HYBRIDIZATION OF BOMBINA BOMBINA AND B. VARIEGATA
(ANURA, DISCOGLOSSIDAE) AT A SHARP ECOTONE IN WESTERN UKRAINE: COMPARISONS ACROSS TRANSECTS AND OVER TIME

ALEXEY YANCHUKOV,1,2,3,4 SERGIY V. MEZHZHERIN,1
SVIATOSLAV Y. MOROZOV-LEONOV,1 NICHOLAS H. BARTON,3 AND BEATE NÜRNBERGER4
1Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, B. Khmelnytskogo Str. 15, 01601 Kiev, Ukraine
2Department of Comparative Anatomy, Jagiellonian University, Ingardena 6, 30-060 Kraków, Poland
3Institute of Evolutionary Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland
4Department Biologie II, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

Abstract.—Bombina bombina and B. variegata are two anciently diverged toad taxa that have adapted to different breeding habitats yet hybridize freely in zones of overlap where their parapatric distributions meet. Here, we report on a joint genetic and ecological analysis of a hybrid zone in the vicinity of Stryi in western Ukraine. We used five unlinked allozyme loci, two nuclear single nucleotide polymorphisms and a mitochondrial DNA haplotype as genetic markers. Parallel allele frequency clines with a sharp central step occur across a sharp ecotone, where transitions in aquatic habitat, elevation, and terrestrial vegetation coincide. The width of the hybrid zone, estimated as the inverse of the maximum gradient in allele frequency, is 2.3 km. This is the smallest of four estimates derived from different clinal transects across Europe. We argue that the narrow cline near Stryi is mainly due to a combination of habitat distribution and habitat preference. Adult toads show a preference for either ponds (B. bombina) or puddles (B. variegata), which is known to affect the distribution of genotypes within the hybrid zones. At Stryi, it should cause a reduction of the dispersal rate across the ecotone and thus narrow the cline. A detailed comparison of all five intensively studied Bombina transects lends support to the hypothesis that habitat distribution plus habitat preference can jointly affect the structure of hybrid zones and, ultimately, the resulting barriers to gene flow between differentiated gene pools. This study also represents a resampling of an area that was last studied more than 70 years ago. Our allele-frequency clines largely coincide with those that were described then on the basis of morphological variation. However, we found asymmetrical introgression of B. variegata genes into B. bombina territory along the bank of a river.

Key words.—Bimodality, cline, ecotone, habitat preference, hybrid zone, introgression, linkage disequilibrium.

Received December 7, 2004. Accepted December 29, 2005.

When ecologically differentiated, parapatric taxa hybridize, the balance of evolutionary forces in the resulting hybrid zone determines the long-term fate of the parental gene pools. Selection against genotypes in the wrong habitat (exogenous selection) and/or against intrinsically incompatible genotypes (endogenous selection) can indefinitely maintain taxon differences in the affected loci and traits. The rate of neutral introgression, however, depends on the barrier to gene flow that is built up by these selected loci. Theoretical models predict that this barrier is primarily a function of the overall strength of selection and of the distribution of selected loci across the genome (Barton 1993). In addition, the form of the ecotone between the habitats to which the hybridizing taxa are respectively adapted should also affect this barrier (Kruuk 1997). The ecotone transition may take the form of a sharp step, a smooth intergradation, or a mosaic of habitat patches of either kind. We present data from a new, unusually narrow transect in the Bombina hybrid zone and explore the relationship between habitat distribution and hybrid zone structure via comparison with four other intensively studied transects. This study also represents a resampling of a transect that was last studied 70 years ago (Horbulewicz 1927) and so provides insight into its stability.

It has long been suggested that an ecotone can determine the position of a hybrid zone under exogenous selection and that spatial variation in ecotone width translates into a similarly variable width of character clines between the hybridizing taxa. For example, Moore and Price (1993) argued that the hybrid zone between two subspecies of a woodpecker (Colaptes auratus) widens regionally because the underlying vegetation gradient also widens. In a pair of North American ground crickets (Allenomobius), it appears that a patchy distribution of local climatic conditions in a mountainous transect leads to a similarly patchy distribution of genotypes (Howard and Waring 1991). Also, the correlation between soil type and genotype in the mosaic hybrid zone between two Gryllus species is very tight on a regional scale, which suggests that any change in the habitat mosaic would alter the hybrid zone as well (Ross and Harrison 2002). Such correlations may be brought about by selection alone, acting either purely by habitat or through a mixture of the exogenous and endogenous forms. They may be further sharpened if the hybridizing taxa show a preference for their respective habitat. We argue that observed large differences in hybrid zone structure among several Bombina transects are due to the joint effect of habitat distribution and habitat preference.

Our analysis uses a population genetic model of endogenous selection and/or exogenous selection at a sharp ecotone (summarized in Barton and Gale 1993) as a point of reference. Individual dispersal is assumed to be random in direction and is modeled as a diffusion process through continuous habitat. The model also applies to patchily distributed habitat as long as the patch density is high relative to individual dispersal distance (Nagylaki 1975). It generates the following set of
predictions. At a single locus, selection of either kind establishes smooth, stable clines in allele frequency (Slatkin 1973; Barton and Gale 1993; Kruuk et al. 1999a). The width of the cline increases with the individual dispersal distance and decreases with selection strength (Slatkin 1973). The dispersal of unrecombined genotypes into a hybrid zone is the main source of linkage disequilibria in the zone’s center. These statistical associations among loci generate interactions among selected loci that increase the effective selection on any one locus. This produces even steeper clines and, possibly, sharp central steps in allele frequencies that are flanked by shallower gradients on either side (Barton 1983). Moreover, linkage disequilibria impose clinal variation and even stepped clines on neutral loci; the total genomewide selection generates a barrier to gene flow for neutral variants, which retards but cannot ultimately prevent their introgression into the opposite gene pool.

Based on this model, the strengths of evolutionary forces in the hybrid zone can be estimated from allelic variation at marker loci (Barton and Gale 1993); linkage disequilibrium in the zone’s center and cline width inform us about the dispersal range and the effective strength of selection per locus. From the existence of a central step and the shallow tails of introgression at the edges, we can infer the relative strengths of locus-specific versus indirect selection and the overall barrier to gene flow. Moreover, heterozygote deficits within hybrid populations may point to the recent arrival of immigrants with a different genetic makeup, assortative mating, and/or strong selection.

In the case of exogenous selection, habitat distributions and individual dispersal patterns other than those assumed above have been shown to affect the structure of hybrid zones (e.g., Slatkin 1973; Kruuk 1997; Cain et al. 1999). Even in the case of random movement, a gradual habitat transition as opposed to a sharp environmental step should produce wider allele frequency clines (Slatkin 1973; Kruuk 1997). The general effect of a habitat preference is to produce a closer match between the habitat and genotype distributions (Kruuk 1997). If, in fact, two types of habitat are juxtaposed along a sharp ecotone, then a habitat preference leads to a reduction of dispersal across the ecotone relative to movement among habitat patches on either side. As a consequence, cline width will shrink relative to the expectation from random movement. On the other hand, if the habitat types are intermingled in a spatial mosaic, then the preference allows a greater proportion of relatively pure genotypes of either kind to disperse into the center of the hybrid zone along a chain of preferred habitat patches. As a consequence, neighboring habitat patches of opposite type might show marked differences in allele frequency. This generates a statistically noisier cline that is wider on average than it would be if the organisms moved randomly (MacCallum et al. 1998). This combined effect of habitat preference and habitat distribution affects the rate at which neutral loci dissociate from the selected genetic background and hence their rate of introgression into the opposite gene pool. It may also affect the rate of introgression of universally favored alleles, although they are expected to cross a hybrid zone fairly rapidly in any case (Barton 1979).

A comparison of hybrid zones between the fire-bellied toads *Bombina bombina* and *B. variegata* is ideally suited to investigate the hypothesis that habitat preference and habitat distribution jointly affect the dynamics of hybrid zones. These taxa are adapted to different kinds of habitat.
bellied toad, *B. bombina*, reproduces in semipermanent ponds, whereas the yellow-bellied toad, *B. variegata*, is a specialist for ephemeral breeding sites. Relative to *B. bombina*, it lays a smaller number of larger eggs (Rafińska 1991) that are distributed in several batches across a given habitat patch and give rise to more rapidly developing tadpoles. Bouts of reproduction are triggered by heavy rainfall during an extended breeding period, and water temperature serves as the proximate cue for choosing a particular oviposition site (Barandun and Reyer 1998). Moreover, due to their sturdier skeleton and thicker skin (Czopkowa and Czopek 1955), *B. variegata* adults seem better adapted to dispersal over land in search for newly available breeding sites. In contrast, *B. bombina* expresses many traits that are seemingly adaptive for reproduction in ponds. The larvae possess a high tail fin and show relatively quiescent behavior (Reichwaldt 1999; Vorndran et al. 2002), that is, traits that are typical of tadpoles in predator-rich habitats. Moreover, the males produce much louder calls and often form large choruses, which presumably attract females from a distance.

On a broader scale, the parapatric distribution ranges of the taxa differ in topography. *Bombina bombina* inhabits the lowlands of Central and Eastern Europe, whereas *B. variegata* is found at higher elevations in the Balkan and Carpathian Mountains. Hybrid zones typically form at altitudinal transitions (Fig. 1A), which are often associated transitions from forested to open terrain. Within them, the distribution of forest cover can play an important role. In the Peščenica hybrid zone in Croatia, *B. bombina* alleles were less frequent in forested sites than in those nearby in open, arable land, after controlling for the effects of aquatic habitat (MacCallum 1994). This fits the observation that *B. bombina* requires exposed, warm breeding sites for successful larval development (Günther and Schneeweiss 1996; B. Nürnberg, F. Zajitschek, S. Zajitschek, and G. Mara, unpubl. ms).

Evidence for an adult habitat preference comes from the joint genetic and ecological analyses of two *Bombina* hybrid zones (Peščenica and Apahida, Romania; MacCallum et al. 1998; Vines et al. 2003). There, correlations between allele frequencies and aquatic habitat existed on a spatial scale that is smaller than the per season dispersal radius of the animals. If toads moved into breeding sites at random, one would have to postulate implausibly strong selection on adults to explain this pattern. Of course, this conclusion does not rule out an
added contribution from habitat-specific natural selection. In fact, it is most likely that the *Bombina* hybrid zone is maintained by a combination of exogenous (Szymura 1993) and endogenous selection (Kruuk et al. 1999b).

There has been a long-standing interest in *Bombina* hybridization (Méhely 1892), and detailed studies on particular transects were carried out decades ago, so that we can directly investigate the stability of these transects by resampling them. The comparison at two transects near Kraków and Przemyśl in Poland revealed that their position did not change between samplings 13 years and more than 50 years apart, respectively (Horbulewicz 1933; Michaowski 1958; Szymura and Barton 1991). Here, we present a new genetic analysis of a transect in the Stryi River valley in Ukraine that was last sampled by Horbulewicz (1927). Other previously studied transects featured either smooth clinal transitions (near Przemyśl, Poland; Szymura and Barton 1986, 1991), clinal transitions with a mosaic structure in the center (near Kostajnica and Pešćenica, Croatia; Szymura 1993; MacCallum et al. 1998), or a large-scale mosaic without a clinal component (near Apahida, Romania; Vines et al. 2003). The Stryi transect neatly adds the opposite endpoint to this spectrum: it combines a sharp ecotone with the smallest cline width found to date in a *Bombina* hybrid zone. Our detailed comparison of all these transects lends support to the hypothesis that habitat preference interacts with habitat distribution to determine the current shape and, ultimately, the fate of hybrid zones.

**Materials and Methods**

**Study Area**

During the spring and summer seasons of 2000–2003, adult toads were sampled from the water bodies in an area of 30 × 40 km in the western Ukraine, on the right bank of the Stryi River, around the towns of Stryi and Morshyn (Fig. 1B). This area is adjacent to the region sampled extensively by Horbulewicz (1927) and was examined by Scherbak and Scherban (1980). Both studies provided evidence for the occurrence of morphologically intermediate individuals. The landscape consists mainly of forested (45%) and agricultural lands (50%), some of it under traditional crop growing and grazing, but also with a large proportion of abandoned fields. The elevation ranges from 190 to 400 m above sea level. There are two main types of forest, a mixed *Fagus-Abies-Picea-Acer* association found on the brown soils at higher elevations and lowland *Quercus-Fagus* forest on alluvial soil in the river valley. The sampling sites were small ponds, drainage ditches, flooded streamside meadows, and puddles in the small ground depressions or more often wheel ruts on
the roads. The spatial limits of each site were assigned arbitrarily, but never exceeded 200 m in diameter.

**Collection of Toads**

A total of 535 toads in years 2000–2001, 282 in 2002, and 329 in 2003 were collected in the water bodies or nearby and anesthetized with MS 222 (3-amino-benzoic acid ethyl ester). Most of them were adults, and only a minor proportion (2%) was clearly juvenile. A photograph was taken of each individual’s belly color spot pattern; a toe was clipped and kept on ice for subsequent allozyme analysis and/or preserved in 95% ethanol for DNA analysis. The site east of Morshyn was sampled repeatedly in each of the four years, and the site east of Dolishnie was sampled twice in 2001 and 2002 (Table 1 of Appendix 1 available online only at http://dx.doi.org/10.1554/04-739.1.s1).

**Allozyme Electrophoresis**

Enzymes from the tissue homogenates were resolved in 7.5% continuous polyacrylamide gels (Peacock et al. 1965) and 12% starch gels in Tris-citrate buffer, pH 7.5 at room temperature) and 2% TBE electrode buffer. Bands were visualised by silver staining. Details of these techniques are available upon request from A. Yanchukov or B. Nürnberg. All seven nuclear loci are unlinked (Nürnberg et al. 2003).

The 424-bp fragment of cytochrome b of mitochondrial DNA was amplified using the following primers: 5’-ACC CAAGCTTCCGGAAAATCGCA-3’ and 5’-CCCTCAAATGTATTTCTCCCAA-3’. Diagnostic restriction fragment length polymorphisms are produced in this fragment by BamHI and HaeIII restriction enzymes. Both polymorphisms were resolved in 2% agarose gel to double-check the haplotype of each individual (Hofmann 2002).

Analysis of the frozen and alcohol-preserved tissue samples was performed in three laboratories (Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, Kiev; Jagiellonian University, Kraków; Ludwig-Maximilians-Universität, Munich) and not all loci could be scored for all sites. The most complete datasets exist for *Ldh-I*, *Mdhl-I*, and mitochondrial DNA (mtDNA). The two SNP-SSCPs (*Bv24.11* and *Bb7.4*) were mostly scored in the zone center, and only a small subset of sites across the transect were scored for *Gpi*, *Np*, and *Ak* (Table 1). Our inference of the within-population genotype distributions is based on at least four loci per individual. However, the sets of loci varied across sites according to the sampling scheme. We therefore investigated the equivalence of loci in terms of their concordance across the hybrid zone (see below).

**Habitat Description**

For each site the elevation, type of terrestrial habitat (forest, grazing land, pasture, arable, or residential area), soil and relief type, presence of other anurans, and aquatic insects were recorded. The following quantitative parameters describing the aquatic habitat were estimated: length, width, and depth of the water body; percentage of the surface constantly shaded; degree of the bank steepness; emergent and submerged vegetation in percent of surface cover; and percentage of the bank vegetation in three height classes (below 15 cm, 15–50 cm, and above 50 cm).

We used a discriminant analysis to reduce this set of 10 variables to a single habitat axis in the following way: 59 of 72 sample sites were a priori assigned to one of the two habitat types (ponds or puddles). The remaining sites were intermediate. Seven variables were transformed to improve their normality (log for continuous variables, arcsine for percentages; Sokal and Rohlf 1995). With the stepwise method in SPSS version 10.0 (SPSS, Inc., Chicago, IL) we identified five significant variables for the discriminant function: width, depth, percent submerged vegetation, and percent of the shore vegetation in the height classes below 15 cm and above 50 cm.

<table>
<thead>
<tr>
<th>Ldh-I/Mdh-I</th>
<th>Gpi/Np</th>
<th>Ak</th>
<th>Bb7.4/ Bv24.11</th>
<th>mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. bombina</em> side (≤−3.64 km)</td>
<td>9 (209)</td>
<td>9 (193)</td>
<td>3 (75)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Center</td>
<td>57 (815)</td>
<td>12 (266)</td>
<td>8 (130)</td>
<td>49 (476)</td>
</tr>
<tr>
<td><em>B. variegata</em> side (&gt;4.6 km)</td>
<td>5 (112)</td>
<td>5 (107)</td>
<td>5 (110)</td>
<td>—</td>
</tr>
</tbody>
</table>

**Table 1.** Number of sites and individuals (in parentheses) sampled for various markers at the periphery and the center of the hybrid zone. Distance is given relative to the approximate center of the hybrid zone, negative for *Bombina bombina* and positive for *B. variegata* side.
cm. These are almost identical to those found to be the most important in discriminating the aquatic sites near Pešćenica in Croatia (MacCallum et al. 1998) and Apahida in Romania (Vines et al. 2003). The discriminant function coefficients are given in Table 2. Finally, the per site discriminant scores were rescaled to fit the interval $H_{aq}$ [0 . . .1], where 0 represents the most extreme pond and 1 the most extreme puddle. Since a rigorous evaluation of the terrestrial habitat features is difficult, we considered only three robust categories: forest, open land, and the transition between the two.

**Recording Individual Genotypes**

Each of the seven nuclear loci had two alleles, fixed alternatively for pure *B. bombina* or *B. variegata*. The two restriction enzymes used to assign the mtDNA haplotype always gave consistent results (Szymura et al. 2000; Hofman 2002). The sum of *B. variegata* alleles gave a nuclear hybrid index for each individual ($H_n$). For the total hybrid index ($H_t$) we also added a score of zero for a *B. bombina* or two for a *B. variegata* mtDNA haplotype, thus giving equal weight to each marker locus. Both indices were scaled to range from zero (pure *B. bombina*) to one (pure *B. variegata*).

**Statistical Methods**

The width of the cline in one dimension; variance in allele frequency; and parameters that describe the concordance of clines, heterozygote deficit, and linkage disequilibria were estimated by maximum likelihood (ML; Edwards 1972; Szymura and Barton 1991; MacCallum et al. 1998). Support limits represent the values of a given parameter for which the natural logarithm of likelihood ($L$) is two units below the maximum (equivalent to 95% confidence limits in large samples). Calculations were performed using the Analyse 1.30 PPC software package by Barton and Baird (1996).

**Cline Analysis**

To simplify the interpretation of the complex spatial structure of the hybrid zone, a common spatial scale must be found to measure clinal variation at all or most of the examined sites. This could be done by dividing the hybrid zone into sections, each with its own cline center and width as implemented in Analyse (Barton and Baird 1996). However, for complex datasets this procedure requires a lot of computing time (e.g., Marshall and Sites 2001). We therefore have used two other simpler methods to find the common transect against which all loci could be tested. First, for the most distant populations at the periphery, for which the shortest distance to the current transect’s center is difficult to determine, distance was simply measured to the center of the morphological hybrid zone, as mapped by Horbulewicz (1927; cf. straight, dotted line in Figs. 1B and 1C). Second, for the large proportion of sites in the densely sampled central area, the value of the total frequency of *B. variegata* alleles between loci, it was assumed that a proportion $(1 - F_{IS})$ of gametes combines at random, while fraction $F_{IS}$ is homozygous for both loci. The average of the two single-locus estimates of $F_{IS}$ was taken into account while calculating the expected proportions of the two-locus genotypes to find the ML estimate of $R$ (MacCallum 1994).

To test whether apparent $F_I$ genotypes were over- or underrepresented within sites, we computed the distribution of gamete frequencies from the estimates of allele frequencies, heterozygote deficits, and linkage disequilibria per site under the assumption of neutral admixture (Barton 2000). To facilitate the computations in Mathematica 5.0 (Wolfram Research, Champaign, IL), we assumed equal allele frequencies and linkage disequilibria across loci. Note that only a subset of genotypes heterozygous at all loci will represent first-generation hybrids, while others may result from the combination of several complementing classes of gametes. We use the following model of Gaussian assortment with mating probability:

$$\text{Exp} \left[ -\frac{(i - j)^2}{2V} \right] P_i P_j,$$

where $P$ is the frequency of gametes in classes $i, j$ [0 . . .4], $i$ and $j$ represent the number of *B. variegata* alleles per gamete, and $V$ is an arbitrary positive parameter adjusted to match the strength of observed heterozygote deficit. The maximum mating probability is therefore one for gametes of the same class. The Mathematica notebook describing the model is available upon request from A. Yanchukov or N. Barton.

**Population Genetic Analysis**

To assess the effect of migration and selection in the hybrid zone, Wright’s inbreeding coefficient ($F_{IS}$) and the pairwise gametic linkage disequilibria ($D$) were estimated across loci and samples. Heterogeneity between loci was estimated as the difference in ln likelihoods ($\Delta n L$) between models with a constant $F_{IS}$ across loci versus $F_{IS}$ fitted separately for each locus. Linkage disequilibrium was standardized by allele frequency

$$R_{ij} = D_{ij}/\sqrt{p_i q_i p_j q_j}.$$

Because heterozygote deficit inflates observed associations

**Table 2.** Standardized canonical discriminant function coefficients of the empirically defined habitat types near Stryi.

<table>
<thead>
<tr>
<th>Overall</th>
<th>Ponds</th>
<th>Puddles</th>
<th>Wilks $\lambda$</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submerged vegetation</td>
<td>1.08</td>
<td>9.5</td>
<td>-7.07</td>
<td>0.53</td>
</tr>
<tr>
<td>Depth</td>
<td>1.02</td>
<td>-1.72</td>
<td>-7.34</td>
<td>0.33</td>
</tr>
<tr>
<td>Shore vegetation &lt;15 cm</td>
<td>-0.5</td>
<td>2.09</td>
<td>5.65</td>
<td>0.23</td>
</tr>
<tr>
<td>Shore vegetation &gt;50 cm</td>
<td>0.47</td>
<td>8.33</td>
<td>0.96</td>
<td>0.23</td>
</tr>
<tr>
<td>Width</td>
<td>-0.40</td>
<td>2.31</td>
<td>3.97</td>
<td>0.21</td>
</tr>
<tr>
<td>Constant</td>
<td>6.55</td>
<td>2.33</td>
<td>-8.23</td>
<td></td>
</tr>
<tr>
<td>Group mean</td>
<td>2.33</td>
<td></td>
<td>-2.01</td>
<td></td>
</tr>
</tbody>
</table>

Note: These are almost identical to those found to be the most important in discriminating the aquatic sites near Pešćenica in Croatia (MacCallum et al. 1998) and Apahida in Romania (Vines et al. 2003). The discriminant function coefficients are given in Table 2. Finally, the per site discriminant scores were rescaled to fit the interval $H_{aq}$ [0 . . .1], where 0 represents the most extreme pond and 1 the most extreme puddle. Since a rigorous evaluation of the terrestrial habitat features is difficult, we considered only three robust categories: forest, open land, and the transition between the two.

**Recording Individual Genotypes**

Each of the seven nuclear loci had two alleles, fixed alternatively for pure *B. bombina* or *B. variegata*. The two restriction enzymes used to assign the mtDNA haplotype always gave consistent results (Szymura et al. 2000; Hofman 2002). The sum of *B. variegata* alleles gave a nuclear hybrid index for each individual ($H_n$). For the total hybrid index ($H_t$) we also added a score of zero for a *B. bombina* or two for a *B. variegata* mtDNA haplotype, thus giving equal weight to each marker locus. Both indices were scaled to range from zero (pure *B. bombina*) to one (pure *B. variegata*).

**Statistical Methods**

The width of the cline in one dimension; variance in allele frequency; and parameters that describe the concordance of clines, heterozygote deficit, and linkage disequilibria were estimated by maximum likelihood (ML; Edwards 1972; Szymura and Barton 1991; MacCallum et al. 1998). Support limits represent the values of a given parameter for which the natural logarithm of likelihood ($L$) is two units below the maximum (equivalent to 95% confidence limits in large samples). Calculations were performed using the Analyse 1.30 PPC software package by Barton and Baird (1996).

**Cline Analysis**

To simplify the interpretation of the complex spatial structure of the hybrid zone, a common spatial scale must be found to measure clinal variation at all or most of the examined sites. This could be done by dividing the hybrid zone into sections, each with its own cline center and width as implemented in Analyse (Barton and Baird 1996). However, for complex datasets this procedure requires a lot of computing time (e.g., Marshall and Sites 2001). We therefore have used two other simpler methods to find the common transect against which all loci could be tested. First, for the most distant populations at the periphery, for which the shortest distance to the current transect’s center is difficult to determine, distance was simply measured to the center of the morphological hybrid zone, as mapped by Horbulewicz (1927; cf. straight, dotted line in Figs. 1B and 1C). Second, for the large proportion of sites in the densely sampled central area, the value of the total frequency of *B. variegata* alleles between loci, it was assumed that a proportion $(1 - F_{IS})$ of gametes combines at random, while fraction $F_{IS}$ is homozygous for both loci. The average of the two single-locus estimates of $F_{IS}$ was taken into account while calculating the expected proportions of the two-locus genotypes to find the ML estimate of $R$ (MacCallum 1994).

To test whether apparent $F_I$ genotypes were over- or underrepresented within sites, we computed the distribution of gamete frequencies from the estimates of allele frequencies, heterozygote deficits, and linkage disequilibria per site under the assumption of neutral admixture (Barton 2000). To facilitate the computations in Mathematica 5.0 (Wolfram Research, Champaign, IL), we assumed equal allele frequencies and linkage disequilibria across loci. Note that only a subset of genotypes heterozygous at all loci will represent first-generation hybrids, while others may result from the combination of several complementing classes of gametes. We use the following model of Gaussian assortment with mating probability:

$$\text{Exp} \left[ -\frac{(i - j)^2}{2V} \right] P_i P_j,$$

where $P$ is the frequency of gametes in classes $i, j$ [0 . . .4], $i$ and $j$ represent the number of *B. variegata* alleles per gamete, and $V$ is an arbitrary positive parameter adjusted to match the strength of observed heterozygote deficit. The maximum mating probability is therefore one for gametes of the same class. The Mathematica notebook describing the model is available upon request from A. Yanchukov or N. Barton.

**Population Genetic Analysis**

To assess the effect of migration and selection in the hybrid zone, Wright’s inbreeding coefficient ($F_{IS}$) and the pairwise gametic linkage disequilibria ($D$) were estimated across loci and samples. Heterogeneity between loci was estimated as the difference in ln likelihoods ($\Delta n L$) between models with a constant $F_{IS}$ across loci versus $F_{IS}$ fitted separately for each locus. Linkage disequilibrium was standardized by allele frequency

$$R_{ij} = D_{ij}/\sqrt{p_i q_i p_j q_j}.$$

Because heterozygote deficit inflates observed associations
(π) was interpolated using ArcView 3.1 GIS Spatial Analyst software (ESRI, Redlands, CA). The program creates a fine-scale lattice, where for any square cell k with side length = 0.1 km, which is the approximate size of the majority of collection sites, the value \( p_k = E[\pi/\{\exp(S_n)\}] \). \( S_n \geq 4 \) km is computed, where \( j = 1, 2, \ldots \) are the cells with known \( \pi \) (in our case the locations of the actual sampling site) within a fixed radius of 4 km (chosen arbitrarily); and \( S_0 \) is the distance from \( k \) to \( j \). The value of \( p_k \) corresponds to the frequency of \( B. variegata \) alleles expected in cell \( k \). After the lattice was built, the cells with values \( 0.49 < \bar{\pi} < 0.51 \) formed an almost continuous curve, which estimates the center of the hybrid zone (Fig. 1C). We define the distance to the cline center as the shortest distance between each site and that curve, thus reducing the data to a one-dimensional transect. This method does not provide confidence intervals around the approximation and ignores differences in sample size. However, it gives comparable results to the more complex fitting routine used by Analyse with a large number of segments (analysis not shown).

According to the theory summarized in the introduction, allele frequency clines are expected to either have a sigmoidal shape or, in the case of strong interactions among selected loci, to form a central step flanked by shallower gradients on either side. We used the latter to infer the cline parameters throughout. The frequency of \( B. variegata \) alleles at position \( x \) in the center of the transect is given by: \( p = \{1 + \tanh[2(x - y)/w]\}/2 \), where \( w \) is the width (defined as the inverse of the maximum gradient) and \( y \) is the position of the cline center (Szymura and Barton 1991). The left and right tails of the cline are described by \( p = e^{\pm \pi \sqrt{2\pi} \sigma_{\alpha \beta}} \) and \( p \), respectively. The parameters \( \theta_{\alpha \beta}, \theta_{\alpha \beta} \) (typically < 1) describe the shallower gradient in allele frequency in the left and right tails, compared with the center. The barrier to gene flow \( (B_{\alpha \beta}, B_{\alpha \beta}) \) is defined as the ratio between the step in gene frequency (\( \Delta p \)) in the center and the gradient of the decay \( (dp/dx) \) on either side of the cline. The Metropolis algorithm (Metropolis et al. 1953) was used to fit the six parameters \( (w, \theta_{\alpha \beta}, \theta_{\alpha \beta}, B_{\alpha \beta}, B_{\alpha \beta}) \) for the cline model (Szymura and Barton 1991). Two additional parameters, the minimum and the maximum gene frequency values \( (p_{\text{min}} \text{ and } p_{\text{max}}) \), in the tails of the cline, were set to those actually observed in the most distant samples or were taken as zero and one, for the SNP-SSCP loci, which were not scored at the periphery of the hybrid zone. Several thousand runs were performed to find the best fit. Random fluctuations of allele frequencies and sampling error need to be taken into account in the computation of the ML cline fits. To this end, we first computed the standardized variance in allele frequencies, \( F_{ST} \), around the fitted spatial cline, given by

\[
F_{ST} = \frac{E[\{p_{x} - p_{o}\}^2]}{p_{x}(1 - p_{x})/n_i},
\]

where \( p_{x} \) and \( p_{o} \) are expected and observed mean allele frequencies per site and \( n_i \) is the total number of alleles sampled at site \( i \). Note that this is analogous to the typical \( F_{ST} \) used to measure spatial variation in allele frequencies, \( F_{ST} = [\text{var}(p_o)]/\{p(1 - p)\} \), in that one expectation (the overall mean frequency) is replaced by another (the cline prediction per site; Szymura and Barton 1991). The ML estimate for \( F_{ST} \) was found assuming a negative binomial distribution of \( p_{o} \), with variance \([F_{ST} + (1/n_i)]p_{x}(1 - p_x)\). We used \( F_{ST}^0 \) to adjust the sample size of each site in the following way: \( 1/n_i = 1/n_e + F_{ST}^0/k \) (Szymura and Barton 1991; MacCallum 1994), where \( k \) is the number of diagnostic loci and \( n_e \) is the effective sample size. The likelihood of the cline model is then recalculated upon substituting \( n_i \) with \( n_e \). This procedure avoids giving undue weight to a few very large samples.

The mean displacement of individuals between their sites of birth and reproduction, \( \sigma \) (henceforth, the mean dispersal range), was calculated as follows from the width of the cline \( (w) \) and the average value of the linkage disequilibrium \( (R) \) in the center of the hybrid zone \( (p = 0.5) \): \( \sigma^2 = w^2 R_p (1 - p) \) (Szymura and Barton 1986, 1991). The recombination rate \( r \) is 0.5 for all pairs or loci (Nürnberg et al. 2003).

**RESULTS**

Our survey of 72 locations spanned the full range of allele frequencies from pure \( B. bombina \) to pure \( B. variegata \). The majority (63%) of the 1146 genotyped individuals were clearly hybrid. Yet, only three specimens were heterozygous at all five loci scored, and so may be \( F_1 \) hybrids. Two of them possessed a \( B. bombina \), and one had a \( B. variegata \) mtDNA haplotype. In three populations in the vicinity of Dolishnie (site A in Fig. 1B) pure marker genotypes of both kinds co-occurred. No pure genotypes of the respective other taxon were observed in peripheral samples, but a single, entirely heterozygous individual was found on the river bank 28 km away from the center of the hybrid zone on the \( B. bombina \) side.

**Concordance of Characters and Residual Variation**

Linkage disequilibria in hybrid zones tend to impose concordance on clines at neutral markers. Deviations of particular loci from the average clinal transition could be caused by direct selection on them. Similarly, selection might be responsible for a scattering of clines to different positions.

To test for the level of concordance among loci in the Stryi transect, the following model was fitted to the cline at a given locus \( i: p_i = \rho + 2pq[\alpha_i + \beta_i(\bar{\pi} - \bar{\pi})] \) (Szymura and Barton 1986). Parameters \( \alpha \) and \( \beta \) in the equation indicate twice the shift in position and the reduction in width, respectively, of that cline relative to the averaged cline across all loci with local allele frequencies \( \bar{\pi} \) and \( \bar{\pi} = 1 - \bar{\pi} \) (Table 3). Recall that not all sites were scored for all loci, so that the parameter estimates vary somewhat in their coverage of the cline and in their precision (cf. Table 1).

All estimates of \( \alpha \) and \( \beta \) were small (only two exceeded 0.1), and their support limits invariably spanned zero. The standardized allele frequency variance among loci around \( \bar{\pi} \), \( F_{ST}^0 \), was less than expected from sampling error: observed \( F_{ST}^0 \) − expected \( F_{ST}^0 \) = −0.018.

**Deviations from Hardy-Weinberg Proportions and Linkage Equilibrium**

The ML estimates of \( F_{ST} \) per locus were estimated by the sum of \( n \) likelihoods in each site (Table 3). We list them separately by locus and sampling period to allow for comparisons among loci. Estimates ranged from near zero to a
maximum of 0.48. Average heterozygote deficits per locus were particularly large for sites sampled in 2002 (range = 0.17–0.48). In the 2000–2001 samples, they were heterogeneous across loci (2ΔlnL = 12.4, df = 3, P < 0.01): the allozymes Ldh-1 and Mdh-1 gave larger estimates than the two SNPs. Standardized linkage disequilibria ($R_{ij} = D_{ij}/\sqrt{P_{ii}P_{jj}}$) were estimated for pairs of loci across all sites. They ranged from 0.18 to 0.42 and were significant throughout (Table 4). The ML estimate over all pairs was 0.278 (95% CL = 0.275, 0.283).

Heterozygote deficits and pairwise disequilibrium were also examined across populations as a function of the mean allele frequency. To obtain robust results, sites with a similar proportion of B. variegata alleles were grouped together. Both $F_{IS}$ and $R$ increase toward intermediate frequencies (Fig. 2), but the maximum of the latter is shifted toward the B. bombina side. Note that the same asymmetry was observed in the Croatian (MacCallum et al. 1998) and Romanian hybrid zones (Vines et al. 2003). Although most of the linkage disequilibrium was observed in the center of the transect, several peripheral populations also possessed particularly large disequilibrium values ($R = 0.99$ [95% CL = 0.44, 1.0] at Bollekiv; $R = 0.6$ [0.43, 0.67] at Zhydachiv, cf. Fig. 1B) due to the presence of individuals with hardly recombined alien genotypes.

### Cline Approximation

The Stryi hybrid zone closely follows a sharp ecotone between upland forest and lowland meadows and pastures. The winding course of this ecological transition at the edge of the Stryi valley possesses a similarly complicated trajectory on the hybrid zone center (cf. Fig. 1C). This observation is consistent with Horbulewicz’s (1927) data, whose depiction of the hybrid zone at nearby Drohobych and Sambir also closely traces the relief with many protrusions and recesses. To reduce this complicated layout to a one-dimensional cline, we made the following approximations. Four sites in the southeastern part of our sampling area were excluded from analysis, because they may actually be closer to a different section of the hybrid zone along the Carpathian ridge. Collections in three other sites (Myrtiuky, Pishchany, and Zhydachiv) along the bank of the Stryi possess unexpectedly high frequencies of B. variegata alleles, which may be due to passive transport down the river, and so were treated separately. Note that small bulges of the fitted cline center in either direction will add noise to the central portion of the one-dimensional cline, because the mean allele frequency of immigrants in nearby sites is then not accurately predicted by the smallest distance to the cline center (for an example see Fig. 3).

The allele frequency plots on a logit scale $Z = \log(p/q)$ confirm the existence of a steep central gradient and two shallower tails at the edges for each locus. Figure 4A shows the average over all loci. The model parameters are given in Table 5. The estimates of cline width ($w$) per locus range from 0.86 km ($Mdh-1$) to 4.25 km ($Ak$), with an overall average of 2.3 km. Consistent with the concordance analysis, there is hardly any scatter in the position of the cline centers

### Table 3. Discordance in position and heterozygote deficit across loci. Estimates of $F_{IS}$ are made for the total dataset and within subsets of samples collected in different years. $\Delta ln L$ tests the heterogeneity between estimates of $F_{IS}$ for different loci.

<table>
<thead>
<tr>
<th>Year</th>
<th>Ldh-1</th>
<th>Mdh-1</th>
<th>Bb7.4</th>
<th>Bv24.11</th>
<th>Gpi</th>
<th>Np</th>
<th>Ak</th>
<th>mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000–2001</td>
<td>0.22</td>
<td>0.12</td>
<td>0.33</td>
<td>0.1</td>
<td>0.07</td>
<td>0.07</td>
<td>6.18</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>0.45</td>
<td>0.3</td>
<td>0.6</td>
<td>0.48</td>
<td>0.62</td>
<td>0.17</td>
<td>0.34</td>
<td>0.38</td>
</tr>
<tr>
<td>2003</td>
<td>0.0</td>
<td>0.25</td>
<td>0.01</td>
<td>0.23</td>
<td>0.01</td>
<td>0.01</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>0.26</td>
<td>0.18</td>
<td>0.34</td>
<td>0.18</td>
<td>0.27</td>
<td>0.07</td>
<td>0.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### Table 4. Pairwise standardized linkage disequilibria ($R$) across all sites. The maximum likelihood estimate of $R$ over the entire dataset (with support limits) is shown in the bottom row. Gaps are left for the pairs of loci that were not scored for the same individuals.

<table>
<thead>
<tr>
<th></th>
<th>Ldh-1</th>
<th>Mdh-1</th>
<th>Bb7.4</th>
<th>Bv24.11</th>
<th>Gpi</th>
<th>Np</th>
<th>Ak</th>
<th>mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ldh-1</td>
<td>0.33</td>
<td>0.29</td>
<td>0.194</td>
<td>0.375</td>
<td>0.372</td>
<td>0.34</td>
<td>0.34</td>
<td>0.365</td>
</tr>
<tr>
<td>Mdh-1</td>
<td>0.305</td>
<td>0.261</td>
<td>0.328</td>
<td>0.413</td>
<td>0.214</td>
<td>0.275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bb7.4</td>
<td>0.24</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.18</td>
</tr>
<tr>
<td>Bv24.11</td>
<td>—</td>
<td>0.31</td>
<td>0.415</td>
<td>0.405</td>
<td>0.227</td>
<td>0.365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gpi</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.235</td>
</tr>
<tr>
<td>Np</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ak</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Average</td>
<td>0.32</td>
<td>0.30</td>
<td>0.25</td>
<td>0.23</td>
<td>0.37</td>
<td>0.34</td>
<td>0.29</td>
<td>0.30</td>
</tr>
<tr>
<td>Total</td>
<td>0.278 (0.275, 0.283)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2. Distribution of (A) $F_{IS}$ and (B) $R$, estimated by the maximum likelihood, in sites grouped by Bombina variegata allele frequency, $P$. Pluses represent 95% confidence support limits.

The ML estimates of the barrier to gene flow on the $B. bombina (B_b)$ and the $B. variegata (B_v)$ side vary considerably among loci (Table 5), partly because the peripheral gradients in allele frequency were estimated from a limited number of sites. More reliable estimates are based on the combined dataset of all loci with peripheral samples ($Mdh-1$, $Ldh-1$, $Gpi$, $Np$, and mtDNA). These were 6.96 and 1.68 cline widths on the $B. bombina$ side and the $B. variegata$ side, respectively. Averaged estimates of $\theta$ were obtained analogously and yielded $\theta_p = \theta_v = 0.25$. These figures represent the proportion of effective selection on a given locus that is due to fitness effects of the locus itself. The estimates above were calculated without allowing for random variation of allele frequencies across loci. Because the fluctuation of allele frequencies around the fitted concordance regressions, $F^2_{ST}$, was even smaller than expected from random fluctuation, we could only use the standardized variance in allele frequency around the prediction of the spatial cline fit, averaged over all sites (see eq. 3). This estimate of $F^2_{ST} = 0.079$ (support limits: 0.057, 0.11) was included to reduce the samples to their effective sizes, $n_e$ (see Materials and Methods). None of the differences among loci in any cline parameter remained significant after that correction, and there was no appreciable change in the shape of the clines. In particular, we found neither asymmetry nor a narrower width of the cline of mtDNA haplotypes compared to the nuclear markers. This is in contrast to the findings of Hofman (2002) near Kraków and Przemęty.

For a direct comparison between transects, we recomputed the cline parameters for the combined dataset from the two Polish transects (Szymura and Barton 1991) with the current version of the Analyse software. Their joint cline is replotted in Figure 4B. There was far less residual variation around the fitted spatial cline in the two Polish transects for six allozyme loci: $F^2_{ST} = 0.018$ (95% CL = 0.012, 0.028). Overall, the Stryi cline is narrower (2.3 km vs. 6.1 km in Poland), yet the tails of introgression imply weaker barriers to gene flow (Stryi: $B < 20$ km, Poland: $B > 50$ km). Estimates of $\theta$ were also generally larger at Stryi than in the Polish and Croatian transects, which in principle indicates stronger selection on the loci themselves. We return to this issue in the Discussion.

The estimate of the mean dispersal range, $s$, was based on the following parameter values in the center of the cline (cf. Materials and Methods): $w = 2.3$, $R = 0.37$ (interpolated by regression on Fig. 2B to $pq = 0.25$), and $r = 0.5$. Our estimate of $s = 0.495$ km/gen$^{1/2}$ is half that derived in a similar way by Szymura and Barton (1991) and MacCallum et al. (1998). Although linkage disequilibrium is about the same as in Poland, the cline is less than half as wide, and so the estimated dispersal distance is correspondingly lower.

River cline.—Five sites were chosen to represent the clinal variation of allele frequencies along the bank of the Stryi River. Given the limited data, only a single cline over all loci was computed. The transition was almost linear on a logit scale. So we used the tanh model (cf. Materials and Methods) to estimate the width of the cline along the river averaged across all loci as $w_{river} = 24.5$ km, which was about 10 times wider than on the remaining transect. Because this cline was surrounded by pure $B. bombina$ sites, it represents asymmetric introgression of $B. variegata$ alleles (Fig. 4D).
Quantitative Description of Habitats and Associations between Genotype and Habitat

At face value, the estimated cline parameters (in particular, cline width, barrier strength, and dispersal range) suggest that the Stryi transect is governed by a different combination of evolutionary forces than any previously studied *Bombina* transect (cf. Discussion). We investigate here whether these observations may be alternatively explained by an active habitat preference in combination with the particular habitat distribution in the Stryi transect, as outlined in the introduction. The Peščenica transect serves as a point of reference, as it is the only other clinal hybrid zone for which habitat information is available (MacCallum et al. 1998).

In the Stryi valley, the cline in allele frequency coincided with a sharp ecological transition in terms of both elevation and vegetation type (forest vs. open terrain; cf. Figs. 1B, 4B). Within 100–200 m of the forest edge, there was an abrupt change in elevation. In contrast, the center of the Peščenica transect curved around an expance of lowland forest such that no elevational transition was associated with the cline's center in that part of the zone (cf. fig. 1 in MacCallum et al. 1998). Moreover, a mosaic of ponds and puddles existed in and around that lowland forest. In contrast, the transition in aquatic habitat in the Stryi valley is sharper and coincides more closely with the center of the cline (Fig. 5). Thus, there is a much sharper juxtaposition of habitat in terms of topography, vegetation cover, and breeding sites in the Stryi transect.

Associations of aquatic habitat and allele frequency on the scale of the toads’ dispersal range support the notion of a habitat preference, as it is unlikely that such associations could be solely maintained by selection on randomly dis-
TABLE 5. Parameter estimates of fitted clines in the Stryi transect. Allele frequencies at the endpoints \((p_{\text{max}})^\text{B}^\text{b}\text{v}\) in \(\text{Bombina} \text{bombina}\) and \(B. \text{variegata}\), respectively. \(\Gamma\) is the gradient in allele frequencies \((w)\) into of the average cline over all loci. \(\beta\) is the deviation (in km) of the cline center at a given locus from that of the average cline over all loci. Parameter estimates are omitted in cases of insufficient data. Likelihood estimates of the best fit are given in the right column. Note that the clines at total allele frequencies were fitted separately and hence the parameters describing them are not the average of the particular locus values.

<table>
<thead>
<tr>
<th>Loci</th>
<th>df</th>
<th>(p_{\text{max}}^\text{B}^\text{b}\text{v})</th>
<th>(\Gamma)</th>
<th>(\beta)</th>
<th>(\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main transect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ldh-1</td>
<td>64</td>
<td>0.03</td>
<td>2.13</td>
<td>3.91</td>
<td>0.05</td>
</tr>
<tr>
<td>Mdh-1</td>
<td>64</td>
<td>0.00</td>
<td>0.06</td>
<td>0.27</td>
<td>0.01</td>
</tr>
<tr>
<td>Gpi</td>
<td>23</td>
<td>0.03</td>
<td>1.07</td>
<td>2.69</td>
<td>0.02</td>
</tr>
<tr>
<td>Ald</td>
<td>11</td>
<td>0.01</td>
<td>0.98</td>
<td>1.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Bb7.4</td>
<td>5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Nuclear</td>
<td>64</td>
<td>0.00</td>
<td>0.04</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>mtDNA</td>
<td>64</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>River cline</td>
<td>5</td>
<td>0.01</td>
<td>0.95</td>
<td>1.11</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Because the allele frequencies of all four loci at Dolishnie were sampled in the same year and also showed a bimodal composition.

Distribution of Genotypes within Populations

Near the forest edge, we discovered populations with strikingly different genotypic compositions, ranging from unimodal to bimodal distributions of the hybrid index. We illustrate the latter type with a particularly large sample (74 adults) that was collected in 2002 near the village of Dolishnie (site B in Fig. 1B). About two-thirds of the individuals had either pure \(B. \text{variegata}\) or \(B. \text{bombina}\) marker genotypes, while the remainder was uniformly distributed across intermediate hybrid indices (Fig. 8A). The collection site consisted of numerous puddles along a road at the forest edge and of a heavily grazed pasture. Even at this small scale of less than 200 m, there was a tendency of \(B. \text{bombina}^\text{--}\)like individuals to be found in the open part of the road and of \(B. \text{variegata}^\text{--}\)like ones closer to the forest. Possibly, the extremely high estimates of \(F_{IS}\) and \(R\) (Table 6) are caused by a combination of habitat-based assortative mating within the Dolishnie site and immigration of pure types from either side of the ecotone. A few other sites at the edge of the forest were sampled in the same year and also showed a bimodal composition.
were ≈0.5, we could calculate the expected proportion of the putative F1 hybrids from the linkage disequilibrium and heterozygote deficit in the sample (see Materials and Methods). The observed fraction of putative F1 hybrids (0.055) is significantly larger than expected \( P_{exp}(F1) = 0.038 \) assuming a Poisson distribution; Table 6; Fig. 8C), which may have resulted from the mating between individuals that do not belong to the local population and might have migrated from elsewhere.

However, most sites near the ecotone had a unimodal distribution of the hybrid index. The combined sample from three sites near the village of Stankiv illustrates this (site D in Fig. 1B). It is located 5 km north from Dolishnie, 1 km away from the nearest forest and represents similar aquatic habitat. At Stankiv, the hybrid index is distributed unimodally (Fig. 8B), as are the proportions of heterozygotes at multiple loci (Fig. 8D), and the observed number of the putative F1 hybrids is not significantly greater than expected \( P = 0.2 \), Table 6).

**Discussion**

The hybrid zone alongside the Stryi River shares a number of critical features with all previously studied clinal transects of the *Bombina* hybrid zone (Kraków, Przemyśl, Pešćenica). It coincides with an altitudinal transition. Recombinant genotypes predominate in the zone's center, yet populations with intermediate allele frequencies show strong linkage disequilibria. Moreover, our unlinked genetic markers show a high level of concordance across populations. Other features have been observed before in at least one *Bombina* hybrid zone. But taken together, the set of characteristics in the Stryi valley is unique. In particular, the narrow cline width contrasts with the estimated weak barrier to gene flow. Also, small-scale correlations between habitat and allele frequency, that is, the observation that had previously indicated a habitat preference, exist only right at the cline center, yet are there unexpectedly strong.

Irrespective of these differences in parameter estimates, the Stryi transect appears to be just as stable in terms of positions and width as the two Polish transects and over an unprecedented interval of more than 70 years (Kraków: 13 years, Przemyśl: > 50 years). However, the agreement is necessarily approximate, because Horbulewicz (1927) did not provide per site data. To display spatial variation, he delineated two belts that specified the course of the hybrid zone along the eastern edge of the Carpathians. A narrow, inner belt defined the region in which hybrid morphologies predominated, whereas the wider belt extended further outward on either side and described the area beyond which only pure types would be seen. The inner belt, which should roughly resemble our \( w \), was much more variable in position and width than the outer one. This is further evidence for spatial variation along the course of the *Bombina* hybrid zone. We
**FIG. 7.** The difference in *Bombina variegata* allele frequency ($\Delta p$) plotted against the difference in habitat score ($\Delta H_{aq}$) for pairs of sites less than 1 km apart. Dashed regression line, all pairs of sites; solid regression line, pairs of sites separated by the ecotone (represented by large datapoints).

**FIG. 8.** Genotypic distributions in two sites: Dolishnie (A, C) and Stankiv (B, D). Panels in the top row give the distribution of the total hybrid index, $H_t$, whereas the bottom row shows the distribution of the number of heterozygous loci per individual.
Parameter estimates come largely from the central portion of 220 km) are based on comparable sampling effort. The other sects. Inspection of Figures 4A and 4B, however, suggests that at least the values on the Dolishnie samples and so will be less precise than in the Polish transects, the level of concordance among loci is the highest particular, the estimates of barrier strength rely on peripheral transects, if we can show: (1) that the habitat distribution is indeed variable and shows the expected correlation with cline width; and (2) that the toads exert a preference in all the transects in question.

As described above, the habitat distribution in the Stryi valley features a sharp ecotone at which abrupt transitions in elevation, forest cover, and aquatic habitat coincide, whereas in Pešćenica the transitions in elevation and forest cover were dissociated in the middle section of the transect, which also featured a mosaic distribution of aquatic sites. This section produced the maximum cline width estimate for the fitted cline center in either direction bias the mean parameters at Stryi. The transposition of a complicated two-dimensional cline into one dimension will have resulted in some imprecision and loss of information. As outlined above, small protrusions of the fitted cline in either direction bias the mean allele frequencies in nearby sites. This should tend to increase the central step and so give the erroneous impression of a stronger barrier. This argument is based on random dispersal. The effect could be lessened in the case of a habitat preference (see below), and, given the unusually small barrier strength, it is not a serious cause for concern. Yet, it may explain the larger residual variance of mean allele frequencies around the fitted spatial cline: \( f_{ST} \) was 0.079 in Stryi compared to 0.018 in the Polish transects.

We now turn to the question of whether the toads’ habitat preference together with the local distribution of habitat could account for the variation in parameter estimates between transects. As we argued in the introduction, an active adult preference tends to generate an approximate replica of the local habitat distribution at the level of genotypes, resulting in an average cline width that is either narrower (sharp ecotone) or wider (spatial mosaic) compared to the case of random movement (Kruuk 1997). This argument serves as a plausible explanation of the observed variation among *Bombina* transects, if we can show: (1) that the habitat distribution is indeed variable and shows the expected correlation with cline width; and (2) that the toads exert a preference in all the transects in question.

As described above, the habitat distribution in the Stryi valley features a sharp ecotone at which abrupt transitions in elevation, forest cover, and aquatic habitat coincide, whereas in Pešćenica the transitions in elevation and forest cover were dissociated in the middle section of the transect, which also featured a mosaic distribution of aquatic sites. This section produced the maximum cline width estimate for this transect. The minimum occurred where open, arable land stretched all the way to the foot of the forest-covered hills (MacCallum 1994). The transect in Kraków is located on a much more gradual elevational transition from the foothills of the Carpathians into the Vistula valley, and the center of the cline just south of the river is not associated with any obvious environmental contour (Szymura and Barton 1986). In contrast, the Przemyśl transect, which is the closest to Stryi, coincides more closely with the edge of the hillside (Szymura and Barton 1991). Finally, the Apahida hybrid zone is located in an open landscape of rolling hills with only small patches of forest and about 20 km away from pure *B. variegata* populations in the mountains. Across all these hybrid zones, cline width is indeed associated with the sharpness of the environmental gradient: it increases from Stryi over the two Polish transects to Pešćenica, while no spatial allele frequency gradient exists at all in the extended habitat mosaic of Apahida (Table 7).

The adult habitat preference for ponds versus puddles has been most thoroughly investigated in the Pešćenica transect (MacCallum et al. 1998), where mark-recapture data and small-scale associations between genotypes and habitat pro-

<table>
<thead>
<tr>
<th>Table 6. Characteristics of populations with unimodal (Stankiv) and bimodal (Dolishnie) distribution of genotype frequencies.</th>
<th>Stankiv unimodal</th>
<th>Dolishnie bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_i )</td>
<td>398</td>
<td>930</td>
</tr>
<tr>
<td>( \bar{p} )</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>( F_{ST} )</td>
<td>0.05 (0.0, 0.2)</td>
<td>0.67 (0.57, 0.75)</td>
</tr>
<tr>
<td>( R )</td>
<td>0.31 (0.20, 0.41)</td>
<td>0.76 (0.70, 0.80)</td>
</tr>
<tr>
<td>( P_{obs} (F_1) )</td>
<td>0.073, ( P = 0.2 )</td>
<td>0.055, ( P = 0.038 )</td>
</tr>
<tr>
<td>( P_{exp} (F_1) )</td>
<td>0.055</td>
<td>0.015</td>
</tr>
</tbody>
</table>
Table 7. Comparison of characteristics and quantitative parameters of the hybrid zones in different geographical regions. The references are (from left to right): present paper; Szymura and Barton (1986, 1991); MacCallum (1994); MacCallum et al. (1998); Vines et al. (2003); Gollmann (1987).

<table>
<thead>
<tr>
<th>Location</th>
<th>Stryi (Ukraine)</th>
<th>Kraków, Przemysł (Poland)</th>
<th>Pešćenica (Croatia)</th>
<th>Apahida (Romania)</th>
<th>Mátra Mountains (Hungary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybridizing Bombina</td>
<td>Carpathian B. variegata and northern B. bombina</td>
<td>Carpathian B. variegata and northern B. bombina</td>
<td>western B. variegata and southern B. bombina</td>
<td>Carpathian B. variegata and southern B. bombina</td>
<td>Isolated B. variegata and southern B. bombina</td>
</tr>
<tr>
<td>Hybridizing populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude (m a.s.l.)</td>
<td>250–420</td>
<td>200–300</td>
<td>100–250</td>
<td>200–500</td>
<td>250–600</td>
</tr>
<tr>
<td>Landscape transition</td>
<td>forest to open land, abrupt change of altitude, vegetation and aquatic habitat</td>
<td>forest to open land, gradual change of altitude, cline center in the river valley (Kraków) or at the altitudinal transition (Przemyśl)</td>
<td>forest to open land, gradual change of altitude, bulge of lowland forest with a mosaic of aquatic habitat types</td>
<td>mosaic, no forest, rolling hillsides</td>
<td>forest to open land, sharp change of altitudes</td>
</tr>
<tr>
<td>Pattern of introgression</td>
<td>clinal</td>
<td>clinal to mosaic</td>
<td>mosaic</td>
<td>one-directional, occasional</td>
<td></td>
</tr>
<tr>
<td>Distribution of hybrid index</td>
<td>unimodal to bimodal</td>
<td>unimodal</td>
<td>unimodal to flat</td>
<td>unimodal to bimodal</td>
<td></td>
</tr>
<tr>
<td>N (individuals)</td>
<td>1146</td>
<td>3014</td>
<td>1764</td>
<td>1034</td>
<td>75</td>
</tr>
<tr>
<td>N (sites)</td>
<td>72</td>
<td>57</td>
<td>147</td>
<td>116</td>
<td>8</td>
</tr>
<tr>
<td>Cline width, w (km)</td>
<td>2.3</td>
<td>6.1</td>
<td>0.06–6.7 km¹</td>
<td>no cline</td>
<td>no cline</td>
</tr>
<tr>
<td>σ (km/gen⁰¹²)</td>
<td>0.50</td>
<td>0.99</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F osob²</td>
<td>0.079 (0.057, 0.11)</td>
<td>0.018 (0.012, 0.028)</td>
<td>0.068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F ST³</td>
<td>−0.018</td>
<td>0.0083</td>
<td>0.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>max F 1SC</td>
<td>0.30 (0.23–0.36)</td>
<td>0.014 (−0.01–0.04)</td>
<td>0.23 (0.05–0.55)</td>
<td>0.21 (0.01–0.5)</td>
<td></td>
</tr>
<tr>
<td>max LD</td>
<td>0.38 (0.34–0.38)</td>
<td>0.22 (0.16–0.29)</td>
<td>0.39</td>
<td>0.38</td>
<td>0.30</td>
</tr>
<tr>
<td>Δρ max²</td>
<td>(0.69)</td>
<td>—</td>
<td>0.16</td>
<td></td>
<td>0.30</td>
</tr>
</tbody>
</table>

¹ Difference between expected and observed F ST from residual variation per locus around the concordance regressions, summed over loci.
² F ST from residual variation of β around spatial cline fit.
³ Width was computed from a model that fitted separate clines for each habitat type. Because this absorbed a large part of the scatter in the cline center, the resulting estimates of w were smaller compared to the fit of a single cline over all sites.
⁴ Estimated difference in mean allele frequency between pairs of ponds and puddles that are no more than 1 km apart. In Stryi, the figure in parentheses is computed only for those pairs that span the cline center.
vided evidence for extensive movement of adults on a larger spatial scale (max: 1.5 km) than the observed genotype-habitat correlation and direct evidence for nonrandom movement between two neighboring sites of different habitat type. Also, in seven of eight regions of the hybrid zone defined on the basis of terrestrial habitat, the *B. variegata* allele frequency was higher in puddles than in ponds. A highly significant, small-scale correlation between habitat and genotype in a series of 13 artificial pools similarly indicates a preference in Apahida (Vines et al. 2003). Although no detailed ecological data on aquatic sites are available for the Polish transects, temporary habitat at Kraków was more common on the *B. variegata* side, whereas artificial ponds existed throughout the region but tended not to be used by *B. variegata* in the south (Szymura and Barton 1986).

In Stryi, allele frequencies in ponds and puddles differ in the expected direction only in three regions with intermediate allele frequencies, which are all located near the cline center. If pairs of sites within 1 km of each other are considered, then a significant regression of $\Delta \rho$ on $\Delta H_{aq}$ rests on six pairs of sites that span the center of the hybrid zone. In other words, the expected patterns arise where the variance in aquatic habitat is greatest. Associations between habitat and allele frequency might similarly exist away from the center, but they will be more difficult to detect, if most aquatic habitat there tends to be either puddle- or pondlike (cf. Fig. 5). Moreover, the very high estimate of $\Delta \rho = 0.69$ across the cline center does not accurately reflect the preference strength either, because right at the central allele-frequency step it is confounded with selection. In essence, the effects of selection and habitat preference cannot be separated with our dataset, if they jointly act to steepen the cline at a sharp ecotone.

More generally, the parental taxa are known to reproduce in puddles and semipermanent ponds, respectively, throughout their distribution ranges, that is, in the absence of the other taxon and even though they must frequently encounter the opposite habitat type (Günther and Schneeweiss 1996; Nöllert and Günther 1996). The suite of taxon differences (see introduction) strongly suggests that this choice is adaptive. As long as hybrid individuals also express habitat preference, if in diminished strength, it should play a role in the hybrid zone. In particular, the coincident transitions of aquatic habitat, vegetation cover, and elevation at Stryi should aid in its manifestation. An indication of this comes from the Dolishnie site, where more *B. variegata*-like individuals were found closer to the forest edge, whereas *B. bombina*-like hybrids tended to be out in the open. The same pattern was also observed by MacCallum (1994) in a Pešćenica site at the forest edge.

If we accept for the moment that regional variation in hybrid zone shape is largely determined by the joint effects of habitat distribution and habitat preference, then does this hypothesis offer new plausible interpretations for the parameter estimates at Stryi? The smaller dispersal estimate in Stryi is inflated by selection in the cline center.

Of course, we do not rule out that intrinsic differences between *Bombina* populations across Europe also play a role. In fact, it would be very likely that allelic variation among regional subgroups of the pure taxa causes variation in the compatibility of the genomes across transects. But our analysis shows that all the necessary components are in place to cause variation in hybrid zone structure as a consequence of the distinctly different habitat distributions across transects.

It is interesting to compare the *Bombina* case with other hybrid systems where comparative mapping of habitat and markers/traits has been carried out across transects. For example, in the contact zone between two subspecies of the land snail *Albinaria hyppolyti*, the narrower transect occurs at a sharp ecotone (Schilthuizen and Lombaerts 1995). But even the wider transect is still narrow relative to the locally very gradual transition of habitat. In fact, the authors argue that exogenous selection might not be very important in this hybrid zone. Narrow clines relative to the environmental gradient were also found in the detailed analysis of two *Gryllus* transects (Ross and Harrison 2002). In these cases, endogenous selection might create a sufficiently strong barrier to gene flow, which is independent of habitat. Therefore the habitat-related traits cannot follow the environmental gradient more closely. Thus, while a hybrid zone may be fixed in place by an environmental gradient, its properties as a barrier to gene flow might still depend on a more complex interplay of forces. Related observations are the existence of particularly steep clines at physical barriers to gene flow (e.g., rivers; Jackson 1992; Raufaste et al. 2005). Note, however, that the Stryi River, which runs perpendicular to the hybrid zone, appears to have the opposite effect: the wide river cline introduces *B. variegata* alleles far into *B. bombina* territory and is most likely a result of passive transport during frequent flooding events. Barriers to dispersal can also be created by a low density of suitable habitats. In the *Bombina* hybrid zone in Austria, genetically distinct populations that were separated by just a few hundred meters had little opportunity to exchange genes because previously suitable intervening habitat had been degraded (Gollmann 1984, 1996). Low-quality habitats may reduce the fitness of those individuals that do actually settle there (thus creating a density trough; Barton 1980), and so center the hybrid zone on it (e.g., as seen in *Podisma* grasshoppers; Barton and Hewitt 1981). All of these cases should differ subtly in the effective selection experienced by neutral genetic variants in the center and in the periphery of the hybrid zone. Their net effect on the strength of barriers to gene flow remains to be investigated in theoretical analyses.
Heterogeneity in Local Population Structure

The Bombina hybrid zone has been cited as a perfect example of a unimodal hybrid zone, where existing isolation mechanisms fail to keep the taxa separate (Jiggins and Mallet 2000). However, the data from Dolishnie and Stankiv provide striking examples of bimodal as well as unimodal local hybrid populations in the center of the same transect. A few bimodal populations were also present in most other Bombina hybrid zones such as Kraków (Szymura and Barton 1991), Pešćenica (MacCallum et al. 1998), Kostajnica (Szymura 1993), in the Slovak karst region (Gollmann et al. 1988), and in the Hungarian Matra Mountains (Gollmann 1987). These may in some cases be transient patterns. In the Slovak karst region, for example, where B. bombina occurs on a high plateau, B. variegata sporadically immigrates into some of these same sites. Also, when heavy rainfall facilitates movement, bimodal aggregations may be found in habitat that serves merely as a stop-over place for both taxa. Similarly, in the hybrid zone between Chorthippus grasshoppers, local extinctions generate vacant patches that are then recolonized by both pure taxa (Bridle and Butlin 2002). As the example of the Dolishnie population shows, habitat heterogeneity within a site may also lead to a bimodal distribution of genotypes. It is unknown whether the effect is strong enough to cause assortative mating and/or whether the genotype distribution is maintained by the continued influx of relatively pure genotypes. Analyses of mating patterns in Bombina hybrid populations have so far been carried out in ecologically homogeneous sites and failed to reject the null hypothesis of random mating (Nürnbergber et al. 2005; Vines and Barton 2003). Given the sporadic occurrence and potentially variable fate of bimodal populations the Bombina hybrid zone and elsewhere, we caution against the inference of long-term dynamics of a complex genetic interaction between differentiated taxa based on snapshot data of local genotype distributions.

To conclude, we have presented an analysis of a very narrow clinal hybrid zone between B. bombina and B. variegata near Stryi, Ukraine, located on a sharp ecotone where transitions in vegetation cover, elevation, and aquatic habitat coincide. We argue that habitat preference plays a significant role in separating the two taxa, as it does in previously studied transects near Pešćenica, Croatia (MacCallum et al. 1998), and near Aplahida, Romania (Vines et al. 2003). In the Stryi valley, we infer that habitat preference results in reduced dispersal across the ecotone and so creates a steeper allele-frequency gradient than would be expected under random movement. A key question remains, however. What is the combined effect of habitat distribution and habitat preference on the barrier to gene flow relative to the case of random movement across an environmental step? At this point we can only speculate. The preference generates assortative mating by habitat and so tends to maintain parental gene combinations. Those individuals that make mistakes introduce relatively less recombinant genomes into the opposite gene pool and should experience relatively stronger selection. A stronger barrier to gene flow results. But in an extended mosaic there will actually be more exchange between the two gene pools because of many more points of contact. The interplay of these forces is complex (Kruuk 1997) and awaits thorough theoretical analysis.

Acknowledgments

Much of the present work was done during the Ph.D. studentship of A. Yanchukov in the Schmalhausen Institute of Zoology in Kiev, Ukraine. D. Kotyashko helped much with the fieldwork, and E. Poljakova and A. Efimenko kindly provided accommodations in Styri. We thank G. Gollmann, R. Harrison, A. Porter, T. Vines, and an anonymous reviewer for helpful comments on an earlier version of the manuscript and C. MacCallum for permission to use the habitat score data for Pešćenica. This work was supported by a Ph.D. stipend from the National Academy of Sciences of Ukraine, short-term stipends from DAAD and Queen Jadwiga Fund at Jagiellonian University of Kraków, and the Royal Society of London NATO/PCO fellowship to AY, KBN 3P04C 055 24 grant to JMS and SH, and DFG grant no. Nu51/2-1 to BN.

Literature Cited

Amphibien und Reptilien Deutschlands. Gustav Fischer Verlag, Jena, Germany.


Corresponding Editor: R. Harrison
ERRATUM

In the March 2006 issue of *Evolution*, Vol. 60, No. 3, there is an incorrect Figure 2A on page 591 in the article “Hybridization of *Bombina bombina* and *B. variegata* (Anura, Dicroglossidae) at a sharp ecotone in western Ukraine: comparisons across transects and over time,” by Alexey Yanchukov, Sebastian Hofman, Jacek M. Szymura, Sergey V. Mezhzherin, Sviatoslav Y. Morozov-Leonov, Nicholas H. Barton, and Beate Nürnberg. This error has no effect on the general message of the figure and thus does not affect the main conclusion of the paper. The entire Figure 2 is provided below.

**Fig. 2.** Distribution of (A) $F_{IS}$ and (B) $R$, estimated by the maximum likelihood, in sites grouped by *Bombina variegata* allele frequency, $P$. Pluses represent 95% confidence support limits.