

Mating patterns in a hybrid zone of fire-bellied toads (*Bombina*): inferences from adult and full-sib genotypes

B Nürnberger¹, NH Barton², LEB Kruuk² and TH Vines^{1,2,3}

¹Department Biologie II, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany;

²Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland

We present two novel methods to infer mating patterns from genetic data. They differ from existing statistical methods of parentage inference in that they apply to populations that deviate from Hardy–Weinberg and linkage equilibrium, and so are suited for the study of assortative mating in hybrid zones. The core data set consists of genotypes at several loci for a number of full-sib clutches of unknown parentage. Our inference is based throughout on estimates of allelic associations within and across loci, such as heterozygote deficit and pairwise linkage disequilibrium. In the first method, the most likely parents of a given clutch are determined from the genotypic distribution of the associated adult population, given an explicit model of nonrandom mating. This leads to estimates of the strength of assortment.

The second approach is based solely on the offspring genotypes and relies on the fact that a linear relation exists between associations among the offspring and those in the population of breeding pairs. We apply both methods to a sample from the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* (Anura: Discoglossidae) in Croatia. Consistently, both approaches provide no evidence for a departure from random mating, despite adequate statistical power. Instead, *B. variegata*-like individuals among the adults contributed disproportionately to the offspring cohort, consistent with their preference for the type of breeding habitat in which this study was conducted.

Heredity (2005) **94**, 247–257. doi:10.1038/sj.hdy.6800607
Published online 10 November 2004

Keywords: amphibia; assortative mating; hybrid zone; parentage inference; prezygotic isolation

Introduction

Prezygotic barriers to reproduction between taxa appear to evolve in a rather capricious manner (Butlin and Tregenza, 1998). They may arise on a surprisingly short time-scale and cause rapid speciation, especially when divergence of the mating system is driven by natural selection (Schluter, 2001) or sexual conflict (Rice, 1998). However, it is also possible, at the other extreme, for prezygotic barriers between two taxa to remain incomplete despite ecological differentiation and the accumulation of genetic incompatibilities, as seen in many hybrid zones (Harrison, 1993). Knowledge of the nature and strength of prezygotic barriers to gene flow is key to understanding differentiation between two taxa, and yet methods of estimating the magnitude of prezygotic isolation in the wild are poorly developed. Here, we present two novel approaches to estimate the strength of assortative mating in natural hybrid populations from genotypic data at marker loci. We illustrate these methods with a data set from the *Bombina* hybrid zone in Croatia.

Depending on the biology of the particular hybridising taxa, it may be possible to infer levels of assortment from field observations of mated pairs (eg in insects: McLain, 1985; Hewitt *et al*, 1987; snails: Johannesson *et al*, 1995; Ribi and Oertli, 2000; and birds: Moore, 1987; Grant and Grant, 1992; Sætre *et al*, 1997), or from the data of one (usually the maternal) parent and its offspring (Hewitt *et al*, 1987; Mallet *et al*, 1998). However, these sampling regimes are often impractical, for example, in the case of sporadic reproduction throughout an extended breeding season and external fertilisation without parental care. Nonetheless, as we demonstrate in this paper, information about mating patterns can still be extracted from genotypes of field-collected offspring (eg egg batches) and from adults that are associated with a given breeding site. Our approach is akin to genetic methods of paternity assignment, but differs from these in that the populations under study may deviate from Hardy–Weinberg and/or linkage equilibrium. It also extends the study of assortative mating in the field to hybrid populations that contain a wide range of recombinants rather than only a few genotypic classes such as parentals, F1s and backcrosses (eg Rieseberg *et al*, 1995; Goodman *et al*, 1999).

Our methods are intermediate between two distinct approaches to the analysis of offspring genetic data. The first, which also require data on samples of adults, are the techniques for assigning parentage (reviewed by Jones and Ardren, 2003) either categorically (eg Sanchez-Cristobal and Chevalet, 1997; Marshall *et al*, 1998) or

Correspondence: B Nürnberger, Department Biologie II, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany.

E-mail: nurnbb@zi.biologie.uni-muenchen.de

³Current address: Zoology Department, University of British Columbia, Vancouver BC, Canada V6T 1Z4

Received 3 February 2004; accepted 13 September 2004; published online 10 November 2004

fractionally, for example, to infer the relative reproductive success of a particular subset of the adult population (Morgan and Conner, 2001; Nielsen *et al*, 2001). One can also use nongenetic (eg behavioural) data to estimate prior probabilities of paternity given known mother-offspring pairs (Neff *et al*, 2001), or in cases where neither parent is known (see also Marshall *et al*, 1998). The second approach involves estimating how many parents contributed to a given batch of offspring in the absence of an adult sample (Emery *et al*, 2001) and, more generally, clustering samples of independent offspring into sibships (Lynch and Ritland, 1999; Thomas and Hill, 2000; Smith *et al*, 2001).

The development of our approach was motivated by our work on the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. During about four million years of independent evolution, these taxa have undergone substantial ecological and molecular divergence (Nei's genetic distance: 0.37–0.59; Szymura, 1993). Nevertheless, a wide spectrum of fertile hybrids is found in typically narrow hybrid zones, wherever their parapatric ranges adjoin in Central and Eastern Europe. The hybrid zone is stabilised in part by endogenous selection against hybrids (Kruuk *et al*, 1999) and, presumably, also by selection against toads in the wrong habitat: *B. bombina* typically reproduce in semi-permanent ponds in the lowlands, whereas *B. variegata* lay their eggs in ephemeral water bodies at higher elevations. The mating systems of the pure types differ in a number of features, including the size and structure of male aggregations and the male call (Lörcher, 1969). Thus, in principle, the stage is set for prezygotic barriers to gene flow.

The differential use of breeding habitat carries over into the hybrid zone, where even adults of mixed ancestry show a habitat preference that correlates with their individual hybrid index, with more *B. bombina*-like hybrids being more likely to be found in ponds and *B. variegata*-like hybrids in puddles (MacCallum *et al*, 1998). This results in partial assortative mating by habitat. Its effect on the genetic structure of hybrid populations is seen most clearly in the transect near Peščenica, Croatia, which has at its centre a mosaic distribution of ponds and puddles. Most toads are found in the sites that are predicted from their marker genotype, but individuals in the 'wrong' habitat inflate linkage disequilibrium (D) and heterozygote deficit (F) locally (Kruuk, 1997; MacCallum *et al*, 1998). But, judging by the difference in allele frequency between habitats, this effect can explain only part of the observed heterozygote deficit. The remainder may be due to nonrandom mating within sites (Kruuk, 1997; MacCallum *et al*, 1998). Here, we want to estimate the net contribution of the divergent mating systems of the pure taxa to prezygotic isolation in the *Bombina* hybrid zone. Laboratory experiments, for example, on female preferences for homotypic male calls, would complement our study and might reveal the traits that mediate any effect seen at the population level (Doherty and Gerhardt, 1984; see also Smajda *et al*, 2004).

Our approach to inferring the mating pattern starts with samples of field-collected egg batches consisting of full sibs. Given a sufficiently large offspring sample, we can accurately reconstruct the parental genotypes at a given diagnostic marker locus: the occurrence of only 'bv' and 'vv' genotypes in the offspring (where b stands

for alleles of *bombina* and v for those of *variegata* origin) implies that one parent was a heterozygote and the other a *B. variegata* homozygote. Next, the parental genotypes inferred at each marker locus are listed across loci. For example, we might infer the joint parental genotype ((bb,bv),(bb,vv),(bv,bv)) for three loci. But, from these data, we cannot determine the multilocus genotype of either parent: in this example, the parents might be ((bb,bb,bv),(bv,vv,bv)) or ((bb,vv,bv),(bv,bb,bv)).

Under ideal conditions of exhaustive sampling of adults and high genotypic variance, there may be only one matching male–female pair to which parentage of the clutch can be assigned. Often, however, limited allelic diversity at the marker loci and/or unsampled adults will render this approach impractical (Marshall *et al*, 1998). Nevertheless, the adult sample does contain information that we can use to determine the composition of the population of mated pairs. Although we may not be able to determine the precise genotypes of the parents, or the full set of genotype frequencies in the parental population, we can estimate parameters such as the mean allele frequency, heterozygosity, and linkage disequilibria. We use two complementary methods. First, we estimate the multilocus genotype frequencies in the population of adults, under the simplifying assumption that it was formed by the mixing of two differentiated populations (Barton, 2000). We use this information to fit a model to the offspring data, which jointly specifies the most likely parental genotypes for each sampled clutch and the maximum likelihood degree of assortative mating for the whole sample. Second, we estimate the composition of the parental population solely from the offspring data, using an extension of the method set out by Barton (2000). This approach provides information about the degree of assortment from levels of correlations between the allelic states of genes in mother and father. Both approaches are outlined below (see Materials and methods) and presented in more detail in Barton (2000) and in Supplementary Appendix II. We apply these methods to a hybrid population from the Peščenica transect in Croatia and demonstrate that they provide surprisingly robust estimates of assortment even for a relatively small data set. Mathematica (Wolfram, 1996) notebooks which implement these methods are available from www.helios.bto.ed.ac.uk/evolgen/.

Materials and methods

Field collections

This study concentrates on a hybrid population (site no. 103) in the Peščenica transect, which is located 20 km south-east of Zagreb, Croatia (see MacCallum *et al* (1998) for a detailed description). In this area, the distance across the hybrid zone between pure *B. bombina* populations on arable flood plains and pure *B. variegata* populations in low, forested hills was approximately 10 km. Site 103 is located on the *B. variegata* side 1.34 km from the estimated centre of the cline (MacCallum, 1994). It is a large, shallow puddle (7.0 × 0.7 m², depth: 0.1 m) that formed on compacted soil on a logging track in lowland forest, and has little aquatic vegetation. Egg clutches and adults were sampled during 10 visits between 29 April and 11 June 1995. Toads were anaesthetised in 0.2% MS222 (3-amino benzoic acid ethyl

ester, Sigma) before a single toe was removed as a tissue sample and immediately transferred to liquid nitrogen. Each individual was given a code in which the first digit refers to the year of first capture (eg '2' = 1992) and the next three digits represent the site code, followed by a running index (eg 3103.5 would be a toad first seen in 1993 at site 103, being the fifth individual in that year's sample from site 103). To maximise the number of families sampled and to minimise disturbance to the populations, egg collections were made on different dates and from batches several metres apart. Whenever possible, 12 eggs from each batch were taken, but some batches contained smaller offspring numbers, for example, after the removal of unfertilised eggs. Batches from which eggs had been removed were flagged to prevent resampling on subsequent visits.

Eggs were transported to the field laboratory, and tadpoles were reared to the age of 3 weeks before being anaesthetised and preserved in liquid nitrogen. This procedure was chosen to allow for both allozyme and PCR-based genotyping. Rearing was carried out at low density and with *ad libitum* food to minimise any confounding effects of selection.

Molecular analyses

Upon return to the laboratory, the samples were stored in a -60°C freezer prior to processing. DNA extraction followed standard protocols. Briefly, the tissue samples were digested overnight with Proteinase K. After centrifugation, the supernatant was extracted once with chloroform. DNA was precipitated in two volumes of ethanol, washed once in 70% ethanol, air-dried and then resuspended in ultrapure water (Merck). Stock solutions were stored at -20°C .

The genetic analysis was based on one microsatellite (*Bv12.19*), one simple length polymorphism (*Bb6.14*), and four SSCP (*Bv24.11*, *Bb3.34*, *Bb3.36*, *Bb7.4*). For the present analysis, we categorise all alleles at these diagnostic, codominant markers as either *b* or *v*, and thus disregard intraspecific variation (1–3 alleles per taxon and locus). The loci are part of a larger set of mapped markers in the *Bombina* genome (Nürnberger et al, 2003). They are unlinked except for one locus pair (*Bv24.11* and *Bb3.34*), which shows weak linkage ($r = 0.36$), and one tightly linked pair (*Bb7.4* and *Bb6.14*, $r = 0.06$). We treat the former pair as unlinked and take linkage into account as required in the latter pair. Given the sensitivity of the statistical analysis to even single isolated errors, the entire data set was assembled twice from independent dilutions of the DNA stocks. Locus/sample combinations that failed to give consistent and unambiguous results even after additional amplifications were left blank.

PCR reactions were set up in a total volume of 30 μl with 50–100 ng template DNA, 50 mM KCl, 10 mM Tris (pH 9.0 at RT), dNTPs (0.2 mM per nucleotide), 10 pmol of each primer, and 0.5 U Taq polymerase (rTaq, Amersham Biosciences). MgCl_2 concentrations varied among loci between 1.5 and 2.5 mM. Amplification was carried out on a Hybaid Touchdown thermocycler with oil overlay. After initial denaturation for 3 min at 94°C , the cycling profile was as follows: 15 s at 94°C , 30 s at $x^{\circ}\text{C}$ and 30 s at 72°C for 32–35 cycles, where x is the locus-specific annealing temperature (53 – 59°C).

The PCR products were separated on polyacrylamide gels and visualised with silver staining. Length polymorphisms were separated in vertical denaturing gels (acryl-bis ratio 19:1) in $1 \times$ TBE buffer at constant voltage. The SSCPs were electrophoresed on native horizontal polyacrylamide gels (acryl-bis ratio 37.5:1) at constant low temperature and constant voltage for 3–4 h (MultiPhor gel rigs, Amersham Biosciences) with an electrode buffer of $2 \times$ TBE and a gel buffer of $1 \times$ Tris-acetate. The locus-specific parameters are available upon request from BN.

Statistical analysis

We apply two methods of inferring the mating pattern, both of which are based on correlations of allelic states within and between individuals or, more precisely, on the moments of the genotypic distribution. Within individuals, correlations between alleles either at a single locus (heterozygote deficit or excess) or at two loci inherited from the same parent (gametic disequilibrium) are examples of second-order moments. Higher-order moments represent correlations between alleles at more than two loci and may exist within and across the haploid (ie maternal and paternal) genomes of a diploid. When correlations arise from the mixing of two differentiated populations, all moments $c_{j,k}$ involving j loci from the maternal and k loci from the paternal genome are expected to be the same, when scaled relative to the differences in allele frequencies between the source populations. This assumption greatly reduces the number of parameters that need to be estimated (Barton, 2000).

We extend this approach to allelic correlations within mated pairs. They are of the form $c_{i,j,k,l}$, where the subscripts indicate the number of loci in the female (i,j) and the male (k,l) of a pair. For example, terms of the type $c_{0,0,0,2}$ represent the standard pairwise linkage disequilibrium and those of type $c_{0,0,1,1}$ the heterozygote deficit ($= Fpq$, where F is the inbreeding coefficient and p and q are frequencies of the *B. variegata* and *B. bombina* alleles, resp.). Terms of the type $c_{0,1,0,1}$ represent the covariance between the state of a gene in the mother and a gene in the father, or, in other words, the covariance between the proportion of *B. variegata* genes in each parent. Other moments give more complex associations between the two parents – for example, between the level of heterozygosity in one and of linkage disequilibrium in the other (such as $c_{1,1,0,2}$). We do not attempt to estimate all these parameters: only a few lower order coefficients are of interest.

In our first approach, the aim is to find the likelihood of a model of mating patterns, given the observed offspring sample, and given an estimate of the composition of the adult population. Direct estimation of adult genotype frequencies from the data is not an option, because the number of possible genotypes is very large (729 for six biallelic loci), such that most of them are unobserved. Generally, the expected frequency of any multilocus marker genotype is a function of the allele frequencies and of the linkage disequilibria in the population. Note that, for a population in Hardy–Weinberg and linkage equilibrium, the computation is greatly simplified, as all correlations are zero and expected multilocus frequencies are just products of the

allele frequencies. In the present context of a hybrid population, however, reliable estimates are only possible if disequilibria including those of higher order are taken into account, even though most of them will be small in magnitude. They can be obtained from a sample of adults at the breeding site, and, in the present study, were estimated by maximum likelihood, assuming a model of population admixture (Barton, 2000).

The second method infers correlations in allelic state in the parents from the moments in the sibships alone, that is, without recourse to the adult sample. While we remain ignorant about the genotypes of any particular parent pair, we can draw inferences about the correlation structure within parent genotypes and, critically, between mated adults. Consider the six possible mating types for a biallelic locus, whose genotypes we code by the number of *B. variegata* alleles (0, 1 or 2) in each parent. For example, a mating between a heterozygote and a *B. variegata* homozygote (denoted by (1,2)) produces '1' and '2' offspring in equal proportion (Figure 1). Note that we conflate (1,2) and (2,1) since these are indistinguishable. The variance within sibships informs us about the heterozygosity in the parents, since only heterozygotes can generate variance within families (horizontal axis in Figure 1). The heterozygosity in the offspring is inversely related to the correlation between parental scores (vertical axis).

More generally, we make use of the fact that the moments among offspring are given by a linear combination of the moments among genes in the parents. For example, if *b* and *v* alleles are coded as 0 and 1, respectively, the variance within families depends linearly on the variance in the population of parents, and on the covariance between parents. We are mainly interested in the parental heterozygosity (*H*) and pairwise linkage disequilibrium (*D*) and in correlations between the parents in the hybrid index (HI), which we define as the number of *B. variegata* alleles across all marker loci. These latter correlations provide estimates of

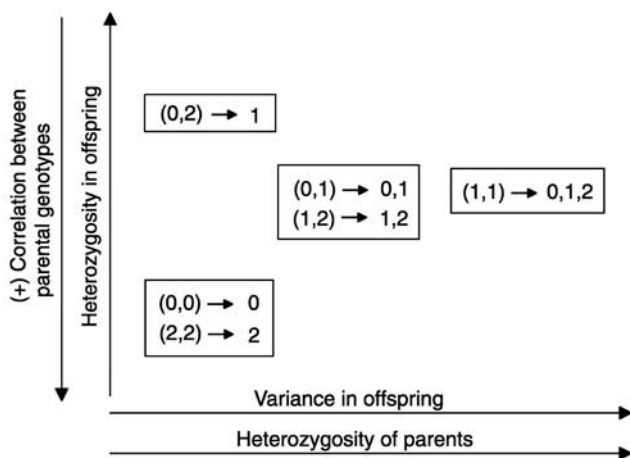


Figure 1 Correlations between offspring and parent genotypes. The six possible mating types at a biallelic locus are ranked in terms of the proportion of heterozygous offspring and the allelic variance within sibships. For example, two heterozygous parents (mating type (1,1)) produce all three possible genotypes among their offspring ($\rightarrow 0,1,2$). Boxed mating types behave identically. Note that nondistinguishable mating types (eg (1,2) and (2,1)) were conflated. See text for the numeric coding of genotypes.

assortment. A full account of the computations is provided in Supplementary Appendix II and at <http://helios.bto.ed.ac.uk/evolgen/barton/index.html>.

Results

The adult sample

The adult sample collected in 1995 from site 103 consisted of 23 toads. Most of these individuals were seen only once. However, some individuals were recaptured, and one male in particular was seen on most sampling dates (individual 2103.10). Genotypes are available for 20 adult individuals and gave a mean frequency of *B. variegata* alleles, \bar{p} , of 0.31. As it became clear that the pool of adults contributing to the egg clutches was much larger than the set of sampled adults (see below), we added the genotypes of another adult sample ($n=8$, $\bar{p}=0.27$) collected in 1995 at the nearest ecologically similar site (183). Site 183 is located at a distance of 800 m from site 103 (ie within the estimated annual adult dispersal range of maximally 1.5 km) and at the same distance from the centre of the hybrid zone as site 103 (MacCallum, 1994). In the following, we base our analyses on the pooled set of 28 genotyped adults (Table 1).

The distribution of the HI, defined as the sum of *B. variegata* alleles across loci per individual, is bimodal. *B. bombina*-like individuals predominate. Eight of them have a HI of zero, whereas no individual has a pure *B. variegata* marker genotype (HI=12, Figure 2a). The number of heterozygous loci per individual has a bimodal distribution as well (Figure 2b). Here, the right mode indicates matings between opposite genotypes that are brought into spatial proximity in this mosaic hybrid zone and produce highly heterozygous offspring. The structure of the adult population is characterised by a strong deficit of heterozygotes ($F=0.342$, Table 2) and strong pairwise linkage disequilibria (standardised $R = D/\sqrt{p_1q_1p_2q_2} = 0.608$). Owing to the statistical associations among loci, it is difficult to test for significant heterogeneity among pairwise estimates of *R*. On the whole, they are broadly similar and, surprisingly, even the tightly linked pair *Bb6.14/Bb7.4* does not stand out from the rest. In all these measures, there are no appreciable differences between the sexes.

The egg sample

We genotyped 168 tadpoles from 18 batches of eggs (Supplementary Appendix I). As a first step, some checks on the data set are required. Given the oviposition habits of *Bombina*, two problems might arise: (a) an apparently homogeneous batch of eggs may consist of two separate sibships and (b) a given pair may distribute its offspring in several batches across the site. We investigate these possibilities at the appropriate stages of the analysis.

Especially in sparsely vegetated temporary habitat, it is possible that more than one female attaches its eggs to a given plant stem. This may not be apparent in the field, but can be inferred from the genotypic data (Vines and Barton, 2003). For unlinked loci, we expect no correlations between genotypes at two different loci within a family. Even for two linked loci, a correlation can only arise when at least one of the parents is heterozygous at both loci. However, mixing of families does generate correlations, which increase in strength the more disparate the genotypic compositions of the two families

Table 1 Genotypes of the pooled adult sample (sites 103 and 183) at six molecular loci

	<i>Toad ID</i>	<i>Bb3.34</i>	<i>Bb6.14</i>	<i>Bb7.4</i>	<i>Bv12.19</i>	<i>Bb3.36</i>	<i>Bv24.11</i>	
Males	5103.1	1	0	0	0	0	0	
	5103.3	0	0	1	0	1	0	
	5103.5	—	—	1	—	—	0	
	5103.6	0	0	0	0	0	0	
	5103.8	0	0	0	0	0	0	
	2103.10	2	1	2	1	2	2	
	5103.14	2	2	2	2	2	1	
	5103.17	0	0	1	0	0	0	
	5103.20	1	1	1	1	1	0	
	4183.4	1	1	1	2	1	1	
	4183.5	1	1	1	2	1	1	
	5183.1	0	0	0	0	0	0	
	5183.2	0	0	0	0	0	0	
	5183.4	0	0	0	0	0	0	
	5183.5	0	0	0	1	0	0	
	Females	5103.4	2	1	1	1	1	2
		5103.7	0	0	0	0	0	0
5103.10		1	1	0	0	1	0	
5103.11		0	0	1	0	0	0	
5103.12		2	1	2	0	2	2	
5103.13		0	0	0	0	0	0	
5103.15		0	0	0	0	1	0	
5103.16		0	0	0	0	0	0	
5103.18		2	1	2	2	2	1	
5103.19		1	1	2	1	1	2	
5103.21		0	0	1	0	0	0	
5183.6		2	0	1	2	2	2	
5183.7		0	0	0	1	1	0	

Listed are the number of *B. variegata* alleles per locus. Dashes indicate missing data. See methods for an explanation of the individual code numbers.

are. We use two measures to investigate mixing: the variance in the hybrid index (ie the sum of pairwise covariances across loci) and the squared covariance between loci, summed over all pairs of loci. The former quantity should reveal mixing of genotypically different families, whereas the latter can detect mixing even when the two families have a similar overall hybrid index. In order to obtain null distributions, the observed genotypes at each locus were randomised 1000 times among the offspring of a family. The tightly linked pair of loci was excluded.

Batches 5, 8 and 18 showed significant correlations based on the variance in the hybrid index (adjusted for multiple tests; Supplementary Appendix I). The squared covariance method identifies only batch 5 as mixed, because the single aberrant genotypes in batch 8 and 18 only generate rather small covariances. We conclude from these tests that the egg batches 5, 8 and 18 are in fact mixtures of different sibships. In the following, we omit the single outliers from batches 8 and 18 and split batch 5 into two (5a and b).

If mixing does occur, it may go unnoticed: in the extreme case, a mixing of families with the same parental genotypes would necessarily be undetectable. Following Vines and Barton (2003), we investigated the power to detect mixed batches by simulations of random mating, given the genotypic composition of the adult population (see below). Among 1000 simulated, randomly joined batches of 5 + 5 eggs, only a little more than 50% of them were detected as mixed by the covariance method, which is most efficient in this case. This moderate resolution, however, does not invalidate our approach. The batches

of greatest concern to us are mixtures of *B. bombina*-like and *B. variegata*-like matings, because they lead to an erroneous over-representation of hybrids in the inferred parent pool. Such incidences of mixing were relatively much better detected in the simulations. Moreover, assortative mating would tend to generate matings that are either rather similar or rather different, which increases the detection rate of ‘troublesome’ incidents of mixing relative to our random mating simulation. From that simulation, we determined the net effect of undetected mixed batches on the inferred parent pool by comparing sets of entirely unmixed batches with sets of batches of which 10% are undetected-mixed (Table 3). As expected, mixing increases the proportion of hybrids among the inferred parents, here mostly at the expense of the (0,0) class of matings, but the effect is moderate.

Joint parental genotypes

As a first step towards inferring the parents, we determine the most likely set of parental genotypes per locus for all families (Table 4). This categorical assignment will, however, not be without error. Recall our earlier example of a sibship that yields for a given locus a mixture of heterozygotes (coded as 1, according to the number of *B. variegata* alleles) and *B. variegata* homozygotes (2). This segregation implies that one parent was a heterozygote and the other a *B. variegata* homozygote (mating type: 1,2). If a single ‘0’ score was added, the parental assignment would change to (1,1). The probability that not all possible offspring genotypes were sampled for a given batch is the highest for (1,1) matings,

as they generate the most diverse set of offspring. For the average batch size of nine and assuming Mendelian ratios, there is a $2 \times (3/4)^9 = 15\%$ chance that a (1,1) mating will be misclassified as either (0,1) or (1,2),

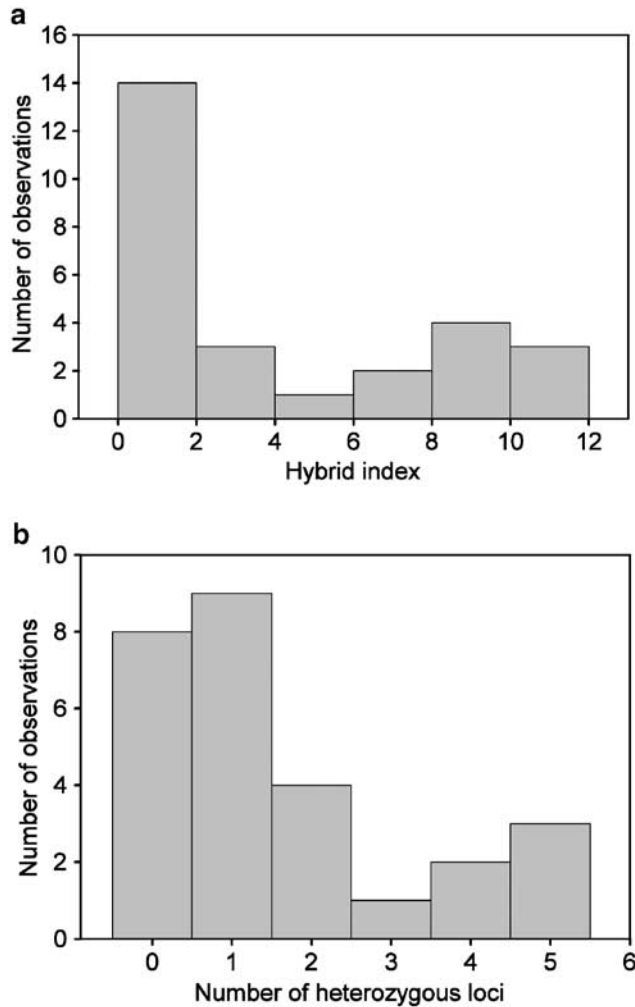


Figure 2 Genetic composition of the adult sample. Shown is the distribution of (a) the hybrid index (= the sum of *B. variegata* alleles across loci per individual) and (b) the number of heterozygous loci per individual. In plot (a), the data have been binned in groups of two: (0,1), (2,3), etc. The poorly characterised individual 5103.5 has been excluded from these plots.

because one of the two homozygotes was not observed. The only other source of error comes from (0,1) and (1,2) matings. In both cases, the probability that one of the two expected genotypes is not observed equals $2 \times (0.5)^9 = 0.4\%$. All of these misclassifications reduce the estimated heterozygosity of the parents in this preliminary assignment (see below). Note, however, that the mating model below explicitly deals with these uncertainties and takes the exact sample sizes per batch into account.

As highlighted in Table 4, there are several pairs and one triplet of batches with identical joint parental genotypes (1+2, 3+4, 7–9, 11+12, 13+14). While batches from three of these sets were sampled on the same date, samples from the remaining two sets were collected 2 days apart (cf. Supplementary Appendix I), and the later collections were at a more advanced developmental stage. These observations suggest that the eggs of some pairs were distributed in several batches. The chance occurrence of repeated joint parental genotypes is indeed unlikely: in 1000 simulations of 18 pairs drawn at random from the estimated adult population, two and four sets of identical joint parental genotypes were observed in less than 17 and 1.1% of the replicates, respectively. We therefore merge the batches within each set into a single family and arrive at a total of 13 separate families. This procedure is conservative, because we end up with the smallest possible number of matings, even though some egg batches with identical joint parental genotypes might actually represent distinct families.

Across the entire offspring data set, there are 10 cases of genotype scores which, if changed, would alter the joint parental genotype of the respective egg batch (highlighted in bold in Supplementary Appendix I). These singletons persisted after additional laboratory tests. As expected, they tend to occur in small clutches. Overall, the frequency of singletons is not greater than expected from Mendelian segregation ($P = 0.033$, for *bv*vv* crosses, but $P > 0.2$ for *bv*bv* and *bv*bb*). As a final check on the data, we inspected the segregation ratios based on the inferred mating type per family and locus. Out of 18 tests (ie six loci, and three kinds of segregating cross), only one significant deviation was found (*b:v* segregation = 18:5, $G = 9.63$, $P = 0.002$) for locus *Bv12.19* and segregations *bv*vv*. Given the large number of tests, we do not consider these significant overall.

Table 2 Mean *B. variegata* allele frequencies (\bar{p}), heterozygote deficit (F) and standardised pairwise linkage disequilibria ($R_{ij} = D_{ij} / \sqrt{p_i q_i p_j q_j}$) for the adult sample

Locus	\bar{p}	F	R_{ij}				
			<i>Bb3.34</i>	<i>Bb6.14</i>	<i>Bb7.4</i>	<i>Bv12.19</i>	<i>Bb3.36</i>
<i>Bb3.34</i>	0.333	0.500*					
<i>Bb6.14</i>	0.204	-0.027	0.715*				
<i>Bb7.4</i>	0.357	0.222	0.619*	0.574*			
<i>Bv12.19</i>	0.296	0.467*	0.563*	0.559*	0.436		
<i>Bb3.36</i>	0.352	0.269	0.790*	0.686*	0.660*	0.610*	
<i>Bv24.11</i>	0.250	0.619*	0.732*	0.520*	0.596*	0.445*	0.611*
Average	0.299	0.342	0.608				

Significant ($\Delta \log(L) > 2$) estimates are indicated by the asterisk.

Table 3 Effect of undetected mixed egg batches on the inference of parental genotypes

Genotypes	Unmixed	10% mixed
0,0	33.4	25.0
0,1	42.2	45.8
0,2	19.2	19.8
1,1	11.6	13.4
1,2	11.8	13.6
2,2	1.8	2.4
0	128.2	115.6
1	77.2	86.2
2	34.6	38.2

Shown are the results of two types of simulations: (a) entirely unmixed batches and (b) 10% of undetected mixtures (5+5 eggs) among the batches. The data are based on 100 simulations of 20 batches each. Assuming random mating, the parents for each family were drawn at random from a simulated population with genotype frequencies, as estimated from the actual adult sample (computations based on the method of moments, see text for details). The upper part of the table shows the average representation of each mating type among the inferred parents (in total: 20 batches × 6 loci = 120 observations). The lower part shows the same data per inferred individual parent genotype.

Table 4 Joint parental genotypes inferred from 19 egg batches (after the splitting of batch 5 and removal of single outliers from batches 8 and 18)

Batch	Bb3.34	Bb6.14	Bb7.4	Bv12.19	Bb3.36	Bv24.11
1	1,2	1,1	1,2	0,2	0,2	0,2
2	1,2	1,1	1,2	0,2	0,2	0,2
3	0,2	0,1	0,1	0,1	0,1	0,1
4	0,2	0,1	0,1	0,1	0,1	0,1
5a	0,1	0,1	0,1	0,1	0,1	0,1
5b	1,2	0,1	2,2	1,2	1,1	2,2
6	2,2	1,2	1,2	0,2	1,2	1,2
7	0,1	0,0	0,0	0,0	0,0	0,0
8	0,1	0,0	0,0	0,0	0,0	0,0
9	0,1	0,0	0,0	0,0	0,0	0,0
10	2,2	1,2	1,2	1,2	1,2	1,2
11	2,2	1,1	1,2	0,2	1,2	1,2
12	2,2	1,1	1,2	0,2	1,2	1,2
13	0,2	0,1	0,2	0,2	1,2	0,2
14	0,2	0,1	0,2	0,2	1,2	0,2
15	0,2	1,1	0,2	1,2	1,2	0,2
16	0,2	1,1	1,1	0,0	0,2	1,2
17	2,2	2,2	2,2	2,2	1,2	1,2
18	1,2	1,1	0,2	0,2	1,2	0,2

See text for further explanation. Batches with identical joint parental genotypes are enclosed by dashed lines.

Closer examination of the joint parental genotypes reveals that they are not a random subset of the adult population (Table 5). While the heterozygote deficit is very similar in parents and adults ($F = 0.34$ versus 0.36, respectively), the mean frequency of *B. variegata* alleles is higher in the parents (0.55 versus 0.30). This difference is highly significant, because 1000 sets of 26 individuals that had been randomly sampled from a simulated adult population with the same genotype distribution as the observed sample yielded $\bar{p} = 0.31$ (standard error: 0.002). It appears that *B. variegata*-like individuals have a higher propensity to mate at this site. We now turn to a more

detailed statistical examination of mating patterns, which takes such associations into account.

Inferring the structure of the adult population

Only three families could be produced by pairs from our adult sample. Moreover, the sampling dates for most of these adults do not agree with those of the clutches. The high proportion of families for which we do not find matching parents suggests a large pool of unsampled adults and rules out direct assignment of parentage to the sampled adults. The problem of unidentified parents persists if we conduct the search more rigorously with a likelihood approach that takes the potential sampling error in the list of joint parental genotypes (Table 4, see above) into account.

As described above, the central difficulty in analysing these family data is that the correct split of the joint parental genotype into a maternal and a paternal component is unknown. Nevertheless, we can determine the most likely parental genotypes of a family from their expected frequency in the adult population and from the observed genotype distribution in the egg sample. To this end, the sample of adult genotypes was used to obtain estimates of allele frequencies and allelic correlations (see Statistics section above). The marked effect of these correlations on the distribution of the hybrid index is illustrated in Figure 3.

The mating model

An estimate of assortative mating was derived using the following model (after Lande, 1981). The probability of observing our families is

$$L[X, Y|F, \alpha, \beta, g] = C^{-n} \prod_n \sum_X \sum_Y g[X]g[Y]P[F_i|X, Y] \times \exp[\alpha(Z_X - \bar{Z})(Z_Y - \bar{Z})] \times \exp[\beta(Z_X - \bar{Z})] \exp[(Z_Y - \bar{Z})],$$

where $g[X]$ and $g[Y]$ are the frequencies of the adult genotypes X and Y in the population, based on the estimated moments. $P[F_i|X, Y]$ gives the probability that these individuals produced family F_i , given their genotype, Mendelian segregation, and the sizes of the egg samples. Z_X and Z_Y represent values of an additive trait (here: the hybrid index) of the genotypes X and Y , respectively, \bar{Z} is the corresponding population mean. The strength of assortment is measured by α , and β represents the bias towards *B. variegata*-like genotypes in the parents relative to the adult sample. We sum over all potential parent pairs per family, and multiply across the 13 families.

The maximum likelihood estimates are $\alpha = 0.007$ (support limits: -0.031 – 0.046) and $\beta = 0.23$ (support limits: 0.114 – 0.368). Thus, there is no evidence of assortative mating, but a strong bias towards *B. variegata*. Figure 4 illustrates these results. We can check the fit of the model by comparing the correspondence between mating types that are predicted from this model with those that are directly inferred from the offspring data. Simulation of pairs using the parameters of this model gives a distribution that is very close to the inferred mating types (Table 6). The greatest discrepancies are seen for mating types (0,2) and (1,1). Their relative over- and under-representation, respectively, in the joint parental genotypes follows the pattern that is predicted for errors in the categorical assignment (s.a.). Note

Table 5 Comparison of adult and inferred parental genotypes

	Genotype/mating type	Bb3.34	Bb6.14	Bb7.4	Bv12.19	Bb3.36	Bv24.11	Total
Adults (N = 27 ^a)	0	15	17	13	16	13	18	92
	1	6	9	9	6	9	4	43
	2	6	1	5	5	5	5	27
Parents (N = 26)	0,0	0	1	1	2	1	1	6
	0,1	2	4	2	2	2	2	14
	0,2	4	0	3	5	2	4	18
	1,1	0	5	1	0	1	0	7
	1,2	3	2	4	3	7	5	24
	2,2	4	1	2	1	0	1	9
	0	6	6	7	11	6	8	44
	1	5	16	8	5	11	7	52
	2	15	4	11	10	9	11	60

The genotypic distribution for both groups is given for each locus and also summed across loci. Data were compiled as for Table 3.

^aThe poorly characterised adult 5103.5 is excluded from the summation of adult genotypes.

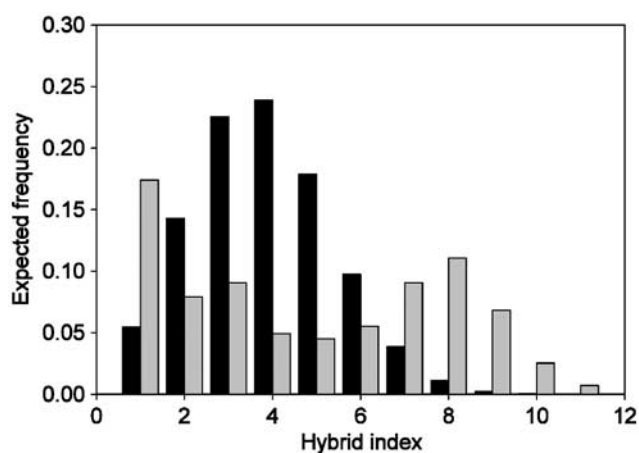


Figure 3 Expected distribution of the hybrid index in adults. Expected genotype frequencies were computed (a) from the allele frequencies at the six marker loci only (ie $R=0$ and $F=0$, black bars) and (b) by taking the estimated moments into account (grey bars). These represent associations among alleles at one or more loci.

that the mean allele frequencies agree closely (0.53 versus 0.55).

Estimating assortment solely from the offspring genotypes

We now turn to our second approach and estimate the degree of assortment from the moments in the offspring sample, without reference to the adults. The results are shown in Table 7, together with the means and standard deviations from simulations. For the latter, 1000 replicate sets of 13 families with the same sizes as in our sample were generated from different base populations: first, one in Hardy–Weinberg proportions and linkage equilibrium, and, second, one with allele frequency, heterozygote deficit, and linkage disequilibrium equal to that estimated here from the offspring genotypes by the method of moments (ie $\bar{p} = 0.542$, $F = 0.263$, $D = 0.191$). For the latter, genotype frequencies were constructed by assuming a hypergeometric distribution for the number of *B. variegata* alleles in gametes, with variance determined by the linkage disequilibrium, and then combining a fraction $1-F$ of gametes randomly, and

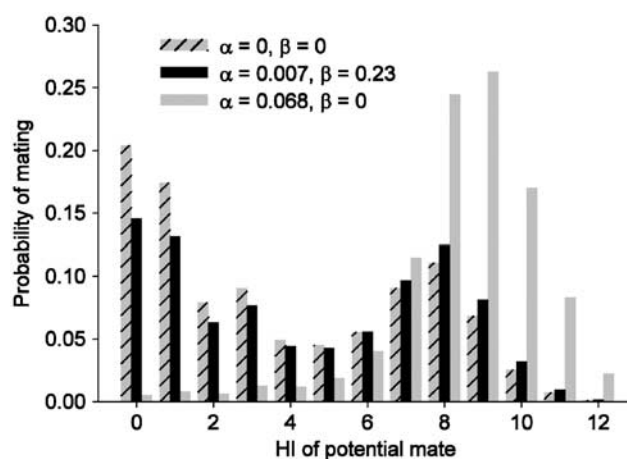


Figure 4 The distribution of mating probabilities for a hypothetical pure *B. variegata* individual (HI = 12). Shown are results for our study population according to the mating model (see text for details). The hatched, grey bars represent the case of random mating ($\alpha=0$) and equal mating propensity for all adults ($\beta=0$). The maximum likelihood parameters ($\alpha=0.007$, $\beta=0.23$) are illustrated by black bars. Note that α in this case does not differ significantly from zero. The plain grey bars show the model fit that ignores any variation in mating propensity (β set to 0) and so gives the erroneous impression of strong assortment ($\alpha=0.068$).

complementing a fraction F into entirely homozygous diploids.

As noted before, the allele frequency estimated for the breeding population was substantially higher than for the sampled adults. The heterozygote deficit in the breeding parents was similar in magnitude to that in the adults (Table 7: $F_{pq} = 0.065$ versus 0.074, standard deviation: 0.042). Linkage disequilibrium is estimated as being substantially stronger among the breeders than among adults ($D = 0.191$ versus 0.122). Estimates as low as 0.122 were observed in only 4.5% of simulated replicates, but, as with heterozygote deficit, it is hard to judge differences in linkage disequilibrium when allele frequencies also differ between breeders and sampled adults (standardised $R = D/\sqrt{p_1q_1p_2q_2} = 0.769$ and 0.581, respectively). Simulations of the null hypothesis of Hardy–Weinberg and linkage equilibrium in the parents show that the observed values of F and D in

the inferred parent population are highly significantly different from zero (Table 7).

The association between *B. variegata* alleles in each parent, $c_{0,1,0,1}$, is small in magnitude, as are all the other associations between parents that reflect nonrandom mating. The standard deviation of $c_{0,1,0,1}$ is 0.034, which corresponds to a standard deviation in the correlation of $0.034/pq = 0.137$. This shows that the method has the power to detect moderate correlations between mates even with only 13 families. On the whole, only one of the cross-parent correlations is notable: the association between linkage disequilibrium in the mother and in the father (D^*D) is 0.013, a value exceeded in only 2.9% of the replicates. However, there are six such tests, so that the estimate is not significant overall.

We note that, for all estimators shown in Table 7, the simulations show some bias, which in most cases is statistically significant. However, the bias is small in all cases relative to the standard deviation, and so is no cause for concern.

Table 6 Comparison of the mating types, parental genotypes and allele frequencies inferred (a) from the joint parental genotypes, (b) from the maximum likelihood mating model, and (c) from the same model with β set to zero

	(a) Joint parental genotypes	(b) Simulated parents $\alpha = 0.007$, $\beta = 0.23$	(c) Simulated parents $\alpha = 0.068$, $\beta = 0$
Mating types			
0,0	0.08	0.06	0.35
0,1	0.18	0.20	0.24
0,2	0.23	0.14	0.07
1,1	0.09	0.18	0.12
1,2	0.31	0.27	0.14
2,2	0.11	0.11	0.05
Genotypes			
0	0.28	0.24	0.51
1	0.33	0.42	0.32
2	0.38	0.32	0.16
\bar{p}	0.55	0.53	0.32

Data were compiled as for Table 3.

Discussion

In this paper, we present methods that allow inference of the mating pattern in hybrid populations to be inferred from genotypes of egg batches and adults at a number of diagnostic marker loci. The composition of the breeding population that produced the egg batches could then be compared with the sampled adults in two ways. First, parameters of an explicit model of mate choice were estimated, by calculating the probability that the egg batches would be observed given that model and estimates of adult genotype frequencies. Second, the associations among breeding parents were estimated based solely on a set of full sibships. This approach should prove useful in any hybrid population in which the actual parents cannot be efficiently sampled, such that no direct assignment of parentage is possible. As we have demonstrated, the genotypic information also allows us to account for complicating features of the organisms' biology, such as the mixing and dispersing of egg batches.

Both methods are quite general, and can be applied to populations of any structure, regardless of whether or not they are in linkage disequilibrium. However, as allelic correlations weaken, the variance of hybrid index decreases and so assortment becomes increasingly difficult to demonstrate. At the other extreme, with strong linkage disequilibria, the population is composed of just a few genotypic classes such as parentals, F1s, backcrosses and so on. Simpler methods that compare the observed frequencies of these categories with those expected under random mating might then be used to determine the strength of assortative mating. Our approach should therefore be most valuable in those hybrid populations with abundant recombinants that are nevertheless structured by linkage disequilibria and in which heterozygote deficit suggests a role for assortative mating.

The tests have surprisingly large power even with our small data set, as the support limits and standard deviations are relatively narrow. This is in part because each family is derived from a sample of four haploid genomes. Moreover, we combine information from several loci and base our inference about the mating

Table 7 Composition of the populations of adults and parents based on the method of moments

Parameter	Moments	Adults	Parents	Simulations $F = 0$, $D = 0$	Simulations $F = 0.263$, $D = 0.191$
Allele frequency	\bar{p}	0.299	0.542	0.541 (0.030)	0.543 (0.072)
Heterozygote deficit	$c_{0,1}$ or $c_{0,0,1,1}$	0.074	0.065	-0.004 (0.025)	0.057 (0.042)
Pairwise D	$c_{0,2}$ or $c_{0,0,0,2}$	0.122	0.191	0.022 (0.040)	0.176 (0.028)
HI*HI	$c_{0,1,0,1}$		-0.027	-0.006 (0.017)	-0.007 (0.034)
HI*H	$c_{0,1,1,1}$		-0.00725	0.0005 (0.0074)	0.0005 (0.014)
HI*D	$c_{0,1,0,2}$		-0.010	-0.0006 (0.0084)	0.0004 (0.0097)
H*H	$c_{1,1,1,1} - c_{0,0,1,1}^2$		-0.003	0.000 (0.010)	-0.002 (0.012)
H*D	$c_{0,2,1,1} - c_{0,0,0,2}c_{0,0,1,1}$		-0.011	0.000 (0.011)	-0.0003 (0.0067)
D*D	$c_{0,2,0,2} - c_{0,0,0,2}^2$		0.013	-0.0010 (0.0136)	-0.0018 (0.0072)

Moments of the type $c_{1,1}$ (or $c_{0,0,1,1}$ or $c_{1,1,0,0}$ in the extended notation) estimate the deficit of heterozygotes (corresponding to F_{pq}). Similarly, terms of the type $c_{0,2}$ (or $c_{0,0,0,2}$ or $c_{0,2,0,0}$ etc.) estimate pairwise linkage disequilibria (corresponding to D). The remaining rows give associations between sets of genes in the two different parents of a breeding pair, involving the individual hybrid index (HI), heterozygosity (H), and pairwise linkage disequilibrium. The significance of these estimates was determined from two sets of simulations of populations for which F and D were either set to zero or to the estimates of the parent population (last two columns: mean (SD)). See text for further details.

pattern on an aggregate measure of genotype, the hybrid index. These aspects contribute to the consistency of the results that both methods provide.

While pre-defined full-sib families are an efficient starting point to the inference of mating patterns, we showed (at the toads' prompting) that less than perfect information about family groups can be accommodated, because mixed egg batches generate detectable correlations in multilocus genotypes even in our minimally polymorphic markers. If a greater number of more variable markers is used, the power of detection is increased (Vines and Barton, 2003), and highly polymorphic markers offer the possibility to assemble sibships from genetic data alone (eg Thomas and Hill, 2000). Thus, our method of inferring patterns of parentage does not depend on recognisable sibling clusters in the field. Moreover while we have focussed on patterns of mating, the genetic composition of the offspring cohort could be affected by other processes as well, such as assortative fertilisations (Howard *et al*, 1998). Their impact could be quantified if, for example, independent mating observations in the field were available. As it stands, our method provides a means to estimate the net contribution of the mating system to the gene flow barrier in a hybrid zone.

In the present example of the Croatian *Bombina* hybrid population, both estimation methods provide no evidence of assortative mating, but do show a significantly higher mating propensity of individuals with a more *B. variegata*-like genotype. While our method provides similar estimates of heterozygote deficit in the adult and in the parent population, it suggests relatively stronger linkage disequilibrium among the parents. This implies that adults with more recombined genotypes are contributing less to the next generation than expected. As we mention above, it is difficult to assess the significance of the observed difference, since we are comparing estimates from populations with different mean allele frequencies.

The fact that we cannot identify potential parent pairs among the adults for most clutches suggests that only a relatively small proportion of the adults that visited site 5103 were sampled. Given the size and structure of this site, our capture efficiency of individuals known to be present during any one visit was likely to be at the high end of the range estimated previously for this transect (~30–80%, MacCallum, 1994). Our data are therefore in agreement with earlier analyses, which demonstrated substantial within-season turnover in the adult population per site (MacCallum *et al*, 1998). Individuals seen early in the season were consistently absent later on, and even towards the end of a 6-week observation period new individuals were detected, while still others were caught repeatedly throughout (MacCallum, 1994). At site 103, the average residence time was approximately 5 days, and the majority of individuals were caught only once. Movement into and out of sites together with a sizeable group of residents within a season were similarly observed in *B. variegata* populations in Switzerland (Barandun and Reyer, 1998). In a recent study on *B. variegata*-like hybrid populations in Romania, the number of toads that were intercepted overnight at fences around five sites was up to four times larger than the mean number sampled within the water during the daytime (Vines *et al*, in preparation). Taken together,

these data suggest that residence times are highly variable and, for a large fraction of the adult cohort per site, very short. Many of the parents that we missed at site 103 may have been such short-term visitors.

Our data indicate that the sample of adults at the study site was a mixture of two subsets of individuals: those that were accepting this habitat for reproduction and others that were simply passing through, perhaps on the way to more suitable breeding sites. Site 103 was a typical puddle, and the disproportionate contribution of *B. variegata*-like adults to the offspring cohort is in agreement with the documented habitat preference in this transect (MacCallum *et al*, 1998). The inferred parental genotypes covered a wide spectrum, such that moderately strong assortative mating should have been detectable if present: the SD of the estimated correlation is 0.137. Yet, our two methods of parentage inference do not provide evidence for any departure from random mating in hybrid populations, despite the anciently diverged mating systems of the two pure taxa. It also suggests that the observed high heterozygote deficit within local adult breeding aggregations in the Croatian transect is due to other causes, such as natural selection against hybrids. A more extensive analysis of assortative mating in hybrid *Bombina* populations is currently under way in our current study area near Cluj, Romania (Vines *et al*, in preparation) to investigate the generality of our findings across a wider range of breeding habitats.

Acknowledgements

We thank J Gilchrist for assistance in the field, and the Perović family for their hospitality in Pešćenica. B Förg-Brey and G Praetzel helped with the development of efficient protocols for the analysis of our newly developed molecular markers. We thank the two reviewers and, in particular, John Brookfield for detailed comments that prompted significant improvements of analysis and presentation. This work was funded by grants from NERC to NB, NERC studentships to LK and TV, the Royal Society, London (LK) and by DFG grant Nu 51/2-1 (BN).

References

- Barandun J, Reyer H-U (1998). Reproductive ecology of *Bombina variegata*: habitat use. *Copeia* **1998**: 497–500.
- Barton NH (2000). Estimating multilocus linkage disequilibria. *Heredity* **84**: 373–389.
- Butlin RK, Tregenza T (1998). Levels of genetic polymorphism: marker loci *versus* quantitative traits. *Phil Trans R Soc Lon B* **353**: 187–198.
- Doherty JA, Gerhardt HC (1984). Acoustic communication in hybrid treefrogs: sound production by males and selective phonotaxis of females. *J Comp Physiol* **154**: 319–330.
- Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR (2001). Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. *Mol Ecol* **10**: 1265–1278.
- Goodman SJ, Barton NH, Swanson G, Abernethy K, Pemberton JM (1999). Introgression through rare hybridisation: a genetic study of a hybrid zone between red and sika deer (genus *Cervus*), in Argyll, Scotland. *Genetics* **152**: 355–371.
- Grant PR, Grant BR (1992). Hybridization of bird species. *Science* **256**: 193–197.

- Harrison RG (1993). *Hybrid Zones and the Evolutionary Process*. Oxford University Press: Oxford.
- Hewitt GM, Nichols RA, Barton NH (1987). Homogamy in a hybrid zone in the alpine grasshopper *Podisma pedestris*. *Heredity* **59**: 457–466.
- Howard DJ, Gregory PG, Chu J, Cain ML (1998). Conspecific sperm precedence is an effective barrier to hybridization between closely related species. *Evolution* **52**: 511–516.
- Johannesson K, Rolán-Alvarez E, Ekendahl A (1995). Incipient reproductive isolation between two morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**: 1180–1190.
- Jones AG, Ardren WR (2003). Methods of parentage analysis in natural populations. *Mol Ecol* **12**: 2511–2523.
- Kruuk LEB (1997). Barriers to gene flow: a *Bombina* (fire-bellied toad) hybrid zone and multilocus theory. PhD Thesis, University of Edinburgh.
- Kruuk LEB, Gilchrist JS, Barton NH (1999). Hybrid dysfunction in fire-bellied toads (*Bombina*). *Evolution* **53**: 1611–1616.
- Lande R (1981). Models of speciation by sexual selection on polygenic traits. *Proc Natl Acad Sci USA* **78**: 3721–3725.
- Lörcher K (1969). Vergleichende bioakustische Untersuchungen an der Rot- und Gelbbauchunke, *Bombina bombina* (L.) und *Bombina v. variegata* (L.). *Oecologia* **3**: 84–124.
- Lynch M, Ritland K (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* **153**: 1753–1766.
- MacCallum C (1994). Adaptation and habitat preference in the hybrid zone between *Bombina bombina* and *Bombina variegata* in Croatia. PhD Thesis, University of Edinburgh.
- MacCallum CJ, Nürnberger B, Barton NH, Szymura JM (1998). Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* **52**: 227–239.
- Mallet J, McMillan WO, Jiggins CD (1998). Estimating the mating behavior of a pair of hybridising *Heliconius* species in the wild. *Evolution* **52**: 503–510.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**: 639–655.
- McLain DK (1985). Clinal variation in morphology and assortative mating in the soldier beetle, *Cauliognathus pennsylvanicus* (Coleoptera: Cantharidae). *Biol J Linn Soc* **25**: 105–117.
- Moore WS (1987). Random mating in the Northern Flicker hybrid zone: implications for the evolution of bright and contrasting plumage patterns in birds. *Evolution* **41**: 539–546.
- Morgan MT, Conner JK (2001). Using genetic markers to directly estimate male selection gradients. *Evolution* **55**: 272–281.
- Neff BD, Repka J, Gross MR (2001). A Bayesian framework for parentage analysis: the value of genetic and other biological data. *Theor Pop Biol* **59**: 315–331.
- Nielsen R, Mattila DK, Clapham PJ, Palsbøll PJ (2001). Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic Humpback whale. *Genetics* **157**: 1673–1682.
- Nürnberger B, Hofman S, Forg-Brey B, Praetzel G, Maclean A, Szymura JM et al (2003). A linkage map for the hybridising toads *Bombina bombina* and *B. variegata* (Anura: Discoglossidae). *Heredity* **91**: 136–142.
- Ribi G, Oertli S (2000). Frequency of interspecific matings and of hybrid offspring in sympatric populations of *Viviparous ater* and *V. contectus* (Mollusca: Prosobranchia). *Biol J Linn Soc* **71**: 133–143.
- Rice WR (1998). Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In: Howard DJ, Berlocher SH (eds) *Endless Forms: Species and Speciation*. Oxford University Press: Oxford, pp 261–270.
- Rieseberg LH, Baird SJE, Desroches AM (1995). Patterns of mating in wild sunflower hybrid zones. *Evolution* **52**: 713–726.
- Sætre G-P, Moum T, Bures S, Kral M, Adamjam M, Moreno J (1997). A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* **187**: 589–592.
- Sanchristobal M, Chevalet C (1997). Error tolerant parent identification from a finite set of individuals. *Genet Res* **70**: 53–62.
- Schluter D (2001). Ecology and the origin of species. *Trends Ecol Evol* **16**: 372–380.
- Smajda C, Catalan J, Ganem G (2004). Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. *J Evol Biol* **17**: 165–176.
- Smith BR, Herbinger CM, Merry HR (2001). Accurate partition of individuals into full-sib families from genetic data without parental information. *Genetics* **158**: 1329–1338.
- Szymura JM (1993). Analysis of hybrid zones with *Bombina*. In: Harrison RG (ed) *Hybrid Zones and the Evolutionary Process*. Oxford University Press: Oxford, pp 261–289.
- Thomas SC, Hill WG (2000). Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* **155**: 1961–1972.
- Vines TH, Barton NH (2003). A new approach to detecting mixed families. *Mol Ecol* **12**: 1999–2002.
- Wolfram S (1996). *The Mathematica Book*, 3rd edn. Cambridge University Press: Cambridge.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)