the invasive potential of the exotic population.

Results

All mitochondrial and nuclear haplotypes observed in the Spanish population were identical to haplotypes from the native range (van Riemsdijk et al. 2017). The two mitochondrial haplotypes found in the Spanish population were identical *O. ophryticus* haplotypes from the Trabzon Province in north-eastern Turkey (Fig. 3.1, Fig. 3.2a, localities 18 and 19 in van Riemsdijk et al., 2017). The Bayesian phylogenetic tree for COI in Fig. 3.2a is adapted from van Riemsdijk et al (2017). For both nuclear markers two haplotypes were recovered, one previously found in *O. nesterovi* and one in *O. ophryticus* (Fig. 3.2b, 3.2c, 3.2d). All Spanish individuals show ancestry of both species. Linking nuclear haplotypes to particular native localities is impossible as both nuclear markers show little intraspecific phylogeographic structure. Species distribution models suggest that conditions at the Spanish introduction site are unfavourable for *O. nesterovi*, but favourable for *O. ophryticus* (Fig. 3.3). A 'hybrid model' based on distribution data for both species suggests that conditions in and around the introduction site are more suitable for admixed newts than for the individual parental species.



#	COI	KIAA	SACS
1	oph_06	08:08	05:05
2	oph_09	08:08	05:07
3	oph_06	08:08	05:07
4	oph_09	03:03	05:05
5	oph_06	03:08	05:07
6	oph_06	08:08	05:05
7	oph_09	03:08	05:05
8	oph_06	08:08	05:05
9	oph_09	03:08	05:05
10	oph_06	08:08	05:05
11	oph_06	08:08	05:07

Figure 3.2: Haplotypes identified in introduced Spanish banded newt *Ommatotriton* samples in the context of a range-wide phylogeography. A Bayesian phylogeny for COI (a). Asterisks indicate posterior probability values > 0.95 and letters correspond to clades indicated in Fig. 3.1. Haplotype networks of KIAA (b) and SACS (c). Haplotypes are coloured according to the species of origin and Spanish ones are shown as purple pie slices. (d) Overview of all haplotypes found in Spain. Haplotype codes correspond to van Riemsdijk et al. (2017).

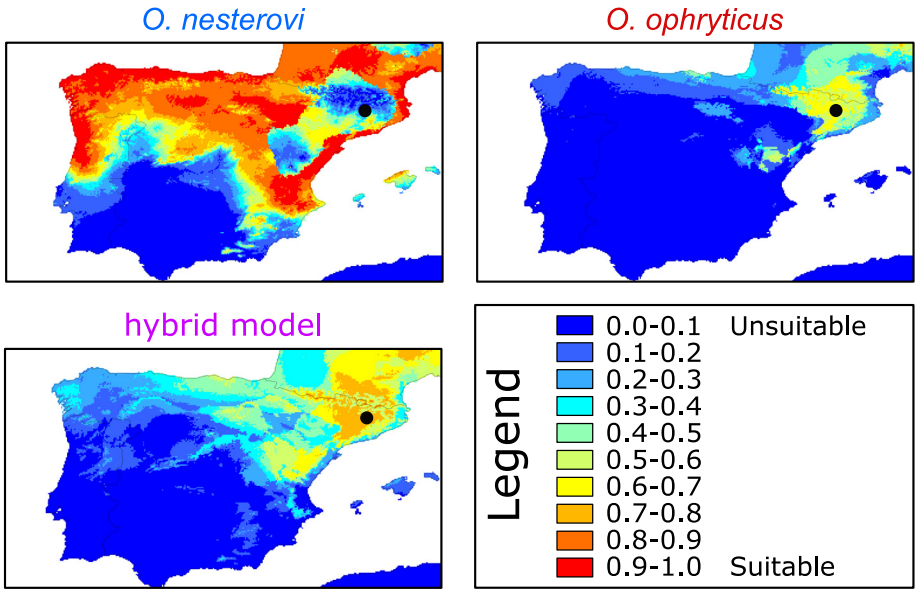


Figure 3.3: Species distribution models for two banded newt species, *O. nesterovi* (upper left) and *O. ophryticus* (upper right), and a 'hybrid model' based on distribution data for both species (bottom left), projected on the Iberian Peninsula. The black dot indicates the introduction site.

Discussion

Mitochondrial DNA suggests that the introduced banded newt population in Spain concerns *O. ophryticus*, in line with the number of rib-bearing vertebrae (Fontelles et al. 2011). However, the nuclear markers show all individuals are genetically admixed between *O. ophryticus* and a second species, *O. nesterovi*, the presence of which was hypothesized based on pictures. Species identification based on mitochondrial data alone has proven, once again, to be misleading (Vences et al. 2005). As all sampled newts are hybrids, the individuals initially introduced in Spain must have been either a mix of pure individuals of *O. nesterovi* and *O. ophryticus*, or hybrids themselves. This is the first evidence that hybrids between these two species are viable.

While the lack of geographic resolution for the nuclear haplotypes prevents us to identify a detailed geographical origin of the source for *O. nesterovi* ancestry, the country of origin is clear as this species is endemic to north-west Turkey. While for *O. ophryticus* the nuclear markers also lack geographic resolution, the mitochondrial marker pinpoints the origin of *O. ophryticus* ancestry to north-east Turkey. Considering that the contact zone between *O. ophryticus* and *O. nesterovi* is positioned c. 400 km to the west of the inferred region of origin of *O. ophryticus*, the Spanish population cannot derive from natural hybrids. Rather, taking into account that *O. ophryticus* and *O. nesterovi* have only recently been recognized (van Riemsdijk et al. 2017; Litvinchuk et al. 2005; Bülbül and Kutrup 2013), the two species likely interbred in the pet trade.

The main impacts of invasive species on natives are competition, predation, transfer of alien pathogens, and hybridisation. The introduced banded newts are likely to compete for resources with native newts and prey on eggs and larvae of native

amphibians to some degree (Borkin et al. 2003). A chytrid fungus, *Batrachochytrium salamandrivorans*, which has proven highly lethal to European Salamandridae, was introduced via the salamander pet trade (Martel et al. 2014). While no elevated amphibian mortality has been reported at the introduction site, monitoring should include screening for chytrid fungus and other pathogens (four introduced alpine newts, *Ichthyosaura alpestris*, in Catalonia were infected with a Ranavirus, see Martínez-Silvestre et al. 2017). Hybridisation with native newt species, like *Triturus marmoratus* or *Lissotriton helveticus* seems a negligible risk, as *Ommatotriton* is not closely related (Arntzen et al. 2015).

The intermingled genomes in the hybrid banded newt population make it hard to approximate the actual niche (Canestrelli et al. 2016). Hybridisation is known to potentially speed up the evolutionary process, leading to increased potential to adapt to new conditions (Allendorf et al. 2001; Dittrich-Reed and Fitzpatrick 2013). Species distribution models highlight the potential of the introduced population to colonize northern Iberia and adjacent France. However, expansion beyond the introduction site might be slowed down because of a low density of suitable breeding ponds surrounding it and an extirpation effort in place (D. Martínez, pers. comm.). However, a contained introduced population could still negatively affect the native amphibian community locally.

The hybrid nature of the Spanish introduced banded newt population makes their impact on native species harder to predict. When the genomes of hybridising species are complementary, the resulting offspring may have a higher fitness than either of the parent species, resulting in a population of 'hopeful monsters' (Fitzpatrick and Shaffer 2007; Dittrich-Reed and Fitzpatrick 2013). Tangible proof of this theory was found for *Ambystoma* salamanders in California, where hybrid larvae between a native and an invasive species show a higher survival rate than either parent species (Fitzpatrick and Shaffer 2007).

Considering the unpredictability of hybrids, we urge the relevant authorities to treat the biological control of the Spanish introduced banded newt population with caution. The efficiency of current measures such as the continued removal of individuals and pond drying should be closely monitored. Even if complete eradication might prove impossible, these measures are likely to reduce the negative effects of the exotic hybrid *Ommatotriton* population in Spain.

Acknowledgements

D. Herrero, D. Martínez, F. Stănescu, D. Székely and A. Tarragó helped collect or provided samples. The PhD position of IvR is supported by the 'Nederlandse organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487.

Compliance with ethical standards

Collection of tissue samples for molecular analyses was authorized by the Servicio de Fauna y Flora, Generalitat de Catalunya, Spain.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16:613–622. doi: 10.1016/S0169-5347(01)-02290-X

Arntzen JW, Beukema W, Galis F, Ivanović A (2015) Vertebral number is highly evolvable in salamanders and newts (family Salamandridae) and variably associated with climatic parameters. Contrib Zool 84:85–113.

Arntzen JW, Olgun K (2000) Taxonomy of the banded newt, *Triturus vittatus*: morphological and allozyme data. Amphibia-Reptilia 21:155–168.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Borkin LJ, Litvinchuk SN, Zuiderwijk A (2003) *Triturus vittatus* (Gray, 1835) - Bandmolch. In: Grossenbacher K, Thiesmeier B (eds) Handbuch der Reptilien und Amphibien Europas. Aula-Verlag, Wiebelsheim, pp 555–605

Bülbül U, Kutrup B (2013) Morphological and genetic variations of *Ommatotriton* in Turkey. Anim Biol 63:297–312. doi: 10.1163/15707563-00002413

Canestrelli D, Porretta D, Lowe WH, et al (2016) The tangled evolutionary legacies of range expansion and hybridization. Trends Ecol Evol. doi: 10.1016/j.tree.2016.06.010

Dittrich-Reed DR, Fitzpatrick BM (2013) Transgressive hybrids as hopeful monsters. Evol Biol 40:310–315. doi: 10.1007/s11692-012-9209-0

Fitzpatrick BM, Shaffer HB (2007) Hybrid vigor between native and introduced salamanders raises new challenges for conservation. Proc Natl Acad Sci 104:15793–15798. doi: 10.1073/pnas.0704791104

Fontelles F, Guixé D, Martínez-Silvestre A, et al (2011) Hallada una población introducida de *Ommatotriton ophryticus* en el Prepirineo catalán. Boletín la Asoc Herpetológica Española 22:153–156.

Kats LB, Ferrer RP (2003) Alien predators and amphibian declines: review of two decades of science and the transition to conservation. Divers Distrib 9:99–110. doi: 10.1046/j.1472-4642.2003.00013.x

Kutrup B, Bülbül U (2011) Comparison of skeletal muscle protein bands and trunk vertebrae count between *Ommatotriton ophryticus nesterovi* and *O. o. ophryticus* populations in Turkey. Turk J Zool 35:579–584. doi: 10.3906/zoo-0904-6

Leigh J, Bryant D, Steel M (2016) PopArt 1.7.

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

Litvinchuk SN, Zuiderwijk A, Borkin LJ, Rosanov JM (2005) Taxonomic status of *Triturus vittatus* (Amphibia: Salamandridae) in western Turkey: trunk vertebrae count, genome size and allozyme data. Amphibia-Reptilia 26:305–323.

Maddison WP, Maddison DR (2015) Mesquite: a modular system for evolutionary analysis.

Martel A, Blooi M, Adriaensen C, et al (2014) Recent introduction of a chytrid fungus endangers Western Palearctic salamanders. Science 346:630–631. doi: 10.1126/science.1258268

Martínez-Silvestre A, Montori A, Oromi N, et al (2017) Detection of a Ranavirus in introduced newts in Catalonia (NE Spain). Herpetol Notes 10:23–26.

Meilink WRM, Arntzen JW, van Delft JJCW, Wielstra B (2015) Genetic pollution of a threatened native crested newt species through hybridization with an invasive congener in the Netherlands. Biol Conserv 184:145–153. doi: 10.1016/j.biocon.2015.01.022

Pimentel D, McNair S, Janecka J, et al (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. Agric Ecosyst Environ 84:1–20. doi: 10.1016/S0167-8809(00)00178-X

Sax DF, Gaines SD (2003) Species diversity: from global decreases to local increases. Trends Ecol Evol 18:561–566. doi: 10.1016/S0169-5347(03)00224-6

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169. doi: 10.1086/379378

van Riemsdijk I, Arntzen JW, Bogaerts S, Franzen M, Litvinchuk SN, Olgun K, Wielstra1 B (2017) The Near East as a cradle of biodiversity: a phylogeography of banded newts (genus *Ommatotriton*) reveals extensive inter- and intraspecific genetic differentiation. Mol Phylogenet Evol 114:73-81. doi: 10.1016/j.ympev.2017.05.028

Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos Trans R Soc Lond B Biol Sci 360:1859–1868. doi: 10.1098/rstb.2005.1717

Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. PNAS 105:11466–11473. doi: 10.1073/pnas.0801921105

Chapter 4

Testing an hypothesis of hybrid zone movement for toads in France

Published in Molecular Ecology, 2019

I. van Riemsdijk[1,2*], R. K. Butlin[3,4], B. Wielstra1,[2,3,5], J. W. Arntzen[1]

[1]Taxonomy and Systematics, Naturalis Biodiversity Center, Leiden, The Netherlands
[2]Institute of Biology Leiden, Leiden University, Leiden, The Netherlands
[3]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, The United Kingdom
[4]Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden
[5]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA
[*]Correspondence: I. van Riemsdijk isolde.vanriemsdijk@naturalis.nl

Abstract

Hybrid zone movement may result in substantial unidirectional introgression of selectively neutral material from the local to the advancing species, leaving a genetic footprint. This genetic footprint is represented by a trail of asymmetric tails and displaced cline centres in the wake of the moving hybrid zone. A peak of admixture linkage disequilibrium is predicted to exist ahead of the centre of the moving hybrid zone. We test these predictions of the movement hypothesis in a hybrid zone between common (*Bufo bufo*) and spined toads (*B. spinosus*), using 31 nuclear and one mtDNA SNPs along a transect in the northwest of France. Average effective selection in *Bufo* hybrids is low and clines vary in shape and centre. A weak pattern of asymmetric introgression is inferred from cline discordance of seven nuclear markers. The dominant direction of gene flow is from *B. spinosus* to *B. bufo* and is in support of southward movement of the hybrid zone. Conversely, a peak of admixture linkage disequilibrium north of the hybrid zone suggests northward movement. These contrasting results can be explained by reproductive isolation of the *B. spinosus* and *B. bufo* gene pools at the southern (*B. spinosus*) side of the hybrid zone. The joint occurrence of asymmetric introgression and admixture linkage disequilibrium can also be explained by the combination of low dispersal and random genetic drift due to low effective population sizes.

Keywords: Admixture linkage disequilibrium, asymmetric reproductive isolation, *Bufo bufo*, *Bufo spinosus*, cline coupling, hybrid zone movement

Introduction

During allopatric speciation, a taxon's range is split by a physical barrier, and the vicariant populations gradually diverge through processes such as mutation, natural selection, and genetic drift (Mayr, 1942; Lande, 1980; Wu & Ting, 2004). Diverged taxa may later meet in secondary contact and form a hybrid zone, for example

when taxa expand their ranges from glacial refugia in the postglacial era (Barton & Hewitt, 1985; Hewitt, 1988, 2011). Upon secondary contact the two populations may have established reproductive isolation, inhibiting exchange of genetic material (Rice, 1998). Conversely, when the fitness of hybrid populations is equal or increased compared to the fitness of parental populations, gene flow is extensive, and the two populations will merge (Anderson & Stebbins, 1954; Arnold, 2004; Seehausen, 2004; Hedrick, 2013). In between these extremes, a reduced hybrid fitness precludes merging and facilitates limited introgression (Mallet, 2005).

When hybrid fitness is reduced independent of the environment, a narrow tension zone exists between the two species (Moore, 1977; Barton & Hewitt, 1985). In the classical hybrid zone literature such tension zones are thought to stabilise where population density is low, or at an ecological barrier to dispersal (Endler, 1977; Barton & Hewitt, 1985). Recent studies suggest that hybrid zone movement is more common than previously thought (Arntzen & Wallis, 1991, 1999; Buggs, 2007; Roy, O'Connor, & Green, 2012; Taylor et al., 2014; Leaché, Grummer, Harris, & Breckheimer, 2017; Wielstra, Burke, Butlin, & Arntzen, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017; Ryan et al., 2018). Hybrid zone movement occurs when one parent species has a higher fitness than the other, such as through a competitive advantage, asymmetric hybrid fitness effects, or environmental change (Buggs, 2007). Introgression caused by hybrid zone movement is thought to affect selectively neutral markers distributed randomly across the genome, resulting in genetic traces of the displaced species in populations of the advancing species (Moran, 1981; Currat, Ruedi, Petit, & Excoffier, 2008). In contrast, under adaptive introgression, alleles from one species with a positive effect in the other species may introgress through the hybrid zone, regardless of whether the zone is stable or moving (Seehausen, 2004; Mallet, 2005; Hedrick, 2013; Schmickl, Marburger, Bray, & Yant, 2017). Adaptive introgression affects only the selected marker and markers that are nearby on the chromosome (physical linkage), or markers that interact functionally with the marker under selection (functional linkage; Barton & Hewitt, 1985; Baird, 2015; Sedghifar, Brandvain, & Ralph, 2016). Only a fraction of the genome is involved in adaptive introgression, whereas hybrid zone movement results in genome-wide introgression (Currat et al., 2008; Wielstra, Burke, Butlin, Avcı, et al., 2017).

The transition from one species to the other through the hybrid zone can be described by a set of allele frequency gradients, i.e. geographical clines (Schmickl et al., 2017). The positions and shapes of geographical clines are indicative of evolutionary processes involved in hybrid zone formation (Barton & Hewitt, 1985). Two types of cline variation are distinguished based on cline position (coincident or displaced) and cline shape (concordant or discordant). Coincident clines share the same cline centre, whereas displaced clines have a cline centre away from the majority of the clines for other loci. Concordant clines are similar in shape (described by e.g. width and tail shape; Szymura & Barton, 1991), whereas discordant clines have a shape divergent from the majority of the clines. Coincident clines are considered evidence of a hybrid zone in a stable position over time (Abbott et al., 2013). Concordant and narrow clines (steep at the centre and with distinct tails) suggest that the hybrid zone is maintained by a balance between strong selection and dispersal, as opposed to displaced and wide clines indicating that selection against hybrids is low (shallow slope without distinct tails; Barton & Gale, 1993). Displaced and discordant clines exist in many species,

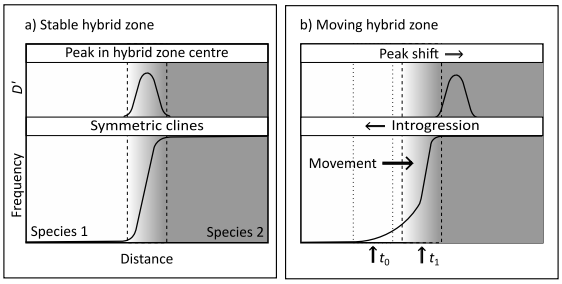


Figure 4.1: Schematic representation of an hypothesis of hybrid zone movement. In contrast to a stable hybrid zone (a) where the peak of admixture linkage disequilibrium (D', y-axis, top plot) is situated in the hybrid zone centre and clines are symmetric (frequency, y-axis, bottom plot, x-axis both plots is distance along a transect), a moving hybrid zone (b) is expected to have a peak of admixture linkage disequilibrium shifted ahead of the hybrid zone and a tail of introgression in the wake of the hybrid zone (t_0 is the previous hybrid zone position, t_1 the current position). Dashed lines indicate the hybrid zone centre.

both for nuclear and mitochondrial markers, and may be caused by sex-biased gene flow or by adaptive introgression (Mallet, 2005; Excoffier, Foll, & Petit, 2009; Toews & Brelsford, 2012; While et al., 2015; Sloan, Havird, & Sharbrough, 2016; Bonnet, Leblois, Rousset, & Crochet, 2017). However, displaced and discordant clines may also be caused by hybrid zone movement, when a trail of genetic material from the overtaken species remains present in the overtaking species (Rohwer, Bermingham, & Wood, 2001; Gay, Crochet, Bell, & Lenormand, 2008; Excoffier et al., 2009; Arntzen, de Vries, Canestrelli, & Martínez-Solano, 2017; Wielstra, Burke, Butlin, Avcı, et al., 2017).

Whether or not markers reflect past or ongoing demographic processes, including hybrid zone movement and adaptive introgression, is dependent on factors such as the number and genomic distribution of barrier loci, mutation rate, and recombination (reviewed in Ravinet et al., 2017). Barrier loci are genes which inhibit the merging of two diverged populations (Abbott et al., 2013; Barton, 2013). A stronger overall barrier to gene flow is established when multiple barrier effects become coincident by processes summarised as coupling (reviewed in Butlin & Smadja, 2017). If genome-wide coupling of barrier effects occurs, the geographic clines for markers are expected to be coincident even if they are not physically or functionally linked to a barrier gene (Barton, 1983; Barton & Gale, 1993; Bierne, Welch, Loire, Bonhomme, & David, 2011; Vines et al., 2016). To overcome the coupling effect of multiple barrier genes and so to cause gene flow across the hybrid zone, the positive selection on a single gene will have to be disproportionally strong (e.g. Vines et al., 2016).

In addition to physical and functional linkage, associations between markers in hybrid zones can be caused by common descent. This is reflected by an increase in admixture linkage disequilibrium in the hybrid zone as alleles originating from the same parent species tend to be found together within the genomes of early generations of hybrid offspring. Recombination during subsequent backcrossing breaks down admixture linkage disequilibrium (Barton & Gale, 1993; Baird, 2015). However, when the hybrid zone is moving, the peak of admixture linkage disequilibrium is predicted to be positioned just ahead of the hybrid zone centre, where individuals with little history of recombination are involved in reproduction (Gay et al., 2008; Wang et al., 2011). This allows us to make predictions based on the hypothesis of hybrid zone movement: there will be a tail of introgression in the wake of the moving zone combined with a shift of the peak in admixture linkage disequilibrium ahead of the

movement. This contrasts with the predictions for a stable situation, where clines are symmetric and the peak of admixture linkage disequilibrium appears in the centre of the hybrid zone (Fig. 4.1).

We test for a genetic footprint of movement in a hybrid zone in northwest France between two morphologically similar but genetically distinct species of toad, the common, *Bufo bufo* (Linnaeus, 1758) and the spined toad *B. spinosus* Daudin, 1803 (Recuero et al., 2012; Arntzen, Recuero, Canestrelli, & Martínez-Solano, 2013; Trujillo, Gutiérrez-Rodríguez, Arntzen, & Martínez-Solano, 2017). While *B. bufo* survived the last glacial maximum in Italy and the Balkans, *B. spinosus* survived in southern France and Spain (Arntzen et al., 2017). Based on four nuclear gene coding markers, twelve microsatellites, and two mitochondrial SNP markers, the width of the hybrid zone (30 km) compared to the dispersal distance (1.3 km generation^{-1}) suggested that selection against hybrids restricts gene flow through the hybrid zone (Arntzen et al., 2016). In the southeast of France, the *Bufo* hybrid zone was hypothesised to move southwards based on traces of introgression of one nuclear gene coding marker (of four tested), and discordance between nuclear, mtDNA, and morphologic clines (Arntzen et al., 2017). In the northwest of France, a similar dataset including microsatellite data, showed introgression towards the north in two nuclear markers out of four, but was

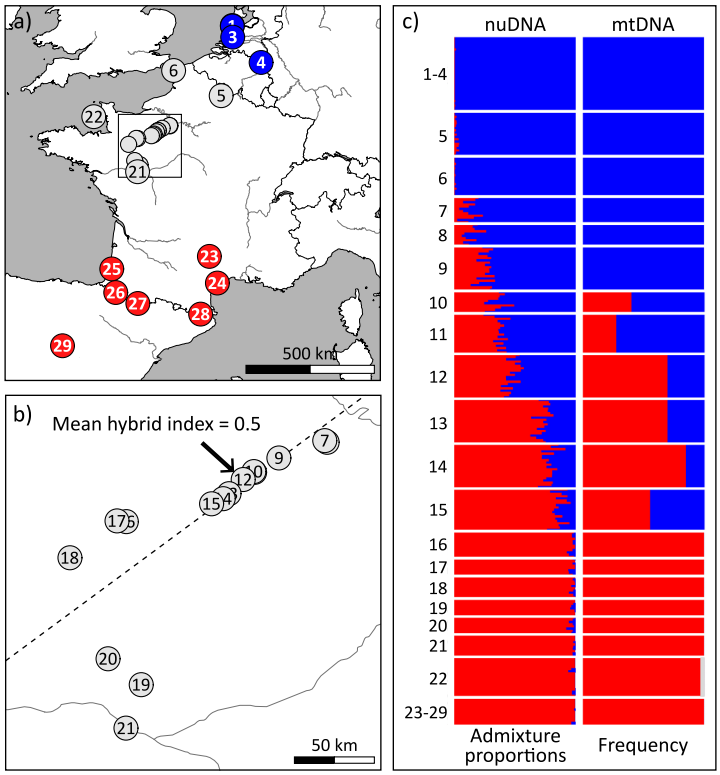


Figure 4.2: Map of Western Europe (a) with reference populations to determine diagnostic nature of markers for *Bufo bufo* (blue circles) and *B. spinosus* (red circles), and transect populations (grey circles). Panel b details the transect orientation (dashed line) and the position of hybrid zone populations. Panel c shows bar plots with individual admixture proportion (Structure Q scores) on the left and mtDNA allele frequency per population on the right. The grey bar refers to one missing mtDNA data point.

not analysed from the perspective of hybrid zone movement at the time (Arntzen et al., 2016). This scenario provides the opportunity to test for the hypothesis of hybrid zone movement in the northwest of France. We improve the resolution using 31 gene coding markers and one mtDNA marker to interpret interspecific gene flow in the light of the average effective selection on a locus, and place our findings in the context of results from the literature.

Materials and Methods

Sample collection and preparation

All DNA extracts were available from previous studies (Arntzen, Wilkinson, Butôt, & Martínez-Solano, 2014; Arntzen et al., 2016, 2017; Arntzen, McAtear, Butôt, & Martínez-Solano, 2018). Eight individuals from four *B. bufo* reference populations and twelve individuals from seven *B. spinosus* reference populations, all positioned > 500 km away from the centre of the hybrid zone (blue and red circles in Fig. 4.2a, 4.2b), served to identify species diagnostic single nucleotide polymorphisms (SNPs) in a test run. Another 268 individuals from 29 populations, including transect populations (grey circles in Fig. 4.2a, 4.2b) and additional samples from the reference populations were genotyped and included in the dataset, to a total of 306 individuals. Distances between transect populations were measured in a straight line, which follows the

Table 4.1: Overview table with the number of individuals sampled per population, the distance in kilometres (km), and the mean hybrid index, pooled for the reference populations (1-4 for *B. bufo* and 23-29 for *B. spinosus*), and for each transect population (5-22).

Population	Individuals sampled	Distance (km)	Hybrid index (HI)
1	10		
2	10		
3	7	0	0.03
4	6		
5	20	174	0.06
6	18	335	0.03
7	12	468	0.16
8	10	469	0.17
9	20	508	0.28
10	10	530	0.38
11	18	530	0.39
12	20	541	0.50
13	20	555	0.74
14	20	561	0.79
15	19	571	0.86
16	12	648	0.96
17	8	644	0.96
18	10	689	0.97
19	8	680	0.98
20	8	695	0.96
21	10	705	0.99
22	18	791	1.00
23	2		
24	2		
25	2		
26	1		
27	2	1222	1.00
28	1		
29	2		

direction of the transect in Arntzen et al. (2016), using a custom R script. Distances for reference populations were measured with Google Earth Pro v.7.3.0 (Table 4.1 and Table S.1).

SNP design based on transcriptome data

Two transcriptomes from liver tissue of a *B. bufo* individual from population 6 (Audresselles, France), and a *B. spinosus* individual from population 18 (Jublains, France) were sequenced commercially by ZF Screens, Leiden, on the Illumina HiSeq 2000 platform. The data were filtered with Trimmomatic v.0.35 software (Bolger, Lohse, & Usadel, 2014), and assembled with Trinity v.2.1.1 (Grabherr et al., 2011; Haas et al., 2013). When using exons from transcriptome data for SNP design in expressed genes, paralogs (gene copies) should be excluded to avoid interpreting SNPs between paralogs as SNPs within orthologues (e.g. Ryynänen & Primmer, 2006). Exon boundaries need to be taken into account, as primers designed across exon boundaries are the main cause of genotyping assay failure (Wang et al., 2008). First, a BLAST search (Altschul, Gish, Miller, Myers, & Lipman, 1990) was used to identify paralogs within the *B. bufo* transcriptome assembly. Next, a pipeline for exon boundary identification (Niedzicka, Fijarczyk, Dudek, Stuglik, & Babik, 2016) was used, employing the well-annotated *Xenopus tropicalis* (Gray, 1864) genome (genome version JGI4.2). The *X. tropicalis* genome was also used to annotate markers (after testing for diagnostic differences between the two species, 30 of 31 markers eventually used were annotated, one remained unidentified and was named exon_1). Finally, a second BLAST of the selected exons against the transcriptome of *B. spinosus* was used to exclude potentially undetected paralogs and to identify SNPs. A detailed description of the SNP design from transcriptome data is presented in the Supplemental Information text.

SNP validation

Fluorescence-based genotyping (Semagn, Babu, Hearne, & Olsen, 2014) was used in the Kompetitive Allele-Specific PCR (KASP) genotyping system at the SNP genotyping facility of the Institute of Biology, Leiden University. Primer design, PCR setup, and data visualisation followed Arntzen et al. (2016). For 96 exon sequences, primers were designed with the Kraken software (LGC genomics, UK). After a first run, three SNPs were excluded as they failed for all individuals of one species, presumably due to a mutation at a primer binding site. Markers were initially considered diagnostic if they showed species-specific alleles in the test dataset. This resulted in 32 nuclear DNA SNPs (Table S.2). Because of limitations due to plate dimensions in the KASP system, the remainder of the samples was analysed with 31 SNPs (by excluding the least well performing marker in the test dataset, scaf4, for which 6 out of 20 reference individuals had missing data) and we added one previously published diagnostic mtDNA SNP (16S; Arntzen et al., 2016). We excluded five individuals from the final dataset with more than ten SNPs missing (presumably because of low quality DNA). After assessment of the final dataset of 306 individuals, 29 of 31 nuclear SNPs were considered species-diagnostic, defined by minor allele frequencies of < 5.0 % in reference populations in the final dataset (Table S.3). Missing data amounted to 5.2% (Table S.4).

Hardy-Weinberg proportions and marker linkage

Because SNPs in genes are more prone to be under direct or indirect selection than non-coding DNA, we were cautious to exclude any markers showing outlier behaviour. Signals of non-random mating or survival were tested for by calculating the Hardy--Weinberg proportions (heterozygote deficit and excess) with the R package genepop based on the program GENEPOP v.1.0.5 (Rousset, 2008). As the Bonferroni correction (Rice, 1989) can be overly conservative (e.g. Narum, 2006), we chose to account for independence of tests within markers (Pc for N = 31). Deviations from Hardy--Weinberg proportions by heterozygote excess were not significant (Pc > 0.05; Table S.5). Deviations by heterozygote deficit were significant for the marker egflam in five populations (Pc < 0.05), hence this marker was excluded in the HZAR cline fitting analysis and admixture linkage disequilibrium calculations (see below).

To infer independence of introgression, close physical or functional linkage of markers should be investigated. As some of the 31 nuclear markers are bound to be on the same chromosome (*Bufo* has 22 chromosomes; Olmo, Gargiulo, & Morescalchi, 1970; Olmo, 1973), we used a test of pairwise linkage disequilibrium (LD), indicative of close physical linkage of markers, based on the log likelihood ratio statistic (G-test) with the R-package genepop (Rousset, 2008). One pair of markers, exon_1 and ttc37, was found to be in significant pairwise LD in two populations (Pc < 0.05; Table S.6). As these populations were located in the hybrid zone, where we expect the effect of admixture linkage disequilibrium to increase the number of markers in pairwise LD, we kept the two markers in downstream analyses, but we checked that this marker pair did not distort the general patterns described below. To investigate functional linkage, a protein-interaction network was analysed with STRING v.10.5 (Szklarczyk et al., 2015) on the basis of the *X. tropicalis* genome, and 30 out of 31 annotated nuclear markers. A protein function description was recorded for each annotated marker (Table S.7). The markers brca2 and rfc1 were found to be functionally linked. As these markers were neither deviating from Hardy-Weinberg proportions nor in pairwise linkage disequilibrium within populations, they were included in all downstream analyses but, again, we checked that this marker pair did not distort the general patterns described below. Note that it remains possible that individual markers are under direct or indirect selection, and the current and future analyses using these markers need to take this into account.

Population structure

Allocation of individuals to genetically defined groups was done using all 31 nuclear markers in Structure v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with ten independent chains of one up to ten genetic clusters (K), a burn in of 10,000, and a chain length of 25,000 under the admixture model following recommendations of Benestan et al. (2016). Other settings were left at default. Credibility intervals (95%) were estimated and convergence was checked comparing log likelihood and similarity of admixture proportion (Structure Q scores) between runs. The results were summarized with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The best value of K was determined using the Evanno method (Evanno, Regnaut, & Goudet, 2005; Table S. 8), and results were visualised using the R

package POPHELPER (Francis, 2017). The admixture proportions are used in the cline analysis (see below) to approach an overall cline shape and position that summarizes all nuclear clines.

Cline analyses

Classic equilibrium cline models, describing a sigmoid change in allele frequency or phenotype across the hybrid zone, were fitted with the R package HZAR (Derryberry, Derryberry, Maley, & Brumfield, 2014) for 30 nuclear markers, mtDNA, and admixture proportion (K = 2), using a set of custom R scripts provided by G. Derryberry. Marker egflam was not analysed with HZAR as it was not behaving according to Hardy--Weinberg proportions. First, 30 maximum likelihood estimation searches were performed with random starting parameters, followed by a trace analysis of 60,000 generations on all models with a delta Akaike Information Criterion corrected for small-sample-size (ΔAICc) below ten. Fifteen model variants can be fitted in HZAR, based on all possible combinations of trait interval (allele frequency at the transect ends; three types) and cline shape. The different cline shapes represent either a single sigmoid curve, or a combination of (1) the central sigmoid portion and (2) shallower exponential decay curves at one or both ends (tails) with slopes shallower than expected from the central portion. We refer to these alternatives as 'tail types' (five types). Tail types were fitted as an exponential tail to the left (L), exponential tail to the right (R), both tails exponential with independent parameters (B), both tails exponential but with the same parameters mirrored on the cline centre (M), and no exponential tails fitted (N). For cline tails estimated separately (B), we assessed tail slope (τ; shallow tails reflected by low parameter values) and distance from the cline centre where a tail started (δ; short distances reflected by low parameter values). When τ and δ are both low on one side and both high at the other side of the hybrid zone, introgression is asymmetric. Convergence of the HZAR analysis was visually assessed in trace plots. Significance of cline model fits was determined by calculating ΔAICc for each best model (ΔAICc > 2 compared to the second best model; Table S.9, S.10, Fig. S.1). Deviation from an equal frequency of left and right tails (as expected in a stable hybrid zone) was tested for all markers with a coincident cline centre with the chi-square test for equal probabilities.

Admixture linkage disequilibrium

Using variance in hybrid index (HI, see also Table 4.1), admixture linkage disequilibrium (D') can be calculated, and subsequently lifetime dispersal distance (σ) weighted for individuals in aquatic (pre-) and terrestrial (post-metamorphosis) life stages, expected cline width under neutrality, and average effective selection on a locus (s*) can be calculated following Barton & Gale (1993). This was done using 29 nuclear markers, excluding marker egflam, which was not within Hardy-Weinberg proportions and marker banp, for which the cline centre was displaced compared to other clines (Fig. S.2). Average effective selection is the selection pressure on a locus at the zone centre due to direct selection or association with other loci under selection. A few input parameters were needed for these calculations (Table S.11). A mean recombination rate between marker pairs of 0.4997 was calculated following formula (6) from Macholán et al. (2007), using the number of chiasmata per bivalent

for *Bufo bufo* (1.95; Wickbom, 1945), and the number of chromosomes for *B. bufo* (N = 22). A generation time of 2.5 years for *Bufo* toads at the latitude of the hybrid zone (mean of three years in females and two years in males; Hemelaar, 1988), and initial secondary contact at 8,000 years ago following Arntzen et al. (2016) were used as input parameters. The width of the hybrid zone was derived from a general sigmoid cline model following the 'no-tails' formula in HZAR (Derryberry et al., 2014) fitted to the HI. At the cline centre, D' was estimated from its regression on p*(1-p), where p is the average over loci of the frequency of the *B. bufo* allele at a sample location. Means and 95% confidence intervals (CI) for cline width, D' at the centre, dispersal distance, expected cline width under neutrality, and effective selection were based on 1,000 bootstrap replicates of the original genotype dataset (with replacement, maintaining original sample size; Table S.11), using custom R scripts. The amplitude and position of the peak in D' were estimated by fitting a Gaussian curve following Gay et al. (2008).

Hybrid index analysis

To visualise potential unequal contribution of both species to the genomic composition of admixed individuals, we used the R package HIest (Fitzpatrick, 2012). This analysis scales the number of markers derived from each parental species (S) to heterozygosity, which is calculated as the fraction of heterozygous markers with variants inherited from both parent species, within an individual (HI). We used 29 markers with an allele frequency difference between species more than 0.8 (Larson, Andrés, Bogdanowicz, & Harrison, 2013). To assess unequal inheritance of mtDNA in hybrid offspring, we coloured the data points by mtDNA type.

Results

The number of genetic clusters (K) supported by Structure was two, concordant with the two toad species (Table S.8). Species admixture was recorded in populations 7-15. Individuals of populations 7-9 north of the hybrid zone centre were hybrid populations with a relatively high frequency of *B. spinosus* nuclear alleles when compared to the number of *B. spinosus* mtDNA haplotypes (Fig. 4.2c). Credibility intervals of the admixture proportion estimates are shown in Fig. S.3.

HZAR clines were partitioned according to tail type (Fig. 4.3, for a plot including all clines see Fig. S.2, Tables S.9, S.10). Eight nuclear marker clines best fitted a left tail of introgression (northwards into *B. bufo*) and for seven markers the fit was significantly better than alternative models (Fig. 4.3a). Two nuclear marker clines had a right tail of introgression (southwards into *B. spinosus*), of which one had a significantly better fit than alternative models (Fig. 4.3b). For four nuclear marker clines, both tails were exponential with independent parameters, of which three had a significantly better fit than alternative models (Fig. 4.3c). For these clines, tail slope (τ) and distance from the cline centre where the tail started (δ) showed an increase in one parameter (e.g. δ) and a decrease in the other (e.g. τ) in all cases. Such results are not straightforward to interpret in terms of directional introgression and these clines were not taken into account when assessing unidirectional introgression. For eleven nuclear markers, clines had mirrored tails of which six had a significantly better fit than alternative models (Fig. 4.3d). For three nuclear markers, clines had no tails, and one marker had

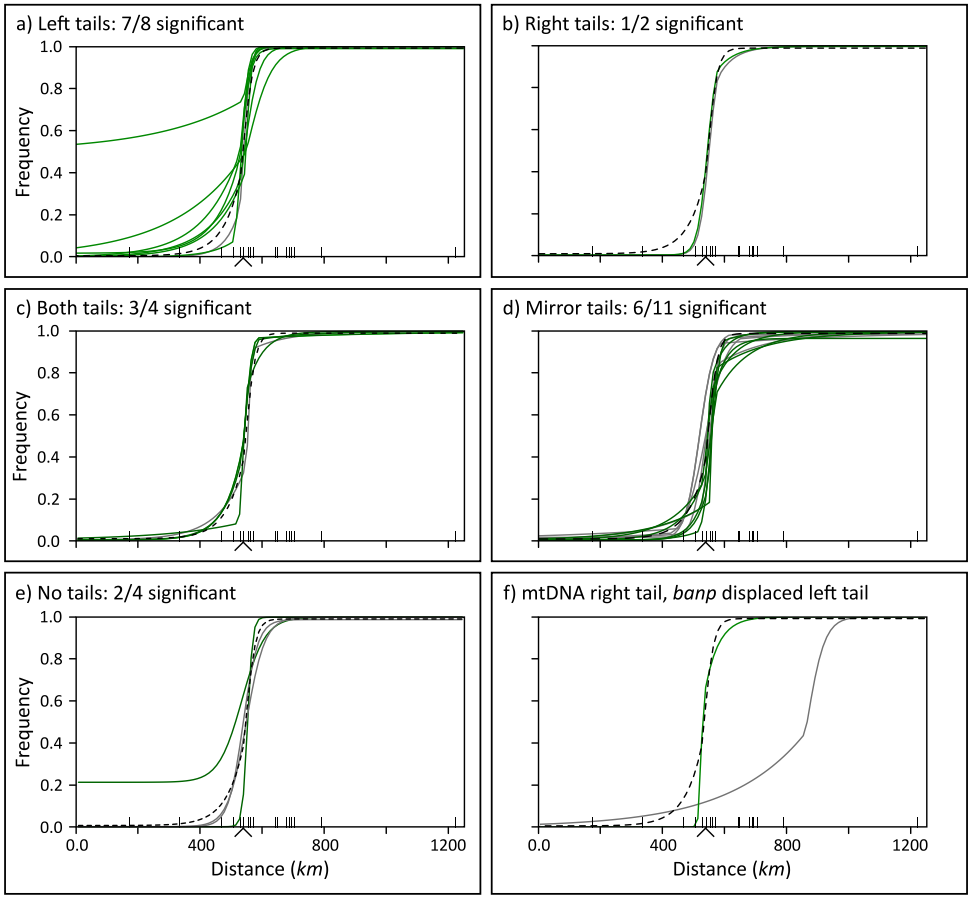


Figure 4.3: Best fitting clines for 30 nuclear markers and the mtDNA marker determined with HZAR. Frequencies of 0 and 1 represent pure *Bufo bufo* and *B. spinosus* genotypes. Left tail clines are in panel (a), right tail clines are in (b), both tail clines in (c), mirror tail clines in (d), and no tail clines are in (e). Panel f shows the mtDNA cline with a significant right tail and the nuclear marker banp with a significantly displaced cline centre and a non-significant left tail. Cline models that fit significantly better than the next best model (ΔAICc > 2) are shown by green lines and the others by grey lines. Sample localities on the transect are shown on the x axis by inside ticks and the arrow (^) indicates the position of the hybrid zone centre based on the admixture proportion cline (dashed line).

a significantly better fit than alternative models (Fig. 4.3e). All nuclear markers had cline centres within 30 km of the admixture proportion cline centre (positioned at 539 km) except for banp, with a cline centre > 300 km to the south, inside the range of B. spinosus (Fig. S.2). This cline had a non-significant left tail into *B. bufo* (Fig. 4.3f). The mitochondrial marker had a significant right tail into *B. spinosus* (Fig. 4.3f). The functionally linked markers brca2 and rfc1 had similar cline models with mirrored tails and overlapping 95% credible cline regions for centre and width, and showed no pattern of introgression deviating from the admixture proportion cline. In summary, seven nuclear markers showed significant introgression of *B. spinosus* into *B. bufo* by cline discordance, whereas two nuclear markers showed significant introgression the other way around, one by cline discordance and one by displacement. The inequality of introgression was non-significant (Chi-square test for equal probabilities, Chi2 =

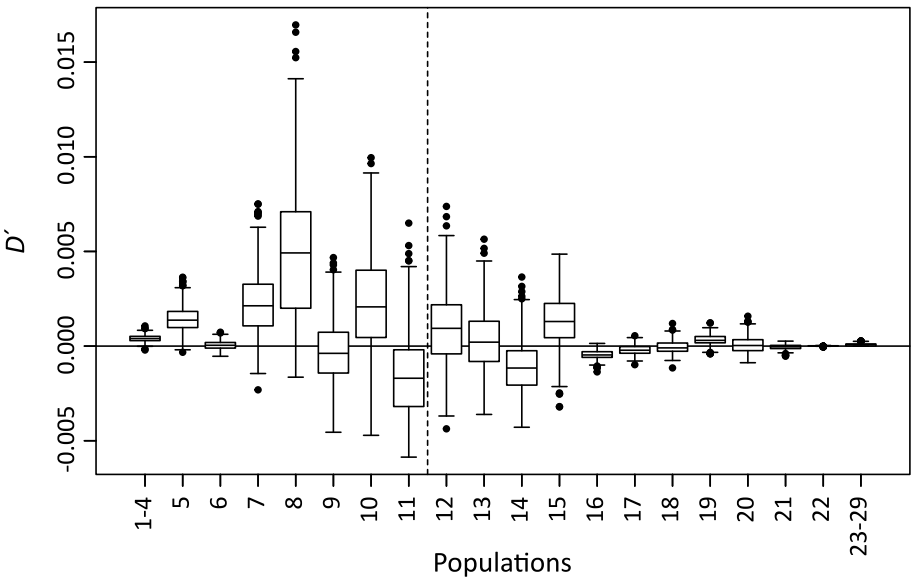


Figure 4.4: Admixture linkage disequilibrium (D') recorded over a transect from the north (*Bufo bufo,* populations 1-4) to the south (*B. spinosus*, populations 23-29). Note that the horizontal axis does not represent linear geographical distance. The dotted vertical line is the position of the hybrid zone centre inferred from the admixture proportion cline. Estimates for D' are obtained by bootstrap analysis and shown by box-and-whisker plots.

2.78, df = 1, P = 0.096).

Populations 9-15 within the hybrid zone had higher values of D' than populations outside the hybrid zone (Fig. 4.4). This is reflected by an increase in pairwise linkage disequilibrium among loci (Table S.6). Populations 7, 8, and 10 north of the hybrid zone centre, and beyond the area of rapid allele frequency change, had larger D' with larger positive deviations from zero than other populations. However, it was not possible to fit a Gaussian curve through these data points, probably because of the large variation in D', and the position and amplitude of the peak of D' remain undetermined.

Estimated lifetime dispersal (σ), based on D' at the cline centre, was 2.2 km generation$^{-1/2}$ (95% CI 0.6-3.7). The observed cline width based on HI was 114 km (95% CI 103-132). Assuming no selection against hybrids, the expected cline width would be 312 km (95% CI 91-532; Table S.11). The average effective selection on a locus (s*) was 0.0017 (95% CI 0.0001-0.0040). Despite variation among clines suggesting weak coupling, the presence of an admixture linkage disequilibrium peak supports the use of s* as an average value to describe the hybrid zone. HIest showed a continuum of *B. bufo* to *B. spinosus* genotypes in which hybrids possessing mtDNA haplotypes typical of either *B. bufo* or *B. spinosus* were about equally frequent (Fig. S.4). The peak of ancestry is at 50% heterozygosity, indicating no asymmetry in the contribution of both species to the hybrid population.

Discussion

The features describing the *Bufo* hybrid zone in northwest France in the current study are broadly in line with previous descriptions, but provide more resolution and

estimates of key parameters. We found that individual marker cline positions and shapes vary (e.g. Fig. S.2), widening the admixture proportion cline. We inferred weak effective selection per locus against hybrids (s*=0.0017), but with low precision of the estimate. Even when effective selection is low and clines are wide, a moving hybrid zone may stagnate at an ecotone, limiting further movement of the zone (Endler, 1977; Moore, 1977; Barton & Hewitt, 1985; Buggs, 2007; Gompert, Mandeville, & Buerkle, 2017). The section of the *Bufo* hybrid zone studied here appears to be located at a weak ecotone in a hilly landscape formed by the 'Collines de Normandie', with *B. bufo* at a lower and *B. spinosus* at a higher altitude (Arntzen et al., 2016). The genetic estimate of lifetime dispersal of 2.2 km generation$^{-1/2}$ diverges from the average maximum observed in field studies, but the field estimate is within the CI for the genetic estimate (1.3 km; Smith & Green, 2005; Daversa, Muths, & Bosch, 2012; Trochet et al., 2014). In amphibians, dispersal distance based on genetic data is more often found to be higher compared to field studies, because mark-recapture, seasonal, or field studies covering a small area are at risk of underestimating (rare) long-distance dispersal (Smith & Green, 2005).

The hybrid zone was calculated to be 114 km wide, based on the hybrid index for 30 nuclear markers, and width was 63 km based on the cline of the admixture proportion for 31 nuclear markers. A narrower width (30 km) was earlier recorded based on the first PCA axis of 12 microsatellites, whilst in the same study cline width for four nuclear markers ranged from 27 to 91 km, and morphological cline widths exceeded 200 km (Arntzen et al., 2016). Microsatellites are noncoding, have higher mutation rates, and they are more suitable for investigating small scale evolutionary processes, therefore they may reflect more recent demographic processes (Ellegren, 2000). This poses the question what is causing a more spatially restricted transition in the microsatellite markers as opposed to a wider transition in nuclear gene markers, as one would expect stronger selection against hybrids and thus narrower clines in markers from coding regions than in microsatellites. Including more markers would cause the hybrid index or admixture proportion clines to get wider, because drift causes the centres of individual marker clines to vary. The larger number of markers in this paper may thus explain the wider zone currently reported compared to the previously reported width based on less markers. Confidence intervals overlap for many of these width estimates and all widths are large relative to dispersal. This indicates the general conclusion of weak selection is robust.

We examined patterns of gene flow and admixture linkage disequilibrium to test the hypothesis of movement of the hybrid zone in two common toad species (*B. bufo* and *B. spinosus*) in northwest France. Introgression from *B. spinosus* into *B. bufo* was more frequent than introgression the other way around, a result that was found before with four nuclear markers (Arntzen et al., 2016, 2017). Disregarding whether cline model fits are significant, eleven markers show introgression from *B. spinosus* into *B. bufo* and two show introgression the other way around, a significant difference (Chi--square test for equal probabilities, $Chi^2 = 6.23$, $P = 0.013$). Cases where introgression is not explicit can also point out methodological difficulties, such as determining unidirectional introgression from cline shape parameters. To better compare the asymmetry of introgression between individual markers, future methods could compare e.g. the area underneath and above the geographic cline on both sides of the cline centre. Further simulations are needed to provide clear expectations for stable

and moving hybrid zones.

Three populations with elevated admixture linkage disequilibrium (D') suggest a peak of early-generation hybrid offspring north of the hybrid zone. Fitting a Gaussian curve through D' proved impossible, as there was too much variance within the data, especially in populations 5, 9 and 11. The Gaussian curve is the shape of the peak expected under a symmetrical cline model (Gay et al., 2008). The absence of a clear peak can be caused by asymmetry in the cline model or insufficient sample size of individuals, populations, or markers, or a combination of both. More data should be generated to be able to determine the shape and position of the peak with higher precision. While the unidirectional introgression is in line with southward movement of the hybrid zone (Buggs, 2007), the northern position of a peak in D' is in line with movement in the opposite direction (Gay et al., 2008; Wang et al., 2011). Since our results do not fully support the original hypothesis of (southward) hybrid zone movement in *Bufo*, alternative explanations have to be considered, in particular for the position of the peak in admixture linkage disequilibrium, for which we compare our result with other case studies.

Concordant and narrow clines suggest that a hybrid zone is maintained by a balance between strong selection and dispersal, as opposed to wide and displaced clines indicating that selection against hybrids is low (Barton & Gale, 1993). By surveying the hybrid zone literature, we found eight studies with empirical data on effective selection (i.e. s*) and cline centre and shape in a wide variety of organisms (Table S.12; Mallet et al., 1990; Szymura & Barton, 1991; Phillips, Baird, & Moritz, 2004; Macholán et al., 2007; Kawakami, Butlin, Adams, Paull, & Cooper, 2009; Carneiro et al., 2013, Baldassarre, White, Karubian, & Webster, 2014; Hollander, Galindo, & Butlin, 2015; Wielstra, Burke, Butlin, & Arntzen, 2017). Six studies report substantial effective selection (0.19 – 0.37) with mostly concordant, coincident and steep clines, and three studies report low average effective selection (0.007 – 0.011) with multiple discordant and displaced clines. Our results for *Bufo*, with low effective selection and substantial cline variability, fall into the latter category.

We also consulted the literature on the direction of introgression and the location of elevated levels of admixture linkage disequilibrium in hybrid zones. Two studies document a peak in admixture linkage disequilibrium opposite to where introgression takes place (Fig. 4.5a), and one study documents a peak of admixture linkage disequilibrium and introgression on the same side (Fig. 4.5b). In the hybrid zone of the gull species *Larus glaucescens* (Naumann, 1840) and *L. occidentalis* (Audubon, 1839), genotypic and phenotypic clines showed introgression towards the north, and a peak in admixture linkage disequilibrium towards the south, indicating a southward movement of the hybrid zone (Fig. 4.5a; Gay et al., 2008). In the hybrid zone of the house mouse subspecies *Mus musculus musculus* Linnaeus, 1758 and *M. m. domesticus* Schwarz & Schwartz, 1943 introgression to the east and a peak of admixture linkage disequilibrium to the west, supported a westward movement of the hybrid zone (Fig. 4.5a; Wang et al., 2011). Conversely, in the hybrid zone of the salamander subspecies *Ensatina eschscholtzii eschscholtzii* Gray, 1850 and *E. e. klauberi* Dunn, 1929, unidirectional introgression coincides with a peak of admixture linkage disequilibrium (Fig. 4.5b; Devitt et al., 2011). This result was explained by asymmetric pre- or postzygotic isolation. On the side of the hybrid zone where reproduction is

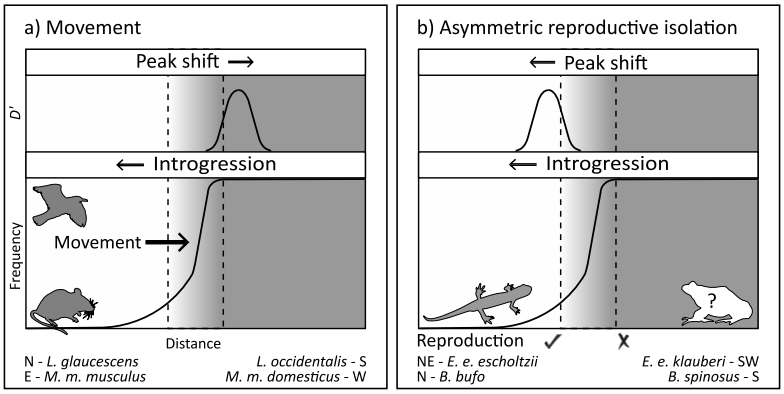


Figure 4.5: Schematic representation of inferred hybrid zone movement in (a) the gulls *Larus glaucescens* and *L. accidentalis* (Gay et al., 2008), and the house mice *Mus musculus musculus* and *M. m. domesticus* (Wang et al., 2011), and asymmetric reproductive isolation in (b) the salamanders *Ensatina eschscholtzii eschscholtzii* and *E. e. klauberi* (Devitt et al., 2011), and possibly also in the toads *Bufo bufo* and *B. spinosus* in the present study. In panel b, the tick mark shows where reproduction can take place and the cross shows where reproduction cannot take place. The y-axis of the top plot shows admixture linkage disequilibrium (D'), the y-axis of the bottom plot shows frequency, and the x-axis of both plots show distance along the transect. Cardinal directions are given by N, E, S and W. The question mark in panel b refers to the uncertainty of unidirectional introgression and of the magnitude and position of the peak of D' in the *Bufo* hybrid zone.

reduced, introgression and admixture of genetically distinct individuals are absent, whereas at the other side where hybridisation succeeds, introgression and admixture of genetically distinct individuals occur. A deficit of hybrids with *E. e. eschscholtzii* mtDNA confirms that offspring of female *E. e. klauberi* and male *E. e. eschscholtzii* are successfully reproducing, whilst offspring of other parental combinations occur rarely.

When comparing patterns of introgression between hybrid zones of different biological systems, dispersal rate and effective population size need to be taken into account (Canestrelli et al., 2016; Ravinet et al., 2017). Introgression generally increases when dispersal rate is high, because more genetically distant populations interbreed (Nichols & Hewitt, 1994). The effect of a higher dispersal on introgression is greater for a narrow hybrid zone where selection against hybrids is high, than for a wide zone where selection against hybrids is low, because in a narrow zone the distance between genetically distinct individuals is smaller (Barton & Gale, 1993). Therefore, we chose to express dispersal in terms of the number of generations needed to cross the hybrid zone, as this summarises width in relation to dispersal, and thus also in relation to selection against hybrids. To include the effect of genetic drift in the comparison of hybrid zones, effective population size can be used. When effective population size is low, an increase in the effect of random genetic drift obscures genetic patterns of demographic and evolutionary history, and clines are predicted to be more randomly distributed in the landscape and vary in shape, despite a high selection against hybrids (Polechová & Barton, 2011). The *Larus* hybrid zone would take five generations of unimpeded dispersal to cross (Fig. 4.5a; Gay et al., 2008). Populations consist of 10,000 – 100,000 individuals (Crochet et al., 2003). The *Mus* hybrid zone would take about 60 generations to cross (Fig. 4.5a; Raufaste et al., 2005; Macholán et al., 2007; Wang et al., 2011). The effective population size of

mice is 500 to 5,000 individuals (Pocock, Hauffe, & Searle, 2005). The *Ensatina* hybrid zone would take seven generations to cross (Fig. 4.5b; Staub, Brown, & Wake, 1995; Devitt et al., 2011). In high quality habitat *Ensatina* population density may be 1,300 individuals ha^{-1} with effective population size unknown (Stebbins, 1954; Rosenberg, Noon, Megahan, & Meslow, 1998).

The *Bufo* hybrid zone would take about 60 generations to cross, and dispersal capacity seems comparable to *Mus*. However, the effective dispersal distance in *Mus* is thought to be underestimated by population structure and migration studies, because effective genetic dispersal in mice is inflated by (human mediated) long-distance movements and frequent extinction-recolonization effects (Barton & Hewitt, 1985; Slatkin, 1985; Macholán et al., 2007). Compared to *Bufo*, the effective population sizes in *Larus* and *Mus* are one to three orders of magnitude higher. Toads breed in large water bodies, and adult *Bufo* population sizes regularly consist of 2,500 – 5,000 individuals, but effective population sizes are about two orders of magnitude lower (Scribner, Arntzen, & Burke, 1997). Population sizes in *Bufo* may be more comparable to *Ensatina*, considering that effective population sizes are often much smaller than census population sizes in amphibians (Funk, Tallmon, & Allendorf, 1999; Zeisset & Beebee, 2003; Vences & Wake, 2007). A combination of low dispersal and small population sizes may impede introgression and cause patterns of admixture linkage disequilibrium to be less pronounced in *Bufo* than in the other hybrid zones.

Drift is strong in populations with low dispersal and small population sizes, generating random variation in allele frequency in different populations. This variation causes clines to differ in shape and position more than would be predicted based on the width of clines of individual alleles (cline wobbling; Polechová & Barton, 2011). Cline wobbling increases the width of the hybrid index (or admixture proportion) cline. A low number of markers, sampling a small portion of the variation in cline shapes and position, may therefore be insufficient to assess patterns of admixture linkage disequilibrium. Further increasing the number of individuals sampled and the number of markers employed would increase our ability to test the hypothesis of hybrid zone movement. It is not currently known how sensitive the methods employed here are to the numbers of individuals or markers included in the dataset. Further investigation through simulations could help to improve our understanding of the limitations of these methods. It will be interesting to see if a higher number of markers generated randomly across the genome, such as RAD sequencing data (Baird et al., 2008), will behave similar or different to these gene-coding markers, as gene--coding markers may be more prone to local selection forces in small populations.

Asymmetric pre- or postzygotic isolation, such as displayed in Ensatina, seems most congruent with the co-occurrence of introgression and a peak of admixture linkage disequilibrium seen in *Bufo* (Fig. 4.5b). However, we find only slight evidence for pre- or postzygotic isolation in the *Bufo* hybrid zone. First, we find marginally higher introgression from *B. spinosus* into *B. bufo* than the other way around. Second, hybrids with both parental types of mtDNA are about equally frequent, and the peak of ancestry is symmetric and centred at 0.5 heterozygosity (Fig. S.4), whereas in *Ensatina* almost all hybrid individuals carried mtDNA of only one parental type. Third, low effective selection within the *Bufo* hybrid zone allows the unimpeded exchange of genes and previous studies showed no indication of asymmetric

incompatibility (Arntzen et al., 2016; Trujillo et al., 2017). The only evidence in favour of asymmetric isolation besides the position of the admixture linkage disequilibrium peak, is the relatively high introgression of *B. bufo* mtDNA into *B. spinosus* (e.g. right tail of introgression in mtDNA). Therefore, asymmetric pre- or postzygotic isolation might be weak, with hybridisation inhibited slightly more on the *B. spinosus* side. But this result can also be explained by other processes, such as sex-biased dispersal or adaptive introgression (Currat & Excoffier, 2005; Toews & Brelsford, 2012).

In conclusion, we cannot reject the null hypothesis of a stable *Bufo* hybrid zone in northwest France, and low levels of dispersal and random genetic drift due to small population sizes are important in shaping the patterns of introgression. Based on earlier studies, this hybrid zone was predicted to move southwards (Arntzen et al., 2016, 2017). Our dataset shows similar characteristics to those of previous studies (e.g. northward introgression), but with additional analyses we showed asymmetric reproductive isolation may provide a more important driver of asymmetric introgression than hybrid zone movement. However, multiple processes appear to be at play in shaping the *Bufo* hybrid zone. One might imagine a situation where asymmetric reproductive isolation and hybrid zone movement co-occur where the strongest process may overrule any genetic patterns related to the other process. In addition, small population sizes and other demographic factors such as dispersal distance per generation may cause patterns to become obscured. With 31 nuclear SNPs, we sampled only a tiny portion of the 6 Gbp *B. bufo* genome (Vinogradov, 1998). Further increasing the number of individuals sampled and the number of markers employed can possibly increase our ability to test the hypothesis of hybrid zone movement. It is not currently known how sensitive the methods presented here are to the number of individuals or markers included in the dataset. Further investigation through simulations could help improve our understanding of the limitations of these methods. Looking back on past literature on hybrid zone movement in various organisms, it will be interesting to see if the hypotheses previously inferred still stand with new data and new analytical approaches. The *Bufo* hybrid zone exemplifies the complex influences of interspecific hybridization on genomic composition and can be used to test various hypotheses of introgression and speciation.

Acknowledgements

We thank four anonymous reviewers for their comments, which helped improve the manuscript. We thank Roland Butôt for maintenance of the DNA collection, Wieslaw Babik for advice on the bioinformatics in general, Marta Niedzicka for advice on the pipeline for extraction of exon sequences, Onno Schaap for running the SNP-line, and Graham Derryberry for providing the HZAR scripts. The PhD position of IvR is supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek'. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487.

Data accessibility

Supplemental text contains details of the SNP design:

MEC_Supplemental_text_vanRiemsdijk _2019.pdf
Supplemental tables are available in:
MEC_Supplemental_tables_vanRiemsdijk _2019.xlsx
Supplemental figures are available in:
MEC_Supplemental_figures_vanRiemsdijk _2019.pdf

In- and output of analyses, and custom R scripts are freely available online:
Dryad link (sponsored by MEC): doi:10.5061/dryad.rg18034

Author contributions

IvR, BW, and JWA designed the study and collected data. IvR and RB analysed the data. IvR wrote the manuscript, with input from all co-authors.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. Journal of Evolutionary Biology, 26, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. Evolution, 8, 378–388.

Arnold, M. L. (2004). Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? The Plant Cell, 16, 562–570. https://doi.org/10.1105/tpc.HistPersp

Arntzen, J. W., de Vries, W., Canestrelli, D., & Martínez-Solano, I. (2017). Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Molecular Ecology, 26, 5663–5675. https://doi.org/10.1111/mec.14273

Arntzen, J. W., McAtear, J., Butôt, R., & Martínez-Solano, I. (2018). A common toad hybrid zone that runs from the Atlantic to the Mediterranean. Amphibia-Reptilia, 39(1), 41-50. https://doi.org/10.1163/15685381-00003145

Arntzen, J. W., Recuero, E., Canestrelli, D., & Martínez-Solano, I. (2013). How complex is the *Bufo bufo* species group? Molecular Phylogenetics and Evolution, 69, 1203–1208. https://doi.org/10.1016/j.ympev.2013.07.012

Arntzen, J. W., Trujillo, T., Butot, R., Vrieling, K., Schaap, O. D., Gutiérrez-Rodriquez, J., & Martinez-Solano, I. (2016). Concordant morphological and molecular clines in a contact zone of the common and spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. Frontiers in Zoology, 13, 1–12. https://doi.org/10.1186/s12983-016-0184-7

Arntzen, J. W., & Wallis, G. P. (1991). Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. Evolution, 45, 805–826. https://doi.org/10.2307/2409691

Arntzen, J. W., & Wallis, G. P. (1999). Geographic variation and taxonomy of crested newts (*Triturus cristatus* superspecies): morphological and mitochondrial DNA data. Contributions to Zoology, 68, 181–203.

Arntzen, J. W., Wilkinson, J. W., Butôt, R., & Martínez-Solano, Í. (2014). A new vertebrate species native to the British Isles: *Bufo spinosus* Daudin, 1803 in Jersey. Herpetological Journal, 24, 209–216.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3, 1–7. https://doi.org/10.1371/journal.pone.0003376

Baird, S. J. E. (2015). Exploring linkage disequilibrium. Molecular Ecology Resources, 15, 1017–1019. https://doi.org/10.1111/1755-0998.12424

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution, 68, 2644–2657. https://doi.org/10.1111/evo.12457

Barton, N. H. (1983). Multilocus clines. Evolution, 37, 454–471. https://doi.org/10.2307/2408260

Barton, N. H. (2013). Does hybridization influence speciation? Journal of Evolutionary Biology, 26, 267–269. https://doi.org/10.1111/jeb.12015

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), Hybrid zones and the evolutionary process (pp. 13–45). New York: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual Review of Ecology and Systematics, 16, 113–148.

Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., … Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. Molecular Ecology, 25, 2967–77. https://doi.org/10.1111/mec.13647

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England), 30, 2114–20. https://doi.org/10.1093/bioinformatics/btu170

Bonnet, T., Leblois, R., Rousset, F., & Crochet, P.-A. (2017). A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution, 71, 2140–2158. https://doi.org/10.1111/evo.13296

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. Heredity, 99, 301–312. https://doi.org/10.1038/sj.hdy.6800997

Butlin, R. K., & Smadja, C. M. (2017). Coupling, reinforcement, and speciation. The American Naturalist, 191, 000–000. https://doi.org/10.1086/695136

Canestrelli, D., Porretta, D., Lowe, W. H., Bisconti, R., Carere, C., & Nascetti, G. (2016). The tangled evolutionary legacies of range expansion and hybridization. Trends in Ecology and Evolution. https://doi.org/10.1016/j.tree.2016.06.010

Carneiro, M., Baird, S. J. E., Afonso, S., Ramirez, E., Tarroso, P., Teotõnio, H., … Ferrand, N. (2013). Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. Molecular Ecology, 22, 2511–2525. https://doi.org/10.1111/mec.12272

Crochet, P.-A., Chen, J. Z., Pons, J., Lebreton, J.-D., Hebert, P. D. N., & Bonhomme, F. (2003). Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? Evolution, 57, 2865–2878.

Currat, M., & Excoffier, L. (2005). The effect of the Neolithic expansion on European molecular

diversity. Proceedings of the Royal Society B: Biological Sciences, 272, 679–688. https://doi.org/10.1098/rspb.2004.2999

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. Evolution, 62, 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Daversa, D. R., Muths, E., & Bosch, J. (2012). Terrestrial movement patterns of the common toad (*Bufo bufo*) in Central Spain reveal habitat of conservation importance. Journal of Herpetology, 46, 658–664. https://doi.org/10.1670/11-012

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. Molecular Ecology Resources, 14, 652–663. https://doi.org/10.1111/1755-0998.12209

Devitt, T. J., Baird, S. J. E., & Moritz, C. (2011). Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine--scale genetic analysis of a hybrid zone. BMC Evolutionary Biology, 11. https://doi.org/10.1186/1471-2148-11-245

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics, 16, 551–558. https://doi.org/10.1016/S0168-9525(00)02139-9

Endler, J. A. (1977). Geographic variation, speciation, and clines. (2nd ed.). Princeton, New Jersey: Princeton University Press.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evolutionary Biology, 12, 131. https://doi.org/10.1186/1471-2148-12-131

Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. Molecular Ecology Resources, 17, 27–32. https://doi.org/10.1111/1755-0998.12509

Funk, W. C., Tallmon, D. A., & Allendorf, F. W. (1999). Small effective population size in the long-toed salamander. Molecular Ecology, 8, 1633–1640. https://doi.org/10.1046/j.1365-294X.1999.00748.x

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution, 62, 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. Annual Review of Ecology, Evolution, and Systematics, 48, 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29, 644–52. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity. Nature Protocols, 8, 1494–512. https://doi.org/10.1038/nprot.2013.084

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology, 22, 4606–4618. https://doi.org/10.1111/mec.12415

Hemelaar, A. (1988). Age, growth and other population characteristics of *Bufo bufo* from different latitudes and altitudes. Journal of Herpetology, 22, 369–388.

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. Trends in Ecology and Evolution, 3, 158–167. https://doi.org/10.1016/0169-5347(88)90033-X

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. Genetica, 139, 617–638. https://doi.org/10.1007/s10709-011-9547-3

Hollander, J., Galindo, J., & Butlin, R. K. (2015). Selection on outlier loci and their association with adaptive phenotypes in *Littorina saxatilis* contact zones. Journal of Evolutionary Biology, 28, 328–337. https://doi.org/10.1111/jeb.12564

Kawakami, T., Butlin, R. K., Adams, M., Paull, D. J., & Cooper, S. J. B. (2009). Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (*Vandiemenella, viatica* species group). Evolution, 63, 139–152. https://doi.org/10.1111/j.1558-5646.2008.00526.x

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

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. The American Naturalist, 116, 463–479.

Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. Evolution, 67, 3653–3661. https://doi.org/10.1111/evo.12205

Leaché, A. D., Grummer, J. A., Harris, R. B., & Breckheimer, I. (2017). Evidence for concerted movement of nuclear and mitochondrial clines in a lizard hybrid zone, 38, 42–49. https://doi.org/10.1111/ijlh.12426

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., … Piálek, J. (2007). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution, 61, 746–771. https://doi.org/10.1111/j.1558-5646.2007.00065.x

Mallet, J. (2005). Hybridization as an invasion of the genome. Trends in Ecology and Evolution, 20, 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Barton, N., Lamas, G. M., Santisteban, C. J., Muedas, M. M., & Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics, 124, 921–936.

Mayr, E. (1942). Systematics and the Origin of Species from the viewpoint of a zoologist. (2nd ed.). New York: Columbia University Press.

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52, 263–277. https://doi.org/10.1086/409995

Moran, C. (1981). Genetic demarcation of geographical distribution by hybrid zones. Proceedings of the Ecological Society Australia, 11, 67–73.

Narum, S. R. (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. Conservation Genetics, 7, 783–787. https://doi.org/10.1007/s10592-005-9056-y

Nichols, R. a, & Hewitt, G. M. (1994). The genetic consequences of long-distance dispersal during colonization. Heredity, 72, 312–317. https://doi.org/10.1038/hdy.1994.41

Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M., & Babik, W. (2016). Molecular Inversion

Probes for targeted resequencing in non-model organisms. Scientific Reports, 6, 24051. https://doi.org/10.1038/srep24051
Olmo, E. (1973). Quantitative variations in the nuclear DNA and phylogenesis of the amphibia. Caryologia, 26, 43–68. https://doi.org/10.1080/00087114.1973.10796525
Olmo, E., Gargiulo, G., & Morescalchi, A. (1970). Il contenuto di DNA nucleare di alcuni Anfibi. Bollettina Di Zoologica, 37, 513–514.
Phillips, B. L., Baird, S. J. E., & Moritz, C. (2004). When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. Evolution, 58, 1536. https://doi.org/10.1554/02-498
Pocock, M. J., Hauffe, H. C., & Searle, J. B. (2005). Dispersal in house mice. Biological Journal of the Linnean Society, 84, 565–583. https://doi.org/10.1111/j.1095-8312.2005.00438.x
Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. Genetics, 189, 227–235. https://doi.org/10.1534/genetics.111.129817
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155, 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x
Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C., Baird, S. J. E., … Boursot, P. (2005). Inferences of selection and migration in the Danish house mouse hybrid zone. Biological Journal of the Linnean Society, 84, 593–616. https://doi.org/10.1111/j.1095-8312.2005.00457.x
Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: finding barriers to gene flow. Journal of Evolutionary Biology, in press, 1450–1477. https://doi.org/10.1111/jeb.13047
Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., … Martínez-Solano, I. (2012). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Molecular Phylogenetics and Evolution, 62, 71–86. https://doi.org/10.1016/j.ympev.2011.09.008
Rice, W. R. (1989). Analyzing tables of statistical tests. Evolution, 43, 223–225.
Rice, W. R. (1998). Intergenomic conflict, interlocus antagonistic coevolution and the evolution of reproductive isolation. In D. J. Howard & S. H. Berlocher (Eds.), Endless Forms Species and Speciation (pp. 261–270). Oxford University Press.
Rohwer, S., Bermingham, E., & Wood, C. (2001). Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. Evolution, 55, 405–422. https://doi.org/10.1554/0014-3820(2001)055[0405:PAMDHV]2.0.CO;2
Rosenberg, D. K., Noon, B. R., Megahan, J. W., & Meslow, E. C. (1998). Compensatory behavior of *Ensatina eschscholtzii* in biological corridors: a field experiment. Canadian Journal of Zoology, 76, 117–133. https://doi.org/10.1139/cjz-76-1-117
Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources, 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x
Roy, J. S., O'Connor, D., & Green, D. M. (2012). Oscillation of an anuran hybrid zone: morphological evidence spanning 50 years. PLoS ONE, 7. https://doi.org/10.1371/journal.pone.0052819
Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. Proceedings of the National Academy of Sciences of the United States of America, 201714950. https://doi.org/10.1073/pnas.1714950115
Ryynänen, H. J., & Primmer, C. R. (2006). Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding

amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. BMC Genomics, 7, 192. https://doi.org/10.1186/1471-2164-7-192

Schmickl, R., Marburger, S., Bray, S., & Yant, L. (2017). Hybrids and horizontal transfer: introgression allows adaptive allele discovery. Journal of Experimental Botany, 10.1093/jxb/erx297. https://doi.org/10.1093/jxb/erx297

Scribner, K. T., Arntzen, J. W., & Burke, T. (1997). Effective number of breeding adults in *Bufo bufo* estimated from age-specific variation at minisatellite loci. Molecular Ecology, 6, 701–712. https://doi.org/10.1046/j.1365-294X.1997.00238.x

Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular Ecology, 25, 2559–2576. https://doi.org/http://dx.doi.org/10.1101/043190

Seehausen, O. (2004). Hybridization and adaptive radiation. Trends in Ecology & Evolution, 19, 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding, 33, 1–14. https://doi.org/10.1007/s11032-013-9917-x

Slatkin, M. (1985). Gene flow in natural populations. Annual Review of Ecology and Systematics, 16, 393–430.

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2016). The on-again-off-again relationship between mitochondrial genomes and species boundaries. Molecular Ecology, 26, 2212–2236. https://doi.org/10.1111/mec.13959

Smith, M. A., & Green, D. M. (2005). Dispersal and the metapopulation in amphibian and paradigm ecology are all amphibian conservation: populations metapopulations? Ecography, 28, 110–128. https://doi.org/10.1111/j.0906-7590.2005.04042.x

Staub, N. L., Brown, C. W., & Wake, D. B. (1995). Patterns of growth and movements in a population of *Ensatina eschscholtzii platensis* (Caudata: Plethodontidae) in the Sierra Nevada, California. Journal of Herpetology, 29, 593–599. https://doi.org/10.2307/1564743

Stebbins, R. C. (1954). Natural history of the salamanders of the plethodontid genus *Ensatina.* University of California Publications in Zoology, 54, 47–123.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., … Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43, D447–D452. https://doi.org/10.1093/nar/gku1003

Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution, 45, 237–261.

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. Current Biology, 24, 671–676. https://doi.org/10.1016/j.cub.2014.01.069

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology, 21, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Trochet, A., Moulherat, S., Calvez, O., Stevens, V., Clobert, J., & Schmeller, D. (2014). A database of life-history traits of European amphibians. Biodiversity Data Journal, 2, e4123. https://doi.org/10.3897/BDJ.2.e4123

Trujillo, T., Gutiérrez-Rodríguez, J., Arntzen, J. W., & Martínez-Solano, I. (2017). Morphological and molecular data to describe a hybrid population of the common toad (*Bufo bufo*) and the

spined toad (*Bufo spinosus*) in western France. Contributions to Zoology, 86, 1–10.
Vences, M., & Wake, D. B. (2007). Speciation, species boundaries and phylogeography of amphibians. In H. Heatwole & M. Tyler (Eds.), Amphibian Biology, vol.6, Systematics (pp. 2613–2660). Chipping Norton, Australia: Surrey Beatty & Sons.
Vines, T. H., Dalziel, A. C., Albert, A. Y. K., Veen, T., Schulte, P. M., & Schluter, D. (2016). Cline coupling and uncoupling in a stickleback hybrid zone. Evolution, 70, 1023–1038. https://doi.org/10.1111/evo.12917
Vinogradov, A. E. (1998). Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry, 31, 100–109. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9482279
Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkovú, P., Vyskočilovú, M. M., … Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. Molecular Ecology, 20, 2985–3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x
Wang, S., Sha, Z., Sonstegard, T. S., Liu, H., Xu, P., Somridhivej, B., … Liu, Z. (2008). Quality assessment parameters for EST-derived SNPs from catfish. BMC Genomics, 9, 450. https://doi.org/10.1186/1471-2164-9-450
While, G. M., Michaelides, S., Heathcote, R. J. P., MacGregor, H. E. A., Zajac, N., Beninde, J., … Uller, T. (2015). Sexual selection drives asymmetric introgression in wall lizards. Ecology Letters, 18, 1366–1375. https://doi.org/10.1111/ele.12531
Wickbom, T. (1945). Cytological studies on Dipnoi, Urodela, Anura, and Emys. Hereditas, 31, 241–346. https://doi.org/10.1111/j.1601-5223.1945.tb02756.x
Wielstra, B., Burke, T., Butlin, R. K., & Arntzen, J. W. (2017). A signature of dynamic biogeography: enclaves indicate past species replacement. Proceedings of the Royal Society Biological Sciences, 284, 1–6.
Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., … Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. Evolution Letters, 1–9. https://doi.org/10.1002/evl3.9
Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. Nature Reviews Genetics, 5, 114–122. https://doi.org/10.1038/nrg1269
Zeisset, I., & Beebee, T. J. C. (2003). Population genetics of a successful invader: the marsh frog *Rana ridibunda* in Britain. Molecular Ecology, 12, 639–646. https://doi.org/10.1046/j.1365-294X.2003.01775.x

Chapter 5

Spatial variation in introgression along the common toad hybrid zone

Unpublished working manuscript

I. van Riemsdijk[1,2*], J.W. Arntzen[1], R.K. Butlin[3], G. Bucchiarelli[4,5], E. McCartney-Melstad[4,5], M. Rafajlovic[6], P. Scott[4], E. Toffelmier[4,5], B. Shaffer[4,5], B. Wielstra[1,2,4]

[1]Taxonomy and Systematics, Naturalis Biodiversity Centre, Leiden, NL
[2]Institute of Biology Leiden, Leiden University, Leiden, NL
[3]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK
[4]Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, USA
[5]Institute of the Environment and Sustainability, UCLA, Los Angeles, USA
[6]Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden
*Correspondence: Isolde van Riemsdijk, isolde.vanriemsdijk@naturalis.nl

Abstract

Consistent patterns of gene flow in multiple transects of a hybrid zone provide strong evidence of universal evolutionary processes. The hybrid zone between the recently recognized common toad species, *Bufo bufo* and *B. spinosus*, runs from the Atlantic to the Mediterranean coast across France, and allows independent replication of transects. We test consistency of patterns of introgression and linkage disequilibrium in two transects, positioned in the northwest and southeast of France, with RAD sequencing data. Genomic and geographic clines are used to identify outlier markers, e.g. with relatively restricted introgression (barrier markers). In the northwest transect, high introgression of neutral markers towards the *B. bufo* side of the zone coincides with a northward shift in the peak of admixture linkage disequilibrium (D'), consistent with asymmetric reproductive isolation. In the southeast transect, introgression is symmetric and a peak of D' coincides with the centre of the hybrid zone, consistent with a stable hybrid zone. Barrier markers were twice as abundant in southeast as in northwest France. However, the proportion of markers identified as barrier markers in both transects was high, suggesting a similar barrier to gene flow. These results are best explained by divergence of genetic groups on the *B. bufo* side of the zone, as suggested by a previous mitochondrial DNA phylogeography. The underlying mechanism may be a difference in sex chromosome morphology driving both the differences and similarities between transects, but remains to be tested. We find the common toad hybrid zone is more complex than previously appreciated.

Keywords: asymmetric reproductive isolation; barrier genes; *Bufo bufo*; *Bufo spinosus*; genomic cline; geographic cline; RAD sequencing; replicate transects

Introduction

Barrier genes restrict gene flow between populations. The effect of barrier genes is differential across the genome, and changes the structure of the genome during hybridisation (Abbott et al., 2013; Barton, 2013; Ravinet et al., 2017). The barrier effect is a reduction of effective migration rate of genetic material, relative to the dispersal of individuals between populations (Ravinet et al., 2017). The barrier effect is reinforced when multiple barrier genes become geographically coincident and fixed in a hybrid zone, referred to as cline coupling (Butlin & Smadja, 2017). When barrier genes are coupled, the remainder of the genome tends to become co-distributed, even though parts of the genome are not physically or functionally linked to a barrier gene, ultimately leading to complete reproductive isolation (Barton, 1983; Barton & Gale, 1993; Bierne, Welch, Loire, Bonhomme, & David, 2011; Vines et al., 2016). Barrier genes and coupled genes are expected to show relative steep transitions and tend to be located together in the hybrid zone (Gompert, Parchman, & Buerkle, 2012; Butlin & Smadja, 2017). Consistent patterns of restricted introgression in the same genes across multiple transects are indicative of a universal species boundary (Teeter et al., 2009; Larson, Andrés, Bogdanowicz, & Harrison, 2013; Larson, White, Ross, & Harrison, 2014; Harrison & Larson, 2014).

Asymmetric introgression in hybrid zones, when gene flow is more prominent in one direction than in the other, can be caused amongst others by hybrid zone movement and asymmetric reproductive isolation. Hybrid zone movement occurs when one species has a competitive advantage over the other. Hybrid zone movement can cause unidirectional introgression from the overtaken species into the overtaking species, reflected by an increase of introgression of neutral markers in the wake of the movement (Barton & Hewitt, 1985; Gay, Crochet, Bell, & Lenormand, 2008; Excoffier, Foll, & Petit, 2009; Taylor, Larson, & Harrison, 2015; Arntzen, de Vries, Canestrelli, & Martínez-Solano, 2017; Wielstra, Burke, Butlin, & Arntzen, 2017; Wielstra, Zieliński, & Babik, 2017). Asymmetric pre- or postzygotic isolation occurs when only particular combinations of individuals are successful in reproduction (Hewitt, 1975). Most of the introgression takes place on the side of the hybrid zone where reproduction is the most successful (Haldane, 1922; Barton, 2001; Devitt, Baird, & Moritz, 2011). As hybrid populations receive higher amounts of gene flow from the side of the hybrid zone where reproduction is most successful, asymmetric reproductive isolation can lead to hybrid zone movement (Buggs, 2007).

To be able to assess patterns of introgression, the behaviour of many geographic clines can be summarised using the expected frequency based on the average allele frequency across loci (e.g. hybrid index; Polechová & Barton, 2011). To further assess the direction of gene flow, admixture linkage disequilibrium can be used. The first generation offspring of two diverged species will be heterozygous for all fixed differences, causing full admixture linkage disequilibrium, rather than the usual linkage disequilibrium caused by physical or functional linkage (Barton & Gale, 1993; Baird, 2015). During backcrossing, recombination breaks down admixture linkage disequilibrium and deflates the peak, whereas migration supplies un-admixed material to the hybrid zone and inflates the peak (Barton & Gale, 1993).

When gene flow into the hybrid zone is symmetric, admixture linkage

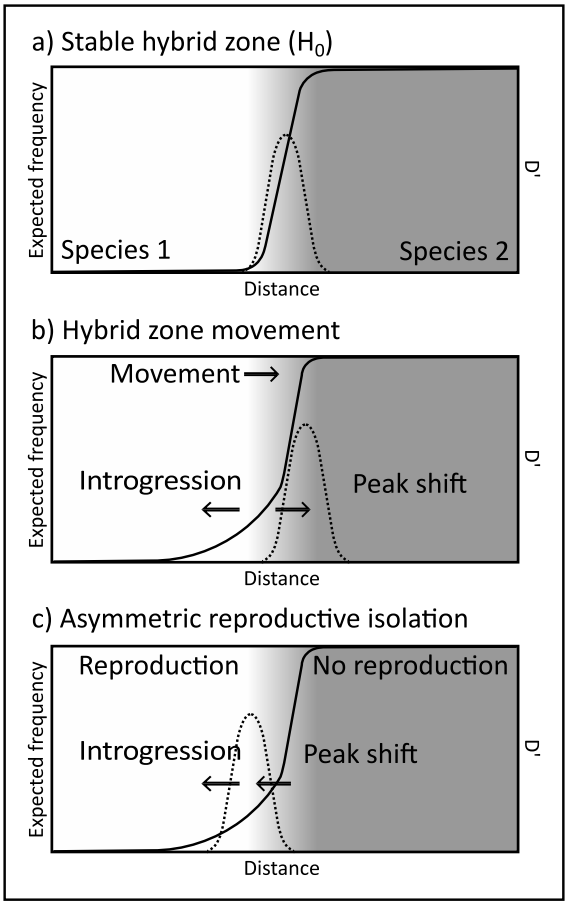


Figure 5.1: Hypotheses of expected allele frequency, which can be represented by the hybrid index (left y-axis, solid line), and a peak in admixture linkage disequilibrium (D', right y-axis, dotted line), against geographic scale (x-axis, distance), under (a) hybrid zone stability (H_0), (b) hybrid zone movement, (c) asymmetric reproductive isolation.

disequilibrium is highest in the centre of the hybrid zone and the peak is shaped following a Gaussian curve (Fig. 5.1a; Gay et al., 2008). The peak is predicted to shift ahead of movement to the side of the hybrid zone where relatively little backcrossing has taken place yet (Fig. 5.1b; Gay et al., 2008; Wang et al., 2011). The peak is predicted to be coincident with the tail of introgression in the case of asymmetric reproductive isolation, because backcrossing allows neutral markers to recombine on both sides of the hybrid zone (Fig. 5.1c; Devitt et al., 2011). What happens to the shape of the peak of admixture linkage disequilibrium during hybrid zone movement or asymmetric reproductive isolation has not received much attention yet (but see Gay et al., 2008). Neither is there a clear hypothesis of what happens to the position of the peak when the hybrid zone moves due to asymmetric reproductive isolation. Sometimes additional evidence may further support hybrid zone movement or asymmetric reproductive isolation, such as a deficit of hybrids with mitochondrial DNA from one of the species supports asymmetric reproductive isolation (Devitt et al., 2011). Again, consistent patterns of such asymmetric introgression across multiple transects are indicative of

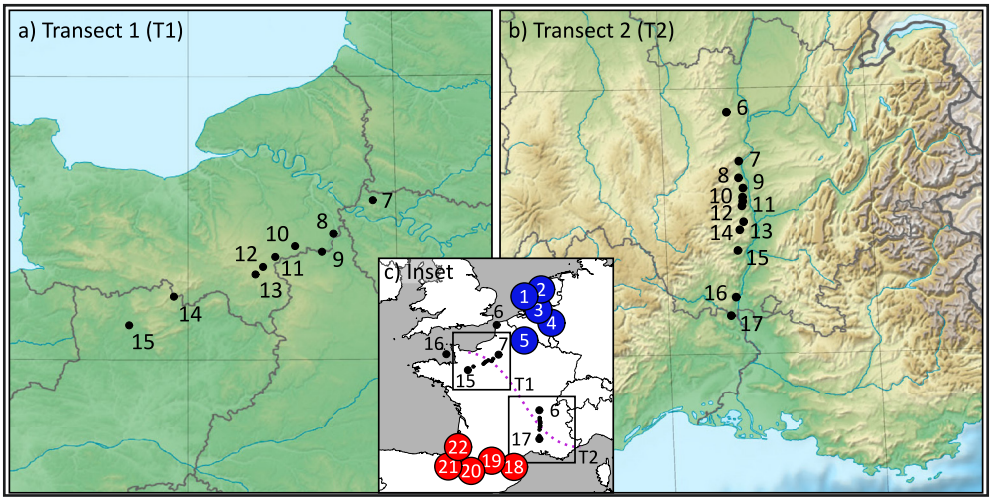


Figure 5.2: Maps of a) transect one in northwest France 1 (T1, population 6-16), b) transect two in southeast France (T2, population 6-17), and c) inset with an overview map. Small black dots are transect populations, large blue dots are reference populations for *Bufo bufo* (population 1-5), and large red dots are reference populations for *B. spinosus* (population 18-22). The purple dashed line indicates the position of the *Bufo* hybrid zone. Base map for panels a, and b was downloaded from https://www.mapsland.com

universal evolutionary processes (Harrison & Larson, 2016).

We study a hybrid zone in France between the common toad, *Bufo bufo* (Linnaeus, 1758) and the spined toad, *B. spinosus* Daudin, 1803, stretching from the coast of the Atlantic Ocean to the coast of the Mediterranean Sea (Fig. 5.2; Arntzen et al., 2018). The length of the hybrid zone allows to test the consistency of patterns of restricted gene flow, asymmetric introgression, and admixture linkage disequilibrium. Previous studies on this hybrid zone reported asymmetric introgression (Arntzen et al., 2017; van Riemsdijk et al., 2019). To infer genome wide patterns of introgression we generated ~ 1,200 diagnostic markers with RAD sequencing. The aim is to test consistency of barrier markers, patterns of asymmetric introgression, and the position of a peak of admixture linkage disequilibrium.

Materials and Methods

3RAD sequencing

DNA extracts of 387 individuals were reused from previous research (Arntzen et al., 2016, 2017, 2018, in prep.). We included five reference populations for *B. bufo*, five reference populations for *B. spinosus*, eleven populations in transect one in northwest France, and twelve populations in transect two in southwest France (Table S.1.; Fig. 5.2). For three pure individuals of each species, libraries were prepared in triplicate (6 individuals x 3 = 18) to assess data error rate, which was estimated to be 0.5 % (Appendix 1). The 3RAD method (Graham et al., 2015; Glenn et al., 2016; Hoffberg et al., 2016) was used to obtain reduced representation genomic libraries. Two restriction enzymes (CLA-I and Sbf-I) were used to cut 50 ng of genomic DNA. A third enzyme (MSP-I) is added to cleave phosphorylated adapter-adapter dimers, avoiding the presence of unwanted adapter or DNA dimers in the library. Internal barcodes

were ligated to the resulting 'sticky ends'. Then external Illumina iTru5 and iTru7 primers, differing by ≥ 3 bp, were ligated to the internal barcodes, in order to trace reads to individuals after pooling (Glenn et al., 2016; Hoffberg et al., 2016).

To 5 µL of DNA extract, 1.5 µL of Cutsmart buffer (New England Biolabs, Ipswich, MA; NEB), 0.5 µL of each digestion enzyme, and 5 µL of purified water were added. Digestion took place for 3h at 37 °C. For internal primer ligation, 1 µL of 100 units/uL DNA ligase, 0.5 µL of ligase buffer (NEB), 1.5 µL of rATP and 2 µL of purified water were added and the samples were heated for two cycles of 20 min on 22 °C and 10 min on 37 °C, and ending with 20 min at 80 °C. The ligated DNA product was purified with diluted SpeedBeads (GE Healthcare Bio-Sciences, Marlborough, MA; see Glenn et al., 2016 and Appendix 2 for details). The external dual index primers were ligated by adding 15 µL of Kapa HiFi HS RM (Kapa Biosystems, Wilmington, MA), 2.5 µl of each iTru5 and iTru7 primer to 5 µl of purified DNA from the previous step (end volume: 25 µL), and a PCR programme of 95 °C for 2 min, and 20 cycles of 98 °C for 20 sec, 60 °C for 15 sec, 72 °C for 30 sec, ending with 72 °C for 5 min, and hold at 15 °C.

After a second purification step, libraries were quantified using the Victor. The DNA concentration of seven libraries contained less than the required concentration after two repeats, thus the total volume was pooled. All other libraries were pooled equimolar. The fragments in the pooled library were size selected for a range of 340-440 bp using a Pippin Prep (Sage Science Inc. Beverly, MA) and quantified using the Victor. The pooled libraries were sequenced on two lanes of an Illumina HiSeq 4000 PE100 at the Vincent J. Coates Genomics Sequencing Laboratory in Berkley, USA.

Data clean-up and assembly

The reads were demultiplexed by the sequencing facility. The average read count per sample was 1.5 million reads, with a maximum of 4.2 million reads and 11 samples had read counts below 0.5 million reads, including the samples which didn't contain the required concentration of DNA at the pooling stage (Table S.2). Twelve samples had low raw read quantities and were excluded (file size <10.000 kb). Cutadapt v. 1.14 (Martin, 2011) was used in three steps to (1) remove 5' primers for each internal barcode combination, (2) remove 3' primers for each internal barcode combination, and (3) remove the Illumina standard adapter and do a quality control. Subsequently, ipyrad v. 0.7.3 (Eaton, 2014) was used to assemble the reads. Settings were: a minimum depth of six at which statistical base calls are made, a maximum of eight heterozygous bases allowed in consensus sequences, and a maximum proportion of shared polymorphic sites in a locus of 0.5 (details in the online supplements). The minimum number of individuals that must have data at a given locus for it to be retained in the final dataset (387 samples, including all triplo samples) varied with 194 (50%), 291 (75%) and 349 (95%) individuals to generate matrixes with different densities for program requirements in downstream analyses. After assembly two more individuals had a very low amount of data present and were excluded from the dataset. The 50% dataset contained 4,863 markers and 39,750 SNPs.

Population structure

First, we assessed the 50% dataset with a single randomly selected SNP per RAD fragment for populations from both transects in a principle component analysis

(PCA), to identify genomic variance in the dataset. We used the package adegenet v. 2.1.1 in R (Jombart, 2008; Jombart & Ahmed, 2011), and based the PCA on allele frequencies whilst replacing missing data with the mean. Population structure was then analysed with Structure v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with the same dataset, but for each transect separately. Preliminary results showed that using a dataset with a different amount of missing data (75% or 90%) gave similar results (not shown). Structure was run with ten replicates of two to ten genetic clusters (K) with a burn in of 10,000 and a chain length of 25,000 under the admixture model using StrAuto (Chhatre & Emerson, 2017). Convergence of the results was checked by investigating log likelihood and parameter stability (Benestan et al., 2016). The results were summarized with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The best value of K was determined with the Evanno method (Evanno, Regnaut, & Goudet, 2005). Structure results were visualised using the R package POPHELPER (Francis, 2017).

Diagnostic SNP selection

Diagnostic SNPs were determined based on genotypes of 37 *B. bufo* and 20 *B. spinosus* individuals from reference populations (Table S.1, custom R script in online supplements). First, we selected all SNPs for which a maximum of 70% of the data was missing in each species to avoid selecting null alleles. Then, all bi-allelic SNPs with only homozygotes of one variant in the reference populations of one species, and vice versa were selected. As sometimes multiple diagnostic SNPs were detected per RAD fragment, we subsequently selected one random SNP per fragment. This resulted in 1,189 diagnostic SNPs.

Hardy-Weinberg and pairwise linkage disequilibrium

We tested for signals of non-random mating success in the dataset containing only diagnostic SNPs, by calculating heterozygote excess and deficit from Hardy-Weinberg equilibrium with the R package 'genepop' based on the program GENEPOP v.1.0.5 (Rousset, 2008). Instead of the conservative Bonferroni test, which accounts for the number of tests performed in total, independence of tests was accounted for within markers (Pc for N=1,189; Rice 1989; Narum 2006). No deviations from Hardy-Weinberg equilibrium with a significant heterozygote excess were present, 14 markers had a heterozygote deficit (Table S.3). These markers were excluded in the HZAR geographic cline fitting analysis and admixture linkage disequilibrium calculations (see below).

BGC outlier detection

To study genome-wide variation of introgression among admixed individuals we used the Bayesian genomic cline model as implemented in the software BGC (Gompert & Buerkle, 2011, 2012; Gompert et al., 2012). The Bayesian genomic cline model is based on the probability that an individual with a certain hybrid index (HI) inherited a gene variant at a locus from one species (φ; in this case *B. bufo*), and the probability of inheriting the other variant from the other species (1 – φ; *B. spinosus*). The probability of *B. bufo* ancestry relative to expected (represented by the HI) is described by cline parameter α. A positive α indicates an increase in the ancestry probability and a negative α indicates a decrease. The cline parameter β measures the average pairwise

linkage disequilibrium based on ancestry for each locus. A positive β indicates an increase in the rate of transition from low to high probability of *B. bufo* ancestry as a function of the HI, or heterozygote disadvantage, whereas a negative β indicates a decrease in the rate of transition, or heterozygote advantage (Gompert & Buerkle, 2011; Parchman et al., 2013). The input files for parental genotypes included only individuals from the reference populations, and the input file for admixed genotypes included individuals with an average admixture proportion (Structure Q score) between 0.05 and 0.95, treated as a single admixed population (Fig. S.1). A single MCMC chain was run for 75,000 steps and samples were taken from the posterior distribution every fifth step following a burn-in of 25,000 steps. Convergence was assessed (Fig. S.2), and we tested for outlier loci using 'estpost' (Gompert & Buerkle, 2011).

Geographic cline analysis

Geographic classic equilibrium cline models were fitted using the R package 'HZAR' (Derryberry, Derryberry, Maley, & Brumfield, 2014) for all diagnostic RAD SNPs. For transect one, populations 6 and 16, which are distant from the 'real' transect, were removed from the dataset. To determine distance between the populations, a custom R script was used, and the direction of the transects was used conform previous publications (Arntzen et al., 2016, 2017). The shape and position of many clines can be summarised in the expected cline (Polechová & Barton, 2011), e.g. the HI cline (Fitzpatrick, 2012). We calculated the HI of all non-outlier markers, and the HI of all heterozygote disadvantage outliers (β > 0) as determined by BGC, and fitted a cline. Thirty maximum likelihood estimation searches were performed with random starting parameters, followed by a trace analysis of 60,000 generations on all models with a delta Akaike Information Criterion corrected for small-sample--size (dAICc) < 10. Fifteen model variants based on all possible combinations of trait interval (allele frequency at the ends of the transects; three types) and tail shape (five types). Convergence was visually assessed in trace plots (online supplements). We used a custom R script to estimate the surface underneath the cline tail towards the left (Q_{left}) and the right (Q_{right}) side of the hybrid zone centre until the equation reached an allele frequency of 0.05 or 0.95 (online supplements, Fig. S.3).

Admixture linkage disequilibrium

Admixture linkage disequilibrium (D') based on the variance in hybrid index, lifetime dispersal distance weighted for pre- and post- metamorphosis (σ) and average effective selection on a locus (s*) were calculated following Barton & Gale (1993) using scripts from van Riemsdijk et al. (2019). Average effective selection on a locus (s*) is the selection pressure on a locus at the zone centre due to association with other loci. The data contained only markers in Hardy-Weinberg equilibrium and markers not indicated as outlier in BGC (832 and 652 loci) for each transect (Table 5.1). The analysis was repeated using only the markers which were found to be heterozygote disadvantage outliers (β > 0; 56 and 121 loci) for each transect. Fixed parameters were: a recombination rate of 0.4997 was calculated following formula (6) from Macholán et al. (2007), using the number of chiasmata per bivalent for *B. bufo* (1.95; Wickbom, 1945), and the number of chromosomes for *B. bufo* (N = 22), a

Table 5.1: Bayesian genomic cline (BGC) results comparison of significant outliers for transect one (T1) and transect two (T2), and the markers which were outliers in both transects (overlap), where significance of outliers is based on the exclusion of 0 in the 99.9% confidence interval (CI). The total number of markers analysed was 1,189.

Outlier	Biological interpretation	T1	T2	Overlap
$\alpha > 0$	Directional introgression from *B. bufo* into *B. spinosus*	110*	174⁺	49
$\alpha < 0$	Directional introgression from *B. spinosus* into *B. bufo*	151*	185⁺	92
$\beta > 0$	Heterozygote disadvantage	56	123	26
$\beta < 0$	Heterozygote advantage	50	105	11

* Significant Chi-squared with 6.4406, df = 1, P = 0.0112
⁺ Not significant Chi-squared with 0.3371, df = 1, P = 0.5615

generation time of 2.5 years for *Bufo* toads at the latitude of the hybrid zone (mean of 3 years in females and 2 years in males; Hemelaar, 1988) and initial secondary contact 8,000 years ago following Arntzen et al. (2016). The width of the hybrid zone was derived from a general sigmoid cline model following HZAR (Derryberry et al., 2014), fitted to the HI. Mean and 95% confidence interval (CI) were based on 1,000 bootstrap replicates of the original genotype dataset (with replacement, maintaining original population size). A Gaussian curve was fitted through the estimates of D' following Gay et al. (2008), and 95% confidence intervals (CI) for the calculated parameters were derived from the bootstrap data.

Results

The first axis (x-axis) of the PCA appears to reflect genetic variance over geographical latitude, ranging from pure *B. bufo* in the north (right side of the PCA plot), to pure *B. spinosus* in the south (left side of the PCA plot; Fig. S.4). The second axis (y-axis) appears to reflect the genetic variance between both transects. The best number of genetic clusters for transect one is K=3, and for transect two K=2 (Fig. S.1). Two genetic clusters correspond to the expectation of two species. A third genetic group identifies the hybrid population in both transects.

The number of individuals representing different hybrid indices in the BGC analysis is well-spread (Fig. S.5). Transect one has a significant higher number of markers with a decrease in the probability of *B. bufo* ancestry relative to the hybrid index ($\alpha < 0$, n = 151), compared to the number of markers with an increase in the probability of *B. bufo* ancestry relative to hybrid index ($\alpha > 0$, n = 110, $\chi 2$ test P = 0.0112; Table 5.1, Fig. S.5). Transect two has a close to equal number of markers with raised and lowered probability of *B. bufo* ancestry ($\alpha > 0$, n = 174; $\alpha < 0$, n = 185, $\chi 2$ test P = 0.5615). The number of markers with an outlier β in transect one ($\beta > 0$, n = 56 and $\beta < 0$, n = 42) is about half the number in transect two ($\beta > 0$, n= 123, and $\beta < 0$, n = 105). About 20 to 40 % of RAD markers which are outlier in transect one, are also outlier in transect (last column Table 5.1). Such overlap would not be likely when the two transects were evolutionarily independent. For example, under random sampling with replacement, the chance of selecting the same 26 markers in both transects to be a barrier marker is close to zero ($2.948e^{-11}$, calculated using the R function dbinom(), with y=26, n=56, and p=132/1,189).

We verify that outlier markers for α or β in the BGC analysis reflect outlier behaviour in the shape and position of the geographic cline. For example, markers which are outlier of heterozygote disadvantage ($\beta > 0$) in the Bayesian genomic

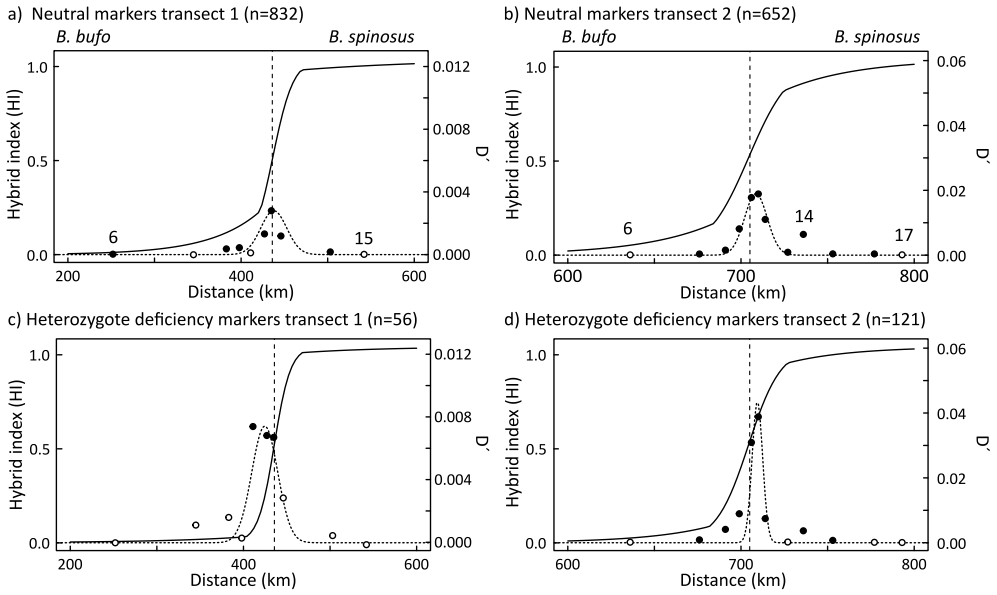


Figure 5.3: Geographic clines for the hybrid index (HI) and admixture linkage disequilbirium peaks (D') for (a) neutral markers according to the genomic clines analysis for transect one and (b) transect two, and (c) for heterozygote disadvantage markers (β > 0) for transect one and (d) for transect two. The x-axis shows distance along the transect. The y-axis on the left shows the HI (solid line), and the y-axis on the right shows D' (dotted line and dots). Note that for transect one, the x-axis is twice as long as for transect two, and the right y-axis for transect one is five times smaller than for transect two. Solid dots show D' significantly different from zero, whilst open dots are not significantly different from zero, based on 95% confidence intervals.

clines analysis also show steep and coincident geographic clines (Fig. S.6). From here onwards, an outlier marker of heterozygote disadvantage is termed a 'barrier marker', and a marker which is not an outlier is a 'neutral marker'. The hybrid index cline based on neutral markers in transect one shows a high level of introgression from *B. spinosus* into *B. bufo* (Q_{left} = 15.0, Q_{right} = 8.8), and shows symmetric introgression in transect two (Q_{left} = 11.0, Q_{right} = 11.0; Fig. 5.3, Table S.4, Fig. S.7). The cline widths for neutral markers in both transects are similar (47 km and 49 km), and the cline shape is generally steep (stepped clines). The hybrid index clines for barrier markers are symmetrical in both transects.

In both transects, admixture linkage disequilibrium (D') for neutral markers shows peaks in the centre of the hybrid zone. The peak in transect one is about five times lower than the peak in transect two (Fig. 5.3). In transect one, barrier markers show a peak of D' on the *B. bufo* side of the hybrid zone, and barrier markers show a peak of D' in the centre of transect two. Population 14 in transect two has a relative high *B. bufo* genetic contribution compared to neighbouring populations in Structure, and shows high D' for both neutral and barrier markers. The lower peaks in transect one underlie the lower estimates of effective selection against hybrids (σ) and lifetime dispersal compared to transect two. The effective selection based on neutral markers (s*) for transect one is 0.0022 (95% CI 0.0012-0.0034) and for transect two 0.0195 (95% CI 0.0143-0.0252; Table S.5). The effective selection based on barrier markers (s*) for transect one is 0.0101 (95% CI 0.0054-0.0152) and for transect two 0.0344

(95% CI 0.0220-0.0470). The lifetime dispersal distance for transect one and two is estimated at 1.8 and 4.0 km per generation, respectively. These results are similar to the previously published values based on 32 diagnostic SNPs (van Riemsdijk, Butlin, Wielstra, & Arntzen, 2019).

Discussion

We studied a hybrid zone between two *Bufo* species to test consistency of barrier markers, patterns of asymmetric introgression, and the position of a peak of admixture linkage disequilibrium (D'). Bayesian genomic cline (BGC) outliers of parameter α indicate directional introgression, and outliers of β indicate heterozygote disadvantage or advantage. Outlier behaviour was verified by assessing the geographic cline shapes, and the terms presented in Table 5.1 are used to identify outlier types. Markers which were not an outlier in the BGC analysis are referred to as 'neutral markers', and heterozygote disadvantage outliers are referred to as 'barrier markers'. These identifications may be false in individual cases, but we assume the larger portion of outliers are indeed linked to markers under selection.

In transect one, directional introgression of markers ($\alpha < 0$) is higher from *B. spinosus* into *B. bufo* than the other way around ($\alpha > 0$). This supports positive selection on a part of the *B. spinosus* genome to replace *B. bufo* variants, or negative selection on *B. bufo* variants to be replaced by *B. spinosus* variants. Neutral markers show unidirectional introgression towards the north, with a peak of admixture linkage disequilibrium (D') at the centre of the hybrid zone. Barrier markers ($\beta > 0$) have a symmetric cline, but the peak of D' is situated north of the hybrid zone. The asymmetries towards the north of the hybrid zone are in agreement with previous results (Arntzen et al., 2017, van Riemsdijk et al., 2019). In transect two, selective introgression of markers is equal on both sides of the hybrid zone. For both neutral and barrier makers the geographic clines are symmetric and the peaks of D' are situated in the centre. The only aspect of asymmetry is one population with an unexpected high frequency of *B. bufo* alleles and two populations with unexpected high D' south of the hybrid zone. A recent dispersal event of *B. bufo* may have occurred near population 14. The results for transect one are in support of asymmetric reproductive isolation, whilst the results for transect two are in support of the null hypothesis of a stable hybrid zone.

The two transects thus do not show the same patterns of asymmetric introgression and linkage disequilibrium. The pattern we find in transect one suggests asymmetric reproductive isolation, but does not exclude the possibility that the hybrid zone is also moving (Buggs, 2007). Transect two shows no signs of asymmetry, even though on a longer time scale it is thought to have moved a considerable distance (Arntzen et al., 2017). How long such movements leave traces in neutral or barrier markers depends on the effective selection against hybrids, but may vary depending on other factors related to hybrid zone movement as well (Currat, Ruedi, Petit, & Excoffier, 2008). The distance a peak of D' can be displaced from the hybrid zone centre is by definition restrained to the area in which introgression occurs. This distance is rather small if the hybrid zone is represented by a steep (stepped) cline, which is the case here. In both transects, the shape of the peak of D' is not well described by a Gaussian curve. A simulation study is required to refine the expectations of the peak of D', or other

related measures (Gay et al., 2008; Wang et al., 2011), under different hybrid zone movement scenarios.

Another difference between both transects is that the number of barrier markers in transect one is about half the number in transect two. However, a number of markers are found to be outliers in both transects by the Bayesian genomic cline analysis, which is unlikely to be the result of random resampling of the same markers (Table 5.1). Markers which show the same barrier behaviour across multiple transects of the hybrid zone indicate a universal species boundary (Harrison & Larson, 2014). We are thus looking for an explanation that unifies asymmetry and symmetry of introgression and the difference in the number of barrier markers in both transects, and explains the overlap of outlier markers in both transects.

Divergence of genetic groups involved in both sections of this *Bufo* hybrid zone could explain these differences. The *B. bufo* in northwest France are part of only one mitochondrial lineage originating from a refugium from the last glacial maximum in the Balkans, whilst the *B. bufo* in southeast France are, in addition, part of two lineages originating from refugia in Italy and the northern Balkans (Garcia-Porta et al., 2012; Recuero et al., 2012; Arntzen et al., 2017). Individuals of *B. spinosus* from France belong to a lineage from an Iberian refugium (Garcia-Porta et al., 2012; Recuero et al., 2012; Arntzen et al., 2017). The reason why additional genetic *B. bufo* clusters in southeast France are not detected in Structure, may be due to the absence of un--admixed individuals of the less widespread lineages in our dataset. An additional genetic cluster for which no un-admixed individuals are included, a "ghost lineage", may remain undetected in population demographic analyses (Rogers & Bohlender, 2015). We assume deep mitochondrial substructure related to geography within *B. bufo* is reflected by nuclear genetic substructure.

Genetic subgroups of *B. bufo* do not yet explain the differences and similarities between the transects. The underlying mechanism may be a difference in (sex) chromosome morphology between the geographic *B. bufo* groups. The *B. bufo* in northwest France are presumably homomorphic for both sexes for all chromosomes, whilst *B. bufo* females in an Italian mitochondrial subclade are heteromorphic for chromosome 6 (Morescalchi, 1964; Birstein & Mazin, 1982; Pisanets et al., 2009; Skorinov et al., 2018). Heteromorphic chromosome 6 in female *B. bufo* are in contrast to the situation in *B. spinosus* from (amongst others) Spain, where males are heteromorphic for chromosome 6 (Skorinov et al., 2018). In northwest France, *B. bufo* with homomorphic chromosome 6 in both sexes are meeting with *B. spinosus* with heteromorphic chromosome 6 in females. In general, taxa with homomorphic chromosomes show a lower level of reproductive isolation than taxa with heteromorphic chromosomes (Lima, 2014). This would explain the increase of introgression on the side of *B. bufo* in northwest France. In southeast France, *B. bufo* with heteromorphic chromosome 6 in males are meeting *B. spinosus* with heteromorphic chromosome 6 in females. A switch of heteromorphic chromosomes from male to female between the two taxa, as could be the case in southeast France, may result in a symmetrical fitness effect depending on dominance of the sex determining genes (Nishioka & Hanada, 1994).

Switches of heteromorphic chromosomes from males to females or vice versa have also been found within other Anuran genera (Hillis & Green, 1990; Uno et al.,

2008; Stöck et al., 2011; Jeffries et al., 2018). Instead of the usual decay which leads to heteromorphic sex chromosomes, homomorphic sex chromosomes in anuran species are maintained by either or both recombination between sex chromosomes and sex chromosome turnover (Perrin, 2009; Stöck et al., 2013; Dufresnes et al., 2015; Rodrigues, Studer, Dufresnes, & Perrin, 2018). If different chromosome morphologies are present within *B. bufo*, this could explain the asymmetry and symmetry in the different transects, and the different number of barrier markers between transects. If indeed a single chromosome is involved, it could also explain the overlap of barrier markers between both transects, as they may be linked to this chromosome. A genetic difference between both transects is thus likely to be found within *B. bufo*, and less likely in *B. spinosus*, but remains to be tested. First, the presence of intraspecific *B. bufo* groups can be studied by using sequence capture or RAD sequencing in multiple sections of the hybrid zone and the wider distribution of *B. bufo*. Laboratory crosses of individuals from the resulting intraspecific *B. bufo* groups and *B. spinosus* should further verify the modes of reproductive isolation.

The differences between both transects could also (partially) be attributed to a difference in the environment. Especially along a long hybrid zone, such as is the case for *Bufo*, the environment will differ. The landscape near transect one is rather homogeneous, whilst the landscape near transect two is more structured (Fig. 5.2). In northern France a tail of introgression to the north could alternatively be interpreted as a tail of introgression in the wake of hybrid zone movement. The occurrence of *B. spinosus* is linked to higher elevation in this section of the zone (Arntzen et al., 2016). Tension zones tend to become trapped at such ecological transitions (Endler, 1977; Barton & Hewitt, 1985). In southeast France the zone appears to be stable. In such a patchy landscape, barrier genes may be coupled to a steep gradient of locally adaptive genes and stabilise the hybrid zone, even when local adaptive selection is moderate (Bierne et al., 2011). The hybrid zone in northwest France may thus be less restricted in terms of introgression because the environment is more homogeneous, whilst the hybrid zone in southeast France may be trapped at a gradient of locally adaptive markers in a heterogeneous environment. Further testing of a relation between elevation and additional environmental gradients could be based on more transects with more similar elevation in different parts of the hybrid zone.

We have highlighted differences and similarities between two transects of the hybrid zone of *B. bufo* and *B. spinosus* using ~1200 nuclear markers. Genomic cline analyses enabled the identification of different outlier groups. Patterns of introgression and admixture linkage disequilibrium suggest asymmetric reproductive isolation in northwest France and a stable hybrid zone in southeast France. We suggest several research avenues to study the geographical variation in the French *Bufo* hybrid zone.

Acknowledgements

We thank Frido Welker, Tara Luckau, the members of the Butlin laboratory at the University of Sheffield, and the Allentoft laboratory at the Natural History Museum in Copenhagen for support and discussion. The PhD position of IvR is supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant

agreement No. 655487. Part of this project was carried out at the Shaffer laboratory in Los Angeles, at the University of California. This study trip has been sponsored by the Leiden University Fund / Swaantje Mondt Fonds (D7102).

Data accessibility

In- and output of analyses, and custom R scripts will be made freely available online.

Author contributions

IvR, JWA, BS, and BW designed the study. IvR and JWA collected samples. IvR performed the laboratory work and data assembly with contributions from GB, ET, PS, ET and MR. IvR analysed the data and wrote the manuscript with input from all authors.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. Journal of Evolutionary Biology, 26, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Arntzen, J. W., de Vries, W., Canestrelli, D., & Martínez-Solano, I. (2017). Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Molecular Ecology, 26, 5663–5675. https://doi.org/10.1111/mec.14273

Arntzen, J. W., McAtear, J., Butôt, R., & Martínez-Solano, I. (2018). A common toad hybrid zone that runs from the Atlantic to the Mediterranean. Amphibia-Reptilia, 39, 41–50. https://doi.org/10.1163/15685381-00003145

Arntzen, J. W., Trujillo, T., Butot, R., Vrieling, K., Schaap, O. D., Gutiérrez-Rodriquez, J., & Martinez-Solano, I. (2016). Concordant morphological and molecular clines in a contact zone of the common and spined toad (*Bufo bufo* and *B. spinosus*) in the northwest of France. Frontiers in Zoology, 13, 1–12. https://doi.org/10.1186/s12983-016-0184-7

Baird, S. J. E. (2015). Exploring linkage disequilibrium. Molecular Ecology Resources, 15, 1017–1019. https://doi.org/10.1111/1755-0998.12424

Barton, N. H. (1983). Multilocus clines. Evolution, 37, 454–471. https://doi.org/10.2307/2408260

Barton, N. H. (2001). The role of hybridization in evolution. Molecular Ecology, 10, 551–568. https://doi.org/10.1046/j.1365-294X.2001.01216.x

Barton, N. H. (2013). Does hybridization influence speciation? Journal of Evolutionary Biology, 26, 267–269. https://doi.org/10.1111/jeb.12015

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), Hybrid zones and the evolutionary process (pp. 13–45). New York: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual Review of Ecology and Systematics, 16, 113–148.

Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., … Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. Molecular Ecology, 25, 2967–77. https://doi.org/10.1111/mec.13647

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072.

https://doi.org/10.1111/j.1365-294X.2011.05080.x
Birstein, V. J., & Mazin, A. L. (1982). Chromosomal polymorphism of *Bufo bufo*: karyotype and C-banding pattern of *B. b. verrucosissima*. Genetica, 59, 93–98. https://doi.org/10.1007/BF00133292
Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. Heredity, 99, 301–312. https://doi.org/10.1038/sj.hdy.6800997
Butlin, R. K., & Smadja, C. M. (2017). Coupling, reinforcement, and speciation. The American Naturalist, 191, 000–000. https://doi.org/10.1086/695136
Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: automation and parallelization of STRUCTURE analysis. BMC Bioinformatics, 18, 1–5. https://doi.org/10.1186/s12859-017-1593-0
Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. Evolution, 62, 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x
Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. Molecular Ecology Resources, 14, 652–663. https://doi.org/10.1111/1755-0998.12209
Devitt, T. J., Baird, S. J. E., & Moritz, C. (2011). Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine--scale genetic analysis of a hybrid zone. BMC Evolutionary Biology, 11. https://doi.org/10.1186/1471-2148-11-245
Dufresnes, C., Borzee, A., Horn, A., Stock, M., Ostini, M., Sermier, R., … Perrin, N. (2015). Sex--chromosome homomorphy in palearctic tree frogs results from both turnovers and X-Y recombination. Molecular Biology and Evolution, 32, 2328–2337. https://doi.org/10.1093/molbev/msv113
Eaton, D. A. R. (2014). PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics, 30, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121
Endler, J. A. (1977). Geographic variation, speciation, and clines. (R. M. May, Ed.) (2nd ed.). Princeton, New Jersey: Princeton University Press.
Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414
Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evolutionary Biology, 12, 131. https://doi.org/10.1186/1471-2148-12-131
Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. Molecular Ecology Resources, 17, 27–32. https://doi.org/10.1111/1755-0998.12509
Garcia-Porta, J., Litvinchuk, S. N., Crochet, P. A., Romano, A., Geniez, P. H., Lo-Valvo, M., … Carranza, S. (2012). Molecular phylogenetics and historical biogeography of the west--palearctic common toads (*Bufo bufo* species complex). Molecular Phylogenetics and Evolution, 63, 113–130. https://doi.org/10.1016/j.ympev.2011.12.019
Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution, 62, 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x
Glenn, T. C., Nilsen, R. A., Kieran, T. J., Finger Jr., J. W., Pierson, T. W., Bentley, K. E., … Faircloth,

B. C. (2016). Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). BioRxiv.
Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. Molecular Ecology, 20, 2111–2127. https://doi.org/10.1111/j.1365-294X.2011.05074.x
Gompert, Z., & Buerkle, C. A. (2012). bgc: software for Bayesian estimation of genomic clines. Molecular Ecology Resources, 12, 1168–1176. https://doi.org/10.1111/1755-0998.12009.x
Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 439–450. https://doi.org/10.1098/rstb.2011.0196
Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., … Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Molecular Ecology Resources, 15, 1304–15. https://doi.org/10.1111/1755-0998.12404
Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics, 12, 7–109. https://doi.org/10.1007/BF02983075
Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. Journal of Heredity, 105, 795–809. https://doi.org/10.1093/jhered/esu033
Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology, 25, 2454–2466. https://doi.org/10.1111/mec.13582
Hemelaar, A. (1988). Age, growth and other population characteristics of *Bufo bufo* from different latitudes and altitudes. Journal of Herpetology, 22, 369–388.
Hewitt, G. M. (1975). A sex-chromosome hybrid zone in the grasshopper *Podisma pedestris* (Orthoptera: Acrididae). Heredity, 35, 375–387. https://doi.org/10.1038/hdy.1975.108
Hillis, D. M., & Green, D. M. (1990). Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. Journal of Evolutionary Biology, 3, 49–64. https://doi.org/10.1046/j.1420-9101.1990.3010049.x
Hoffberg, S., Kieran, T., Catchen, J., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources, 16, 1264–1278. https://doi.org/10.1111/jnc.13494
Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., … Perrin, N. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature Communications, 9, 4088. https://doi.org/10.1038/s41467-018-06517-2
Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129
Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521
Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources, 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387
Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. Evolution, 67, 3653–3661. https://doi.org/10.1111/evo.12205
Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. Molecular Ecology, 23, 1668–1678. https://doi.org/10.1111/

mec.12601
Lima, T. G. (2014). Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. Nature Communications, 5, 1–8. https://doi.org/10.1038/ncomms5743
Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., … Piálek, J. (2007). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution, 61, 746–771. https://doi.org/10.1111/j.1558-5646.2007.00065.x
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17, 10. https://doi.org/10.14806/ej.17.1.200
Morescalchi, A. (1964). Il corredo cromosomico dei Bufonidi Italiani. Bolletino Di Zoologia, 31, 827–836. https://doi.org/10.1080/11250006409441116
Narum, S. R. (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. Conservation Genetics, 7, 783–787. https://doi.org/10.1007/s10592-005-9056-y
Nishioka, M., & Hanada, H. (1994). Sex of reciprocal hybrids between the Hamakita (XX-XY type) population and the Murakami (ZW-ZZ type) population of *Rana rugosa*. Sci. Rep. Lab. Amphibian Biol., Hiroshima Univ., 13, 35–50.
Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., … Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology, 22, 3304–3317. https://doi.org/10.1111/mec.12201
Perrin, N. (2009). Sex reversal: a fountain of youth for sex chromosomes? Evolution, 63, 3043–3049. https://doi.org/10.1111/j.1558-5646.2009.00837.x
Pisanets, E. M., Litvinchuk, S. N., Rosanov, J. M., Reminniy, V. Y., Pasynkova, R. A., & Suryadnaya, N. N. M. A. S. (2009). Common toads (Amphibia, Bufonidae, *Bufo bufo* complex) from the Ciscaucasia and north of the Caucasus: the new analysis of the problem. Zbirnik Prats' Zool. Mus., 40, 83–125. [in Russian]
Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. Genetics, 189, 227–235. https://doi.org/10.1534/genetics.111.129817
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155, 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x
Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: finding barriers to gene flow. Journal of Evolutionary Biology, 30, 1450–1477. https://doi.org/10.1111/jeb.13047
Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., … Martínez-Solano, I. (2012). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Molecular Phylogenetics and Evolution, 62, 71–86. https://doi.org/10.1016/j.ympev.2011.09.008
Rice, W. R. (1989). Analyzing tables of statistical tests. Evolution, 43, 223–225.
Rodrigues, N., Studer, T., Dufresnes, C., & Perrin, N. (2018). Sex-chromosome recombination in common frogs brings water to the fountain-of-youth. Molecular Biology and Evolution, 35, 942–948. https://doi.org/10.1093/molbev/msy008
Rogers, A. R., & Bohlender, R. J. (2015). Bias in estimators of archaic admixture. Theoretical Population Biology, 100, 63–78. https://doi.org/10.1016/j.tpb.2014.12.006
Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources, 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Skorinov, D. V, Bolshakova, D. S., Donaire, D., Pasynkova, R. A., Litvinchuk, S. N., & Litvinchuk, B. (2018). Karyotypic analysis of the spined toad, *Bufo spinosus* Daudin, 1803 (Aphibia: Bufonidae). Russian Journal of Herpetology, 25, 253–258. https://doi.org/10.30906/1026-2296-2018-25-4-253-258

Stöck, M., Croll, D., Dumas, Z., Biollay, S., Wang, J., & Perrin, N. (2011). A cryptic heterogametic transition revealed by sex-linked DNA markers in Palearctic green toads. Journal of Evolutionary Biology, 24, 1064–1070. https://doi.org/10.1111/j.1420-9101.2011.02239.x

Stöck, M., Savary, R., Betto-Colliard, C., Biollay, S., Jourdan-Pineau, H., & Perrin, N. (2013). Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). Journal of Evolutionary Biology, 26, 674–682. https://doi.org/10.1111/jeb.12086

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: windows on climate change. Trends in Ecology and Evolution, 30, 398–406. https://doi.org/10.1016/j.tree.2015.04.010

Teeter, K., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., C.Nachman, M. W., & Tucker, P. K. (2009). The variable genomic architecture of isolation between hybridizing species of house mice. Evolution, 64, 472–485. https://doi.org/10.1111/j.1558-5646.2009.00846.x

Uno, Y., Nishida, C., Oshima, Y., Yokoyama, S., Miura, I., Matsuda, Y., & Nakamura, M. (2008). Comparative chromosome mapping of sex-linked genes and identification of sex chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems. Chromosome Research, 16, 637–647. https://doi.org/10.1007/s10577-008-1217-7

van Riemsdijk, I., Butlin, R. K., Wielstra, B., & Arntzen, J. W. (2019). Testing an hypothesis of hybrid zone movement for toads in France. Molecular Ecology, in press, 0–3. https://doi.org/10.1111/mec.15005

Vines, T. H., Dalziel, A. C., Albert, A. Y. K., Veen, T., Schulte, P. M., & Schluter, D. (2016). Cline coupling and uncoupling in a stickleback hybrid zone. Evolution, 70, 1023–1038. https://doi.org/10.1111/evo.12917

Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkovú, P., Vyskočilovú, M. M., … Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. Molecular Ecology, 20, 2985–3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x

Wickbom, T. (1945). Cytological studies on Dipnoi, Urodela, Anura, and *Emys*. Hereditas, 31, 241–346. https://doi.org/10.1111/j.1601-5223.1945.tb02756.x

Wielstra, B., Burke, T., Butlin, R. K., & Arntzen, J. W. (2017). A signature of dynamic biogeography: enclaves indicate past species replacement. Proceedings of the Royal Society Biological Sciences, 284, 1–6.

Wielstra, B., Zieliński, P., & Babik, W. (2017). The Carpathians hosted extra-Mediterranean refugia-within-refugia during the Pleistocene Ice Age: genomic evidence from two newt genera. Biological Journal of the Linnean Society, 122, 605–613.

Discussion

This thesis focusses on hybrid zone dynamics. I will highlight important results that apply specifically to the genera studied. Using the effective selection against hybrids to compare the barrier effect between taxa is a promising new lead. The shape of a peak of D' may provide a further avenue to investigate the underlying mechanisms of asymmetric introgression in dynamic hybrid zones. Future perspectives of research on dynamic hybrid zones are discussed.

Specific findings

In Chapter 1 I show that the banded newts (*Ommatotriton*) represents three distinct species, rather than two as in previous treatments. A phylogenetic tree based on mitochondrial DNA reveals high divergence between the species and they do not share nuclear DNA genotypes. As the morphology of *O. nesterovi* and *O. ophryticus* is highly similar, they were considered cryptic species. However, building on the information from Chapter 1, a recent publication supports the species status of *O. nesterovi* and *O. ophryticus* by subtle divergence in phenotype (Uzum et al., 2019).

In Chapter 2, hybrid offspring of *O. nesterovi* and *O. ophryticus* were found in Spain. All Spanish individuals possessed mitochondrial haplotypes of *O. ophryticus*, which was in line with the identification of one introduced Spanish individual as *O. ophryticus* based on the number of rib-bearing vertebrae (Fontelles, Guixé, Martínez--Silvestre, Soler, & Villero, 2011). However, the nuclear markers showed all individuals contain both *O. nesterovi* and *O. ophryticus* haplotypes. Hybrid offspring is able to successfully reproduce as eleven out of twelve individuals from Spain were not first generation offspring. The geographic origin of the *O. ophryticus* mitochondrial haplotypes, which occur 400 km eastward from the species boundary. This implies that the hybridisation has not happened in a natural population, but must have taken place during or after captivity.

Finding hybrid offspring was surprising, because we found no admixture in the natural range of *Ommatotriton* (Chapter 1). We could narrow down the area of an expected hybrid zone to a 60 km gap (Fig. 6.1a). The possibility remains that low levels of introgression near this gap remained undetected, as only two nuclear markers were used. Two other newt genera with distributions similar to that of *Ommatotriton,* however, have shown considerable movement of up to 600 km (Nadachowska & Babik, 2009; Wielstra et al., 2017). If there is a hybrid zone present between *O. nesterovi* and *O. ophryticus*, selection against hybrids may be very high, or the hybrid zone may be trapped in an area with low population density (Endler, 1977; Barton & Hewitt, 1985). Therefore I developed a set of ~ 200 nuclear markers based on transcriptome data (data unpublished).

In Chapter 3 I address the hypothesis of movement in a common toad hybrid zone by employing 31 nuclear gene based markers. Introgression from the southern *B. spinosus* into *B. bufo* is not significantly asymmetric, and the peak of admixture linkage disequilibrium appears to coincide with the larger portion of introgression. If this pattern was significant, it would rather point to asymmetric reproductive isolation, with hybridisation more successful north of the hybrid zone than south of the hybrid

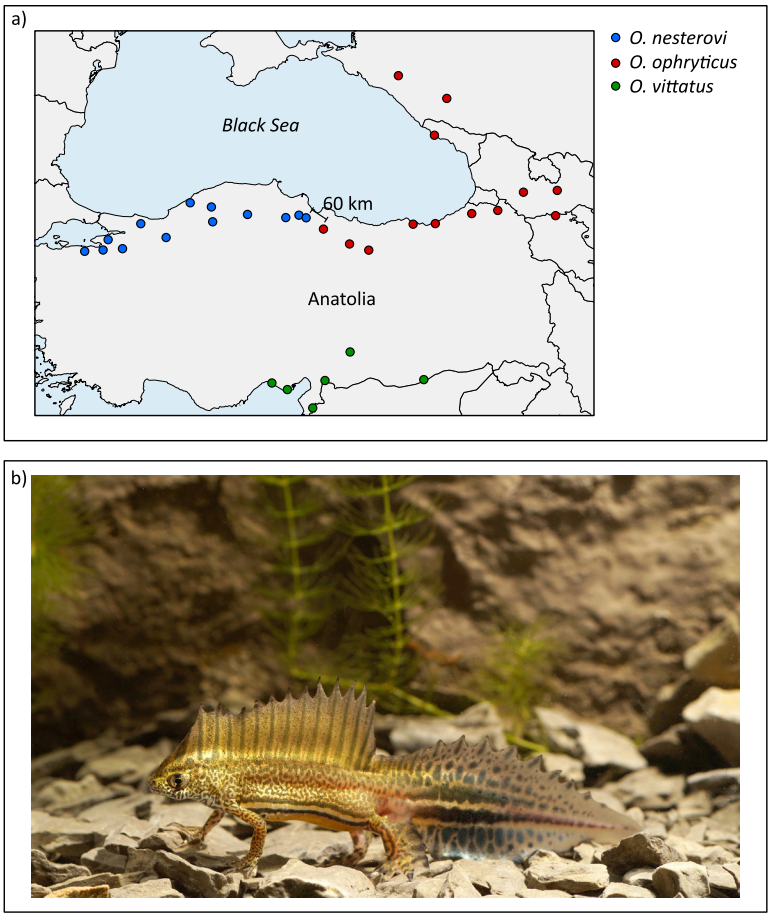


Figure 6.1: Map (a) showing northern part of the *Ommatotriton* distribution, with the 60 km gap between *O. nesterovi* and *O. ophryticus.* (b) Photograph of an *O. ophryticus* male in breeding season from Trabzon (Turkey). Photograph kindly provided by Michael Fahrbach.

zone. In Chapter 3 I propose an alternative explanation of low population sizes and slow dispersal, which would have increased genetic drift and obscured patterns of asymmetry.

The data for Chapter 3 is gene-coding, which may be more prone to selective pressures. In addition, the length of the *Bufo* hybrid zone makes it ideal to draw multiple transects. In Chapter 4 a replicate transect was thus included to test consistency of barrier genes (genes restricting introgression) and patterns of introgression and admixture linkage disequilibrium using a RAD dataset. I show asymmetric introgression in northwest France, with higher levels of introgression and admixture linkage disequilibrium coincident at the northern side of the hybrid zone, following predictions of asymmetric reproductive isolation. This strengthens the non-significant patterns found in Chapter 3. Introgression is symmetric in southeast France, and a peak of admixture linkage disequilibrium is found in the centre of the hybrid zone, following the prediction of a stable hybrid zone. The number of genomic outliers suggestive of heterozygote advantage and disadvantage is about half in

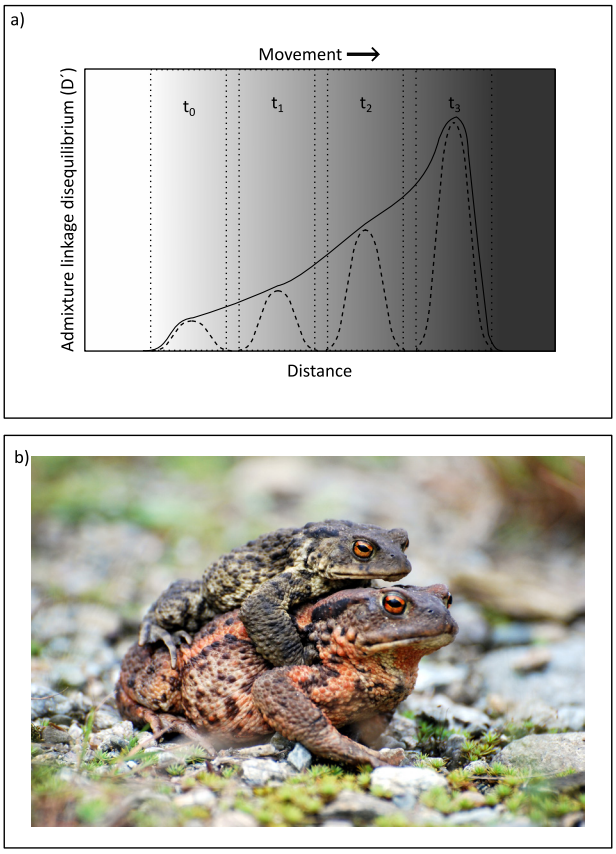


Figure 6.2: Hypothesis (a) for the shape of a peak of admixture linkage disequilibrium (D'; y-axis) under hybrid zone movement, with the position of the hybrid zone centre at four different time points (starting at t0 and ending at t3) indicated between dotted lines, distance on the x-axis. The peaks of D' under a stable situation, are expected to quickly disappear when migration no longer adds unadmixed individuals to the location. These peaks are drawn as dashed lines. The peak of D' expected during movement at time point t3 is drawn as a solid line. (b) Photograph of a *B. bufo* pair in amplexus with the male on top, from Cerna Sat (Romania). Photograph is kindly provided by Ioan Ghira.

northwest France compared to southeast France, and is in support of a difference in the number or strength of barrier genes between both transects. However, more markers are behaving as barrier markers in both transects than would be expected from random resampling. This suggests at least part of the barrier effect is shared between both transects.

Any explanation of these results should unify asymmetry and symmetry in different transects, explain the difference in the number of barrier markers in both transects, and explain the overlap of outlier markers in both transects. The involvement of divergent genetic groups involved in the hybrid zone may be underlying the differences between the transects. Only one mitochondrial lineage, originating from a refugium from the last glacial maximum in the Balkans, is involved on the *B. bufo* side of the zone in transect one, whilst three different mitochondrial lineages, originating from refugia in Italy, the northern Balkans and the Balkans, are involved in transect two (Fig. 6.2a; Garcia-Porta et al., 2012; Recuero et al., 2012; Arntzen et

al., 2017). The lineage of *B. spinosus* involved in the hybrid zone originated from the same Iberian refugium (Garcia-Porta et al., 2012; Recuero et al., 2012; Arntzen et al., 2017). This indicates that a deeper genetic structure can be found on the *B. bufo* side of the hybrid zone.

To explain our results, however, an underlying mechanism is needed. This mechanism might be found in a change in chromosome morphology related to the intraspecific variation in *B. bufo*. In northwest France, *B. bufo* with presumably homomorphic sex chromosomes are meeting with *B. spinosus* with a heteromorphic sex chromosome in females (Birstein & Mazin, 1982; Pisanets et al., 2009; Skorinov et al., 2018). In southeast France, *B. bufo* with presumably heteromorphic sex chromosome in males are meeting *B. spinosus* with a heteromorphic sex chromosome in females (Morescalchi, 1964; Skorinov et al., 2018). These different sex chromosome systems explain the different introgression patterns in the two transects, and a lower amount of barrier markers found in northwest France than in southeast France. Because the same chromosome is involved in all interactions, this would also explain the overlap between the barrier markers found in both transects.

The spatial variation in introgression, potentially related to intraspecific variation in sex-determination system, shows that the common toad hybrid zone is more complex than previously appreciated.

The role of sex chromosomes in Anuran speciation

A pair of chromosomes with a difference in morphology, are usually the sex chromosomes. Homomorphy of sex chromosomes and changes of sex chromosome systems (when the heterozygosity of sex chromosomes switches from male to female or vice versa), are well studied in amphibians. In XY systems where males are the heterogeneous sex, the genes causing 'maleness' are situated on the Y chromosome. Recombination between Y and X chromosomes are predicted to become suppressed. Genes on the Y chromosome that are not related to male development will be lost (Mueller's ratchet), and the Y chromosome will degenerate, making its' morphology visibly different from the X chromosome. Homomorphic sex chromosomes, such as is the case in *B. bufo* in northwest France, are thought to be maintained by recombination in phenotypic females with a male sex chromosome combination (phenotypic female with both X and Y chromosomes) which generates new Y haplotypes and thus prevents evolutionary decay of the Y chromosome (Perrin, 2009).

Changes in sex chromosome system occur exceptionally often in Anurans, even within genera (Hillis & Green, 1990; Uno et al., 2008; Stöck et al., 2011; Jeffries et al., 2018; Skorinov et al., 2018). In *Rana temporaria*, it has been shown that, indeed, recombination between sex chromosomes is not dependent on the genetic sex of an individual, but on the phenotypic sex (Rodrigues, Studer, Dufresnes, & Perrin, 2018). In the *Bufo viridis* subgroup recombination between sex chromosomes is also linked to sex chromosome homomorphy (Stöck et al., 2013). In the Eurasian Hyla radiation, homomorphy of the sex chromosomes was driven by both recombination between the sex chromosomes and by sex chromosome turnover (Dufresnes et al., 2015).

The possibility of a chromosome turnover still needs to be verified for the *Bufo* hybrid zone. Having both homomorphic and heteromorphic sex chromosome systems

within the same species complex may make *Bufo* a great system to study the role of sex chromosomes in Anuran speciation.

Effective selection against heterozygotes

The average effective selection against hybrids on a locus (s*) can be calculated using a few markers. However, since the introduction of s* in 1993 by Barton and Gale, the amount of genetic data available greatly increased, which would make the estimate more precise. I developed a script to calculate s* for many markers in a reproducible manner. Chapter 3 compares s* from hybrid zones in various taxa (Table 6.1). When s* is lower, it appears individual clines are more often displaced. This makes sense as with the alleviation of selection against hybrids, clines are free to move away from the hybrid zone centre. In hybrid zones where drift causes variation in the position of the individual (narrow) clines by locally driving alleles to fixation, such as I suggest may be the case in the *Bufo* hybrid zone, the cline of the expected frequency becomes wider (Polechová & Barton, 2011). This means that s* will, in general, be underestimated in hybrid zones where the effect of drift is high.

On the other hand, Table 6.1 shows that when the amount of markers used for the calculations increases, s* appears to decrease. This may be because using more markers makes the estimate more precise. It was possible to compare the effective selection against heterozygotes estimated using both the 32 nuclear marker dataset from Chapter 3 and the 1,189 nuclear marker dataset from Chapter 4 for the transect in northwest France. We show that the estimated s* is similar when expanding the dataset from 31 loci to 1,189 loci. For the *Bufo* system, apparently, tens of loci already capture the variability well enough to estimate s*. However, the confidence interval is narrower with 1,189 loci than with 31 loci.

To test if increasing amounts of markers decreases s*, larger datasets can be generated for earlier published hybrid zones with high estimates of s*. Such datasets probably do not need to contain more than a hundred markers.

Table 6.1: Estimated effective selection against heterozygotes (confidence intervals for s*). The table is ordered from low to high s*, and a line is drawn between low estimates of s* in studies which also report displaced clines, and high estimates of s* reporting only coincident clines.

Comon name	Estimated s*	Displaced	No. loci	Citation
Bufo bufo and *B. spinosus*	0.0001-0.004	Yes	31	van Riemsdijk, et al. (2019)
Bufo bufo and *B. spinosus*	0.001-0.003	Yes	1,189	van Riemsdijk, et al. in prep.
Malurus melanocephalus subsp.	0.002-0.03	Yes	103	Baldassarre, et al. (2014)
Two ecotypes of *Littorina saxatilis*	0.005-0.32	Yes	57	Hollander, et al. (2015)
Triturus anatolicus & *T. ivanbureschi*	0.004-0.019	Yes	49	Wielstra, et al. (2017)
Vandiemenella viatica sp.	0.058-0.405	No	11	Kawakami, et al. (2009)
Two lineages of *Carlia rubrigularis*	0.22-0.49	No	8	Phillips, et al. (2001)
Heliconius melpomene & *H. erato*	0.23	No	4	Mallet, et al. (1990)
Mus m. musculus & *M. m. domesticus*	0.28-0.48	No	7	Macholán, et al. (2007)
Bombina bombina & *B. variegata*	0.15-0.58	No	6	Szymura & Barton (1991)
Pontia daplidice & *P. edusa*	0.47-0.64	No	17	Porter, et al. (1997)
Oryctolagus c. cuniculus & *O. c. algirus*	0.5-0.64	No	17	Carneiro, et al. (2013)

Asymmetric introgression & admixture linkage disequilibrium

Alleles originating from the same parent species are found together within the genomes of hybrid offspring, causing admixture linkage disequilibrium. The benefit of plotting the admixture linkage disequilibrium (D') is that it shows where most recently fresh genetic material from one of the parent species, which has not had the chance to recombine during backcrossing, has come into the hybrid zone. This peak is hypothesised to be shifted ahead of the hybrid zone movement, to the opposite side of the trail of introgression (Gay, Crochet, Bell, & Lenormand, 2008; Wang et al., 2011). During asymmetric reproductive isolation the peak is hypothesised to be more coincident with the tail of introgression (Devitt, Baird, & Moritz, 2011). However, the distance a peak of D' can move within the hybrid zone centre is per definition restrained to the area in which introgression occurs. If the hybrid zone is a stepped cline, this distance becomes rather small, and the highest D' may be coincident with the place where the expected frequency cline is most steep, regardless of what process occurs.

In a neutral situation, the peak of D' is expected to follow a Gaussian curve. What shape would be expected during movement or asymmetric reproductive isolation is not clear yet. In a hybrid zone between two gull species, it appears however that the peak of D' is not following a Gaussian curve, but rather a Gaussian curve with a tail, a type of β distribution (Gay et al., 2008; Fig. 6.2). The further the hybrid zone has moved away from its original position, the longer recombination has taken place to decrease the effect of D', and the lower the local D' will be. In both transects from Chapter 4, the shape of the peak of D' is not very well described by a Gaussian curve, either. I would argue the shape of the peak under asymmetric reproductive isolation could be highly dependent on where the clines of the barrier genes are located. As barrier genes are causing the incompatibility between the species, the place where the cline centres of the barrier genes coincide is the place where most likely highest D' will be measured. What will happen to the shape of the peak of D' during hybrid zone movement caused by asymmetric reproductive isolation, is not clear yet either.

The concept of a peak of D' was more nuanced in the paper by Gay et al. (2008), as they described an excess of intermediates in the tail of the hybrid zone movement and a deficit of intermediates ahead of the hybrid zone movement. The two gull species have different morphologies and intermediates were defined by the morphology in the hybrid populations. The observation supporting this concept was (1) that an excess of individuals with intermediate morphology were found in the tail of the movement, and (2) a deficit of intermediates was found at the leading edge of the hybrid zone where only individuals with a morphology of either species (and no individuals of hybrid morphology were present) were found. In absence of morphological differences, other measures are available to assess both excess and deficit of intermediates, besides the measure of D' we used for the *Bufo* hybrid zone. For example, the squared correlation of alleles at two sites (r^2) could be used, which shows less inflation than D' in small samples (Weir, 1996; Ardlie, Kruglyak, & Seielstad, 2002; Weiss & Clark, 2002; Wang et al., 2011).

The shape of the peak of D' may be indicative of different processes, but to hypothesise without further evidence would be pure conjecture. Extensive simulations

may provide a first step on the way to be able to clearly outline the expectations of a peak of D' under different scenarios.

Future perspectives

This thesis contributed to our knowledge of the *Ommatotriton* and *Bufo* systems. Many new questions have arisen for both these systems, and for the study of dynamic hybrid zones in general.

For *Ommatotriton*, the question remains if there is introgression between *O. nesterovi* and *O. ophryticus*. More populations may be sampled in the 60 km gap between the two species distributions, and I developed a set of ~ 200 nuclear markers based on transcriptome data to asses introgression (data unpublished).

The existence of intraspecific *B. bufo* groups can be studied using many nuclear markers. A possible setup would be to use sequence capture or RAD sequencing in samples from different sections of the hybrid zone and include samples covering the wider distribution of *B. bufo* to study nuclear phylogeographic patterns. This allows to determine the genetic groups involved in the hybrid zone first, and then determine diagnostic markers based on reference populations. After establishing which genetic groups are involved in the hybrid zone, one can also make an informed decision about the setup of a chromosome study by karyotype analyses, linkage map, or genome sequencing. Crossings within and between *B. bufo* groups and *B. spinosus* can be used to investigate mechanisms involved in genetic sex determination and patterns of reproductive isolation.

The expectations about the shape of a peak of D' in a dynamic hybrid zone remain speculative. Two lines of investigation can be set out to determine the shape of this peak in more detail. The first is to simulate different scenarios of hybrid zone movement and asymmetric reproductive isolation and test for different variables, such as selection against hybrids, dispersal, and recombination rate. The other is to gather some published datasets, which show asymmetric introgression and for which good additional evidence for the cause of the asymmetry is available. Both simulations and published datasets can further inform expectations of the shape of the peak of D', or expectations of other measures such as r^2.

Instead of estimating measures of linkage disequilibrium it is possible to directly infer linkage of markers if reliable genome resources such as dense linkage maps or genome assemblies are available. Chromosome or ancestry blocks are stretches of chromosomes which can be assigned to the parent species and reflect inheritance patterns (Fisher, 1954; Baird, Barton, & Etheridge, 2003; Luikart, England, Tallmon, Jordan, & Taberlet, 2003; Wall & Pritchard, 2003; Gravel, 2012; Harris & Nielsen, 2013; Hellenthal et al., 2014; Sedghifar, Brandvain, Ralph, & Coop, 2015; Sedghifar, Brandvain, & Ralph, 2016). A junction, the border of two chromosome blocks, is a break in a chromosome caused by a recombination event, and is inherited as a point mutation (Fisher, 1954; Baird, 2006). The more generations of backcrossing occur, the more recombination has taken place, the more junctions will be present, and the smaller the genome blocks will become. The size of chromosome blocks can thus be used to investigate the amount of time passed since hybridisation first occurred (Baird, 1995; Chapman & Thompson, 2002; Pool & Nielsen, 2009).

Where chromosomes contain barrier genes, such as heterozygote disadvantage or Dobzhansky-Muller incompatibilities, chromosome blocks are predicted to be long and junctions few (Barton, 1983; Sedghifar et al., 2016; Hvala, Frayer, & Payseur, 2018). For example, a linkage map based on a single cross of two divergent lineages of *Rana temporaria* showed low recombination rate linked to genetic incompatibilities between the parental populations (Palomar et al., 2016). When a hybrid zone has moved, longer chromosome blocks can be found at the leading edge (Wang et al., 2011; Seixas, Boursot, & Melo-Ferreira, 2018).

These chromosome block studies all have one thing in common; the availability of genomic resources. During the last years, sequencing methods have become cheaper, more advanced, and more reliable. It is now possible to complete a full-scale genomic analysis within a PhD project for many non-model organisms (Matz, 2017). Amphibian genomes vary widely in size and contain many repetitive regions, which has made it more difficult to accumulate genomic resources for amphibians (Liedtke, Gower, Wilkinson, & Gomez-Mestre, 2018; Nowoshilow et al., 2018). Long reads (> 10,000 base pairs [bp]) can be used to bridge repetitive regions, and reliably assemble such large genomes.

Two well-known methods are available for sequencing long reads; Pacific Biosciences (PacBio), and Oxford Nanopore Technologies (ONT) platforms (Matz, 2017). The accuracy for PacBio has improved from 82% to 87%, and maximum read length is 50,000 bp (Koren and Phillippy, 2015, and references therein; Watson and Warr, 2019). The accuracy of ONT has improved from 70-80% to over 85%, and maximum read length is up to 800,000 bp (Quick, Quinlan, & Loman, 2014; Risse et al., 2015; Michael et al., 2018; Watson & Warr, 2019). For chromosome block analysis it is not necessary to have a highly accurate reference genome. As polymorphism in non-model populations is usually higher than 1%, the reference based on a single individual is inaccurate with respect to the genomes of the rest of the population, no matter how accurate the reference genome is (Matz, 2017; Koren, Phillippy, Simpson, Loman, & Loose, 2019).

At the start of 2019, five genome assemblies of considerable quality have been published for amphibians (Table 6.2). The *Xenopus tropicalis* genome is the most complete assembly available for amphibians. The genome was originally sequenced

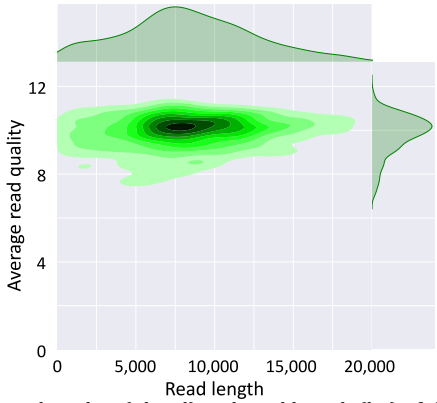


Figure 6.3: Plot for average read quality (phred) and read length (bp) of the raw reads obtained from a single MinION flow cell run for liver tissue of *B. bufo*.

Table 6.2: Amphibian genome assemblies available in February 2019, ordered by scaffold N50. The N50 is a measure of assembly quality, and represents the minimum contig length of the contigs covering half the total genome assembly length.

Species (Genome size)	Scaffold N50 (kbp)	Reference
Xenopus tropicalis (1.4 Gbp)	135,000	Hellsten et al. (2010)
Ambystoma mexicanum (32 Gbp)	3,000	Nowoshilow et al. (2018)
Nanorana parkeri (2.3 Gbp)	1,050	Sun et al. (2015)
Rhinella marina (2.5 Gbp)	168	Edwards et al. (2018)
Rana catesbeiana (6 Gbp)	69	Hammond et al. (2017)

with short read sequences and transcriptome data, and has been updated and improved on since publication (Table 6.2: Hellsten et al., 2010; Vize and Zorn, 2016). The *Nanorana parkeri* and *Rana catesbeiana* genomes have, too, been sequenced with short reads and transcriptome data but are much less complete than the *Xenopus assembly* (Sun et al., 2015; Hammond et al., 2017). For the *Ambystoma mexicanum* and *Rhinella marina* genome assemblies, PacBio long read sequencing and short read sequencing or transcriptome data were used. This resulted in more complete assemblies than for *N. parkeri* and *R. catesbeiana*, but the assemblies are still nowhere near the quality of the *Xenopus* assembly (Edwards et al., 2018; Nowoshilow et al., 2018).

The use of ONT sequencing for amphibian genomes has not yet been explored, whilst its reads are of similar quality and may be much longer than PacBio reads. I tested whether ONT could be applied to assemble the *B. bufo* genome, and successfully sequenced high quality reads (unpublished data; Fig. 6.3). The application of long read sequencing to amphibian study systems has great potential for future evolutionary research.

Conclusion

The three species in the genus of *Ommatotriton* have split a long time ago in a relatively short time span. If introgression occurs between *O. nesterovi* and *O. ophryticus* in their natural range, it is geographically limited. Hybridisation between *O. nesterovi* and *O. ophryticus* can occur, and I narrowed down an area in the natural distribution where introgression between the two species could take place. I have developed genetic tools for detailed analysis of genetic introgression.

Previously identified asymmetries in introgression along the transect of *Bufo bufo* and *B. spinosus* were confirmed for northwest France, but in southeast France the zone appears to be symmetric. There appears to be asymmetric reproductive isolation between the two *Bufo* species, and a difference in reproductive isolation with *B. spinosus* for different genetic groups within *B. bufo*. It appears the variation in introgression is linked to differences in chromosome morphology.

The availability of the two datasets for the *B. bufo* hybrid zone allowed to compare the estimated effective selection against hybrids (s*) based on 31 and 1,189 markers. From this, it appears datasets of no more than a hundred markers may already enable stable estimations of s*. With the availability of a script to estimate s* for larger datasets, it is possible to reproducibly estimate s* and compare the barrier effect within or between different taxonomic groups.

The position of a peak of admixture linkage disequilibrium (D') has been useful in studying hybrid zones, but many questions remain about what is expected for the shape and position under different scenarios of hybrid zone movement and asymmetric reproductive isolation. The shape of the peak under hybrid zone movement may be more like a peak with a tail than a Gaussian curve. Where possible, a more nuanced hypothesis of a heterozygote excess and deficit may be more informative for hybrid zones involving movement. By using simulations and previous datasets, expectations of the shape of the peak and behaviour of different statistics needs to further be tested.

With the increasing availability of genome resources, a chromosome block approach will be applicable to many more non-model systems, and will allow detection of more detailed patterns of differential introgression along the genome. This will aid in the identification of barrier genes and the causes underlying asymmetries in hybrid zones. I anticipate that researchers will be able to answer many of the open questions about amphibian evolution with the aid of long read sequencing.

We have surpassed the point where hybrid zones are thought to be in a stable situation by default.

References

Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics, 3, 299–309. https://doi.org/10.1038/nrg777

Arntzen, J. W., de Vries, W., Canestrelli, D., & Martínez-Solano, I. (2017). Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Molecular Ecology, 26, 5663–5675. https://doi.org/10.1111/mec.14273

Baird, S. J. E. (1995). A simulation study of multilocus clines. Evolution, 49, 1038–1045. https://doi.org/10.2307/2410429

Baird, S. J. E. (2006). Phylogenetics: Fisher's markers of admixture. Heredity, 97, 81–83. https://doi.org/10.1038/sj.hdy.6800850

Baird, S. J. E., Barton, N. H., & Etheridge, A. M. (2003). The distribution of surviving blocks of an ancestral genome. Theoretical Population Biology, 64, 451–471. https://doi.org/10.1016/S0040-5809(03)00098-4

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution, 68, 2644–2657. https://doi.org/10.1111/evo.12457

Barton, N. H. (1983). Multilocus clines. Evolution, 37, 454–471. https://doi.org/10.2307/2408260

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual Review of Ecology and Systematics, 16, 113–148.

Birstein, V. J., & Mazin, A. L. (1982). Chromosomal polymorphism of *Bufo bufo*: karyotype and C-banding pattern of *B. b. verrucosissima*. Genetica, 59, 93–98. https://doi.org/10.1007/BF00133292

Carneiro, M., Baird, S. J. E., Afonso, S., Ramirez, E., Tarroso, P., Teotõnio, H., … Ferrand, N. (2013). Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. Molecular Ecology, 22, 2511–2525. https://doi.org/10.1111/mec.12272

Chapman, N. H., & Thompson, E. A. (2002). The effect of population history on the lengths of ancestral chromosome segments. Genetics, 162, 449–458.

Devitt, T. J., Baird, S. J. E., & Moritz, C. (2011). Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine--scale genetic analysis of a hybrid zone. BMC Evolutionary Biology, 11. https://doi.org/10.1186/1471-2148-11-245

Dufresnes, C., Borzee, A., Horn, A., Stock, M., Ostini, M., Sermier, R., … Perrin, N. (2015). Sex--chromosome homomorphy in palearctic tree frogs results from both turnovers and X-Y recombination. Molecular Biology and Evolution, 32, 2328–2337. https://doi.org/10.1093/molbev/msv113

Edwards, R. J., Tuipulotu, D. E., Amos, T. G., O'Meally, D., Richardson, M. F., Russell, T. L., … White, P. A. (2018). Draft genome assembly of the invasive cane toad, *Rhinella marina*. GigaScience, 7, 1–13. https://doi.org/10.1093/gigascience/giy095

Endler, J. A. (1977). Geographic variation, speciation, and clines. (R. M. May, Ed.) (2nd ed.). Princeton, New Jersey: Princeton University Press.

Fisher, R. A. (1954). A fuller theory of "junctions" in inbreeding. Heredity, 8, 187–197.

Fontelles, F., Guixé, D., Martínez-Silvestre, A., Soler, J., & Villero, D. (2011). Hallada una población introducida de *Ommatotriton ophryticus* en el Prepirineo catalán. Boletín de La Asociación Herpetológica Española, 22, 153–156.

Garcia-Porta, J., Litvinchuk, S. N., Crochet, P. A., Romano, A., Geniez, P. H., Lo-Valvo, M., … Carranza, S. (2012). Molecular phylogenetics and historical biogeography of the west--palearctic common toads (*Bufo bufo* species complex). Molecular Phylogenetics and Evolution, 63, 113–130. https://doi.org/10.1016/j.ympev.2011.12.019

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. Evolution, 62, 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Gravel, S. (2012). Population genetics models of local ancestry. Genetics, 191, 607–619. https://doi.org/10.1534/genetics.112.139808

Hammond, S. A., Warren, R. L., Vandervalk, B. P., Kucuk, E., Khan, H., Gibb, E. A., … Birol, I. (2017). The North American bullfrog draft genome provides insight into hormonal regulation of long. Nature Communications, 8, 1–8. https://doi.org/10.1038/s41467-017-01316-7

Harris, K., & Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genetics, 9, 1–20. https://doi.org/10.1371/journal.pgen.1003521

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic atlas of human admixture history. Science, 343, 747–751. https://doi.org/10.1126/science.1243518

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., … Rokhsar, D. (2010). The genome of the western clawed frog *Xenopus tropicalis*. Science, 328, 633–636. https://doi.org/10.1126/science.1183670

Hillis, D. M., & Green, D. M. (1990). Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. Journal of Evolutionary Biology, 3, 49–64. https://doi.org/10.1046/j.1420-9101.1990.3010049.x

Hollander, J., Galindo, J., & Butlin, R. K. (2015). Selection on outlier loci and their association with adaptive phenotypes in *Littorina saxatilis* contact zones. Journal of Evolutionary Biology, 28, 328–337. https://doi.org/10.1111/jeb.12564

Hvala, J. A., Frayer, M. E., & Payseur, B. A. (2018). Signatures of hybridization and speciation in genomic patterns of ancestry. Evolution, 72, 1540–1552. https://doi.org/10.1111/evo.13509

Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., … Perrin, N. (2018). A

rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature Communications, 9, 4088. https://doi.org/10.1038/s41467-018-06517-2

Kawakami, T., Butlin, R. K., Adams, M., Paull, D. J., & Cooper, S. J. B. (2009). Genetic analysis of a chromosomal hybrid zone in the Australian morabine grasshoppers (*Vandiemenella, viatica* species group). Evolution, 63, 139–152. https://doi.org/10.1111/j.1558-5646.2008.00526.x

Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Current Opinion in Microbiology, 23, 110–120. https://doi.org/10.1016/j.mib.2014.11.014

Koren, S., Phillippy, A. M., Simpson, J. T., Loman, N. J., & Loose, M. (2019). Reply to 'Errors in long-read assemblies can critically affect protein prediction.' Nature Biotechnology, 37, 127–128. https://doi.org/10.1038/s41587-018-0005-y

Liedtke, H. C., Gower, D. J., Wilkinson, M., & Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-018-0674-4

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews. Genetics, 4, 981–994. https://doi.org/10.1038/nrg1226

Macholán, M., Munclinger, P., Šugerková, M., Dufková, P., Bímová, B., Božíková, E., … Piálek, J. (2007). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution, 61, 746–771. https://doi.org/10.1111/j.1558-5646.2007.00065.x

Mallet, J., Barton, N., Lamas, G. M., Santisteban, J. C., Muedas, M. M., & Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics, 124, 921–936.

Matz, M. V. (2017). Fantastic Beasts and how to sequence them: ecological genomics for obscure model organisms. Trends in Genetics, 34, 121–132. https://doi.org/10.1016/j.tig.2017.11.002

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., … Ecker, J. R. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. Nature Communications, 9, 1–8. https://doi.org/10.1038/s41467-018-03016-2

Morescalchi, A. (1964). Il corredo cromosomico dei Bufonidi Italiani. Bolletino Di Zoologia, 31, 827–836. https://doi.org/10.1080/11250006409441116

Nadachowska, K., & Babik, W. (2009). Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). Molecular Biology and Evolution, 26, 829–841. https://doi.org/10.1093/molbev/msp004

Nowoshilow, S., Schloissnig, S., Fei, J., Dahl, A., Andy, W., Pang, C., … Roscito, J. G. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature Publishing Group, 554, 50–55. https://doi.org/10.1038/nature25458

Palomar, G., Ahmad, F., Vasem, A., Matsuba, C., Nicieza, A. G., Cano, M., … Sciences, L. (2016). Comparative high-density linkage mapping reveals conserved genome structure but variation in levels of heterochiasmy and location of recombination cold spots in the common frog. https://doi.org/10.1534/g3.116.036459

Perrin, N. (2009). Sex reversal: a fountain of youth for sex chromosomes? Evolution, 63, 3043–3049. https://doi.org/10.1111/j.1558-5646.2009.00837.x

Phillips, B. L., Baird, S. J. E., & Moritz, C. (2004). When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. Evolution, 58, 1536. https://doi.org/10.1554/02-498

Pisanets, E. M., Litvinchuk, S. N., Rosanov, J. M., Reminniy, V. Y., Pasynkova, R. A., & Suryadnaya, N. N. M. A. S. (2009). Common toads (Amphibia, Bufonidae, *Bufo bufo* complex) from the Ciscaucasia and north of the Caucasus: the new analysis of the problem. Zbirnik Prats' Zool. Mus., 40, 83–125. [in Russian]

Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. Genetics, 189, 227–235. https://doi.org/10.1534/genetics.111.129817

Pool, J. E., & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics, 181, 711–719. https://doi.org/10.1534/genetics.108.098095

Porter, A. H., Wenger, R., Geiger, H., Scholl, A., & Shapiro, A. M. (1997). The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. Evolution, 51, 1561. https://doi.org/10.2307/2411208

Quick, J., Quinlan, A. R., & Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinIONTM portable single-molecule nanopore sequencer. GigaScience, 3, 1–6. https://doi.org/10.1186/2047-217X-3-22

Recuero, E., Canestrelli, D., Vörös, J., Szabó, K., Poyarkov, N. A., Arntzen, J. W., … Martínez--Solano, I. (2012). Multilocus species tree analyses resolve the radiation of the widespread *Bufo bufo* species group (Anura, Bufonidae). Molecular Phylogenetics and Evolution, 62, 71–86. https://doi.org/10.1016/j.ympev.2011.09.008

Risse, J., Thomson, M., Patrick, S., Blakely, G., Koutsovoulos, G., Blaxter, M., & Watson, M. (2015). A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. GigaScience, 4, 1–7. https://doi.org/10.1186/s13742-015-0101-6

Rodrigues, N., Studer, T., Dufresnes, C., & Perrin, N. (2018). Sex-chromosome recombination in common frogs brings water to the fountain-of-youth. Molecular Biology and Evolution, 35, 942–948. https://doi.org/10.1093/molbev/msy008

Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular Ecology, 25, 2559–2576. https://doi.org/http://dx.doi.org/10.1101/043190

Sedghifar, A., Brandvain, Y., Ralph, P., & Coop, G. (2015). The spatial mixing of genomes in secondary contact zones. Genetics, 201, 243–261. https://doi.org/10.1534/genetics.115.179838

Seixas, F. A., Boursot, P., & Melo-Ferreira, J. (2018). The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biology, 19, 1–20. https://doi.org/10.1186/s13059-018-1471-8

Skorinov, D. V, Bolshakova, D. S., Donaire, D., Pasynkova, R. A., Litvinchuk, S. N., & Litvinchuk, B. (2018). Karyotypic analysis of the spined toad, *Bufo spinosus* Daudin, 1803 (Aphibia: Bufonidae). Russian Journal of Herpetology, 25, 253–258. https://doi.org/10.30906/1026-2296-2018-25-4-253-258

Stöck, M., Croll, D., Dumas, Z., Biollay, S., Wang, J., & Perrin, N. (2011). A cryptic heterogametic transition revealed by sex-linked DNA markers in Palearctic green toads. Journal of Evolutionary Biology, 24, 1064–1070. https://doi.org/10.1111/j.1420-9101.2011.02239.x

Stöck, M., Savary, R., Betto-Colliard, C., Biollay, S., Jourdan-Pineau, H., & Perrin, N. (2013). Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). Journal of Evolutionary Biology, 26, 674–682. https://doi.org/10.1111/jeb.12086

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., … Zhang, Y.-P. (2015). Whole--genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences of the United States of

America, 112, E1257-62. https://doi.org/10.1073/pnas.1501764112
Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. Evolution, 45, 237–261.
Uno, Y., Nishida, C., Oshima, Y., Yokoyama, S., Miura, I., Matsuda, Y., & Nakamura, M. (2008). Comparative chromosome mapping of sex-linked genes and identification of sex chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems. Chromosome Research, 16, 637–647. https://doi.org/10.1007/s10577-008-1217-7
Üzum, N., Avci, A., Olgun, K., Bülbül, U., Fahrbach, M., Litvinchuk, S. N., Wielstra, B. (2019). Cracking cryptic species: external characters to distinguish two recently recognized banded newt species (*Ommatotriton ophryticus* and *O. nesterovi*). In press.
van Riemsdijk, I., Butlin, R. K., Wielstra, B., & Arntzen, J. W. (2019). Testing an hypothesis of hybrid zone movement for toads in France. Molecular Ecology, in press, 0–3. https://doi.org/10.1111/mec.15005
Vize, P. D., & Zorn, A. M. (2016). *Xenopus* genomic data and browser resources. Developmental Biology, 426, 194–199. https://doi.org/10.1016/j.ydbio.2016.03.030
Wall, J. D., & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. Nature Reviews Genetics, 4, 587–597. https://doi.org/10.1038/nrg1123
Wang, L., Luzynski, K., Pool, J. E., Janoušek, V., Dufkovú, P., Vyskočilovú, M. M., … Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. Molecular Ecology, 20, 2985–3000. https://doi.org/10.1111/j.1365-294X.2011.05148.x
Watson, M., & Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. Nature Biotechnology, 37, 124–126. https://doi.org/10.1038/s41587-018-0005-y
Weir, B. S. (1996). Genetic Data Analysis II. Sunderland: Sinauer Publications.
Weiss, K. M., & Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. Trends in Genetics, 18, 19–24. https://doi.org/10.1016/S0168-9525(01)02550-1
Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzüm, N., Bozkurt, E., … Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. Evolution Letters, 1, 93–101. https://doi.org/10.1002/evl3.9

Summary

Hybrid zones occur where two species meet and produce offspring (hybrids). Typically, hybrids show a considerable reduction in fitness. When two species have diverged in allopatry (different geographic locations), meet again after they expand their distributions, and if the two are not yet completely reproductively isolated, the hybrid zone classifies as a secondary contact. Newly evolved gene variants of one species may be incompatible with the gene variants in the other species, and may underlie (incomplete) reproductive isolation. When such incompatibilities become linked in the genome, e.g. because they are located on the same stretch of DNA, or are distributed throughout all the parental populations, they may cause a genome--wide reduction of gene flow between the two species, compared to the dispersal rate of individuals. This is referred to as the 'barrier effect'. The genes that underlie the barrier effect may also be called 'barrier genes'.

In the classic hybrid zone literature, hybrid zones are thought to stabilise (geographically) soon after they are formed. However, recent studies show that hybrid zones can shift across considerable distances after their formation. During hybrid zone movement, one species replaces the other because of a fitness advantage of one of the species over the other. A genetic footprint of neutral gene variants of the displaced species remaining in populations that have been overtaken by other species would provide evidence of hybrid zone movement.

During meiosis (egg or sperm formation) genes segregate and randomly combine again at fertilization. During this process, the DNA can break and be recombined, resulting in new combinations of genes. Linkage disequilibrium is the association of multiple genes, and decays with segregation and recombination. The closer two gene variants are on a chromosome, the less likely recombination will take place. Genes which are different between two species involved in a hybrid zone, show a peak of linkage disequilibrium in the hybrid zone centre. The position of this peak indicates where, geographically, recombination of the parental species' genes has taken place the least during hybridisation. If a hybrid zone moves, this peak is expected to be positioned on the side of the hybrid zone in the direction of movement. In this thesis, two potentially dynamic hybrid zones are studied to further characterise the genetic footprint and expected peak of linkage disequilibrium caused by hybrid zone movement.

Two species of banded newts (*Ommatotriton nesterovi* and *O. ophryticus*) are thought to meet in a hybrid zone in the north of Turkey. The two other pairs of salamander species in this region (*Triturus anatolicus* and *T. ivanbureschi*, and *Lissotriton vulgaris* and *L. kosswigi*) show a genetic footprint due to hybrid zone movement over hundreds of kilometres. In this thesis I confirm the species status of the two banded newt species based on mitochondrial DNA and two nuclear DNA sequences. The location of the hybrid zone is narrowed down to a 60 km wide region where no samples were available. No genetic footprint is found within the natural distribution outside of this 60 km wide region. Thus, if a hybrid zone is present between *O. nesterovi* and *O. ophryticus*, it is narrower than 60 km and it may be geographically stable. An introduced population of hybrid banded newts in Spain provides evidence that the two species can produce fertile offspring. This increases the likelihood that a

hybrid zone indeed exists between the two species.

Common and spined toads (*Bufo bufo* and *B. spinosus*) meet in an 800 km long hybrid zone that runs diagonally across France. A genetic footprint north of the hybrid zone was previously recorded and linked to southward hybrid zone movement, with *B. bufo* overtaking *B. spinosus.* To test hypotheses of hybrid zone movement, a transect in northwest France was studied with 31 nuclear markers. A genomic footprint towards the north indicated southward movement, whilst a peak of linkage disequilibrium in the north is in line with northward movement. The contrasting results suggest that stronger reproductive isolation on the *B. spinosus* side of the hybrid zone than on the *B. bufo* side, may be more likely than hybrid zone movement.

Replication of transects along a hybrid zone helps to identify factors driving hybrid zone movement. Two distant transects, one in the northwest and one in the southeast of France, were studied using 1200 nuclear markers. Asymmetries which were previously found in the literature and in this thesis, were confirmed for the hybrid zone in northwest France, but not in the southeast, where the gene flow appears to be symmetric, indicating the hybrid zone is stable here. Barrier markers, genetic regions which may be associated with barrier genes, were identified by their relatively restricted gene flow. The number of shared markers which behave as a barrier marker in both transects is higher than expected when the barrier effect was established in each transect separately. This overlap suggests that the two species have evolved a universal genetic barrier to gene flow. The differences in patterns of gene flow between the transects may be caused by genetic divergence within *B. bufo,* documented in previous phylogeographical work.

We clearly can no longer think of hybrid zones as static upon formation; hybrid zones evolve! The position of a peak of linkage disequilibrium generated testable hypotheses, but many questions remain about what is expected for the shape and position of this peak under different scenarios of hybrid zone movement. With the increasing availability of genome resources, detection of more detailed patterns of differential introgression along the genome in hybrid zone studies will reveal the genetic architecture of speciation and the evolution of hybrid zones.

Samenvatting

Hybride zones ontstaan waar twee soorten elkaar tegenkomen en nakomelingen (hybrides) produceren. Hybrides hebben vaak een behoorlijk verlaagd voortplantingssucces. Hybride zones kunnen geclassificeerd worden als secundair contact, wanneer twee soorten afgesplitst zijn in allopatrie (verschillende geografische locaties), zij elkaar na verspreiding weer tegenkomen, en zij nog niet volledig reproductief geïsoleerd zijn. Gedurende het secundaire contact kan blijken dat combinaties van nieuwe varianten van genen van de twee soorten een nadelig effect hebben, en een verlaging van voortplantingssucces van de hybriden veroorzaken. Waneer zulke incompatibiliteiten dicht bij elkaar liggen in het genoom, of dezelfde verspreiding krijgen in de ouderlijke populaties, of beide, kunnen ze een reductie van de uitwisseling van het hele genoom tussen de twee soorten teweeg brengen wanneer de uitwisseling van genen vergeleken wordt met de dispersiesnelheid van individuen, ook wel het 'barrière-effect' genoemd. De genen die het barrière-effect veroorzaken worden ook wel 'barrière-genen' genoemd.

In de klassieke literatuur over hybride zones wordt ervan uitgegaan dat hybride zones zich (geografische) snel stabiliseren na vorming. In de recentere literatuur wordt steeds vaker rekening gehouden met de mogelijkheid dat hybride zones over aanzienlijke afstanden kunnen bewegen. Gedurende het bewegen van een hybride zone vervangt de ene soort de andere doordat er sprake is van een competitief voordeel van de ene soort over de andere. Een genetische voetafdruk, wanneer neutrale genvarianten van de weggevaagde soort achterblijven in populaties van de overheersende soort, kan bewijs leveren voor beweging van een hybride zone.

Gedurende meiosis (ei- en sperma- formatie) segregeren genen en worden bij fertilisatie willekeurig gecombineerd. Het DNA kan hierbij breken en recombineren, wat resulteert in nieuwe combinaties van genen. 'Linkage disequilibrium' is de associatie van genen, welke afneemt door segregatie en recombinatie. Hoe dichter bij elkaar twee genen op een chromosoom liggen, hoe minder recombinatie er plaats kan vinden. Genen die verschillen tussen twee soorten, laten een piek van linkage disequilibrium zien in het centrum van een hybride zone. De positie van deze piek geeft een indicatie waar (geografisch gezien) recombinatie het minst voor is gekomen gedurende hybridisatie. Als een hybride zone beweegt, wordt deze piek verwacht aan de zijde van de zone in de richting van de beweging. In deze thesis worden twee potentieel dynamische hybride zones bestudeerd om een beter beeld te krijgen van de genetische voetafdruk en de verwachtte piek van linkage disequilibrium die veroorzaakt worden door de beweging van een hybride zone.

Twee soorten van de bandsalamander (*Ommatotriton nesterovi* en *O. ophryticus*) hebben mogelijk een hybride zone in het noorden van Turkije. De twee andere paren van salamander-zustersoorten (*Triturus anatolicus* en *T. ivanbureschi*, en *Lissotriton vulgaris* en *L. kosswigi*) die in deze regio voorkomen laten patronen zien van beweging van de hybride zone over honderden kilometers in de vorm van een genetische voetafdruk. In deze thesis bevestig ik dat de twee soorten onderscheiden kunnen worden op basis van mitochondriale en nucleaire DNA-sequenties. De locatie van de hybride zone wordt benaderd tot een 60 km brede regio waar geen monsters genomen zijn. Echter, er is geen genetische voetafdruk gevonden in het natuurlijke

verspreidingsgebied buiten deze 60 km brede regio. Als er een hybride zone is tussen *O. nesterovi* en *O. ophryticus*, dan is die smaller dan 60 km en is het mogelijk dat de zone stabiel is. Een geïntroduceerde populatie van hybride bandsalamanders in Spanje bewijst dat de twee soorten vruchtbare nakomelingen kunnen krijgen. Dit vergroot de waarschijnlijkheid dat een hybride zone tussen deze twee soorten wel degelijk bestaat.

Twee soorten van de gewone pad (*Bufo bufo* en *B. spinosus*) komen elkaar tegen in een 800 km lange hybride zone die zich diagonaal over Frankrijk uitstrekt. Een genetische voetafdruk was eerder geobserveerd en gelinkt aan beweging van de hybride zone, waarbij *B. bufo B. spinosus* vervangt. Om de hypothese van een bewegende hybride zone te testen, is een transect in noordwest Frankrijk bestudeerd met 31 nucleaire markers. Een genetische voetafdruk ten noorden van de zone suggereert zuidwaards bewegende hybride zone, terwijl een piek van linkage disequilibrium aan de noordelijke zijde van de hybride zone een noordwaardse beweging suggereert. Deze tegensprekende resultaten wijzen op een grotere reproductieve isolatie aan de *B. spinosus* zijde van de hybride zone dan aan de *B. bufo* zijde.

Meerdere transecten door een hybride zone kunnen helpen bij het identificeren van de factoren die de beweging van de hybride zone drijven. Twee ver uiteengelegen transecten, in noordwest en zuidoost Frankrijk, zijn bestudeerd met 1200 nucleaire markers. Eerder in de literatuur en in deze thesis vastgestelde asymmetriën zijn hiermee bevestigd in noordwest Frankrijk, maar niet in zuidoost Frankrijk, waar de introgressie eerder symmetrisch blijkt te zijn, wat wijst op een stabiele hybride zone in dit gebied. Barrière-markers, genetische regios die geassocieerd kunnen worden met barrière genen, zijn geïdentificeerd door hun relatief gelimiteerde overdracht tussen de twee soorten. Er zijn meer markers geidentificeerd als barrière marker in beide transecten dan verwacht als het barrière-effect onafhankelijk ontstaan was in beide delen van de hybride zone. Deze overlap van barrièremarkers tussen beide transecten suggereert dat de twee soorten een universele barrière voor reproductie hebben, die over de gehele hybride zone werkt. De verschillen in het patroon van introgressie tussen de transecten kan veroorzaakt zijn door genetische verschillen tussen groepen binnen *B. bufo,* die beschreven zijn in eerder fylogeografisch werk.

We laten hiermee het uitgangspunt dat hybride zones stabiele situaties zijn, achter ons; hybride zones evolueren! De positie van een piek van linkage disequilibrium levert testbare hypotheses, maar veel vragen over de vorm en positie in verschillende scenario's van bewegende hybride zones blijven onbeantwoord. Door de groeiende beschikbaarheid van genomische informatiebronnen ligt de detectie van variatie in patronen van introgressie binnen het genoom binnen handbereik, en wordt het mogelijk de genetische architectuur van soortvorming in steeds groter detail te onderzoeken.

Curriculum vitae

Isolde van Riemsdijk was born on the 4th of January 1991 in Gouda in the Netherlands. After finishing high school at the Coornhert Gymnasium in Gouda in 2009, she started her academic education with a BSc and MSc in Biology on the subject of Evolution and Biodiversity at Wageningen University, the Netherlands, from 2009 to 2014. The subject of her MSc thesis was 'dating analysis of a phylogenetic tree of Annonaceae'. Isolde conducted an MSc internship of half a year at BioArch at the University of York in the UK, to identify bone specimens from an archaeological collection, for which she received an Erasmus grant. After finishing her MSc, she did a research project on bioinformatics tools for fragmented genome assemblies at the University of Leipzig under the supervision of Prof. Dr. K. Nowick, and PhD candidate R. Kolora. In 2015, she started her PhD on the subject of hybrid zone dynamics in amphibians at the Naturalis Biodiversity Center in Leiden, the Netherlands, under supervision of Dr. J.W. Arntzen and Dr. B. Wielstra. This position was supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014). During her PhD, Isolde visited the Shaffer Laboratory at the University of California for half a year, to set up a project using restriction-associated DNA sequencing, for which she received funding from the Leiden University Fund / Swaantje Mondt Fonds.

Publications

van Riemsdijk, I., Arntzen, J.W., Bucchiarelli, G., McCartney-Melstad, E., Rafajlovic, M., Scott, P., Toffelmier, E., Shaffer, H.B., Wielstra, B., (in prep.). Spatial variation in introgression along a toad hybrid zone in France.

van Riemsdijk, I., Struijk, R. P. J. H., Pel, E., Janssen, I. A. W., Wielstra, B. (subm.). Hybridisation complicates the conservation of *Natrix* snakes in the Netherlands.

van Riemsdijk, I., Butlin, R. K., Wielstra, B. Arntzen, J. W. (2018). Testing an hypothesis of hybrid zone movement for toads in France. Molecular Ecology, 28, 1070-1083.

van Riemsdijk, I., van Nieuwenhuize, L., Martínez-Solano, I., Arntzen, J. W., & Wielstra, B. (2017). Molecular data reveal the hybrid nature of an introduced population of banded newts (*Ommatotriton*) in Spain. Conservation Genetics, 19, 249-254.

van Riemsdijk, I., Arntzen, J. W., Bogaerts, S., Franzen, M., Litvinchuk, S. N., Olgun, K., & Wielstra, B. (2017). The Near East as a cradle of biodiversity: a phylogeography of banded newts (genus *Ommatotriton*) reveals extensive inter- and intraspecific genetic differentiation. Molecular Phylogenetics and Evolution, 114, 73–81.

Miller, J. A., **van Riemsdijk, I.**, Khalik, M. Z., Scager, D. J., & Schilthuizen, M. (2016). Comment on Falade et al. (2016) DNA-barcoding of *Clarias gariepinus*, *Coptedon zillii* and S*arotherodon melanotheron* from Southwestern Nigeria. F1000Research, 5, 2654.

Welker, F., Soressi, M. A., Roussel, M., **van Riemsdijk, I.**, Hublin, J.-J., & Collins, M. J. (2016). Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. STAR: Science & Technology of Archaeological Research, 3, 15–27.

Acknowledgements

Writing this thesis would not have been possible without the support of many people.

First, I want to thank both my supervisors. My supervisor Pim Arntzen has a vast knowledge of many amphibian species, and has been a guiding figure throughout the PhD trajectory. My supervisor Ben Wielstra always encouraged me to take a step further. He is a great example for me. This PhD position was supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek' (NWO Open Programme 824.14.014), granted to my supervisors.

I want to thank Menno Schilthuizen. His research group, and especially Barbara Gravendeel and Martin Rücklin, was always ready to get me out of the office for some good scientific conversations.

For my 6-month visit to the Shaffer laboratory at the University of California in Los Angeles I want to thank the entire group, but especially Peter Scott, Gary Bucciarelli, Tara Luckau, Erin Toffelmier, Evan McCartney-Melstad, Mario Colon, and Kevin Neal. I would like to thank Brad Shaffer for this great opportunity to be part of this diverse group of researchers. This study trip has been sponsored by the Leiden University Fund / Swaantje Mondt Fonds (D7102).

I want to thank all my co-authors, listed with the different papers, but in specific Roger Butlin. Roger and I had long conversations about the many analytical and theoretical aspects of this thesis. He always motivated me to go the extra mile. Thank you Roger!

I am grateful to the laboratory technicians of Naturalis, Roland Bûtot, Elza Duijm, Marcel Eurlings, and Frank Stokvis from Naturalis, and Onno Schaap from Leiden University for their practical support.

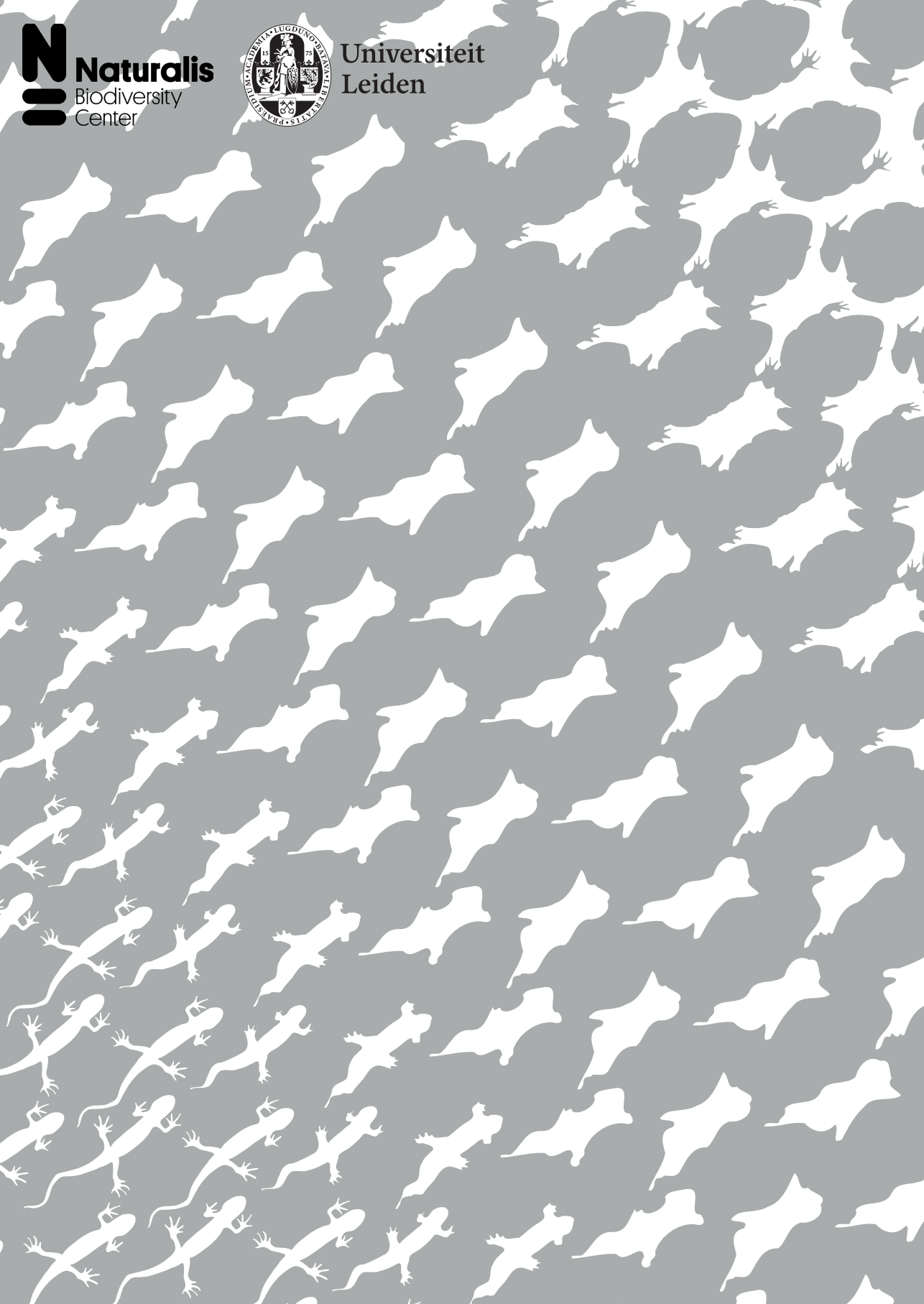
I would like to thank the PhD students from Naturalis, Anita, Kevin, Kasper, Sofia, Dewi, Richa, Zac, Andres, Roderick, Eka, Diego, Lisette, David, Tiedo, Larissa, Leon, Marcel, Saroj, and all the others, and Michael from Leiden University. My work has benefited from the hard work of many students, Davy, Eran, Emily, Julia, Jan Willem, Laurens, Mattijn, Ruben, and Steven. It was fun to work with you!

I would like to thank my mentors from the past, Lars Chatrou, Freek Bakker and Matthew Collins. Special thanks to Katja Nowick, who provided me with the opportunity to do an internship on genomics at her laboratory and encouraged me to develop my bioinformatics skills and share them with the world.

Much gratitude should go to my friends, Anoek, Caspara, Irene, Loes, Mariëlle, Marlies, Meike, Minke, Rogier, Simone, Tineke, and Tineke, who were always there to cheer me up. Jessie, who I see both as a friend, and as a shining example of a scientist, thank you.

Ik ben mijn kleine broertjes, Rombout en Wolbert, en mijn grote zus, Beatrijs, hun families, en mijn ouders, Bregje en Anton, bij wie ik altijd op kan laden met een kopje thee en een goed gesprek, heel dankbaar. Jullie zijn allemaal een belangrijk klankbord en grote steun voor mij. Ook Ans, Chris, Tami, Marijke, en mijn grootouders, dankjulliewel.

Tot slot mijn partner, Frido Welker, die me bij heeft gestaan gedurende deze wetenschappelijke reis. Dankjewel voor je ondersteuning, vertrouwen en bemoediging, waar en wanneer ik het nodig had. Du er mit geni.


Naturalis
Biodiversity
Center
Universiteit
Leiden